



Informatica®

10.5.2

Guía de transformaciones de Developer

Informatica Guía de transformaciones de Developer
10.5.2
Abril 2022

© Copyright Informatica LLC 2009, 2022

Este software y la documentación se proporcionan exclusivamente en virtud de un acuerdo de licencia independiente que contiene restricciones de uso y divulgación. Ninguna parte de este documento puede ser reproducida o transmitida de cualquier forma o manera (electrónica, fotocopia, grabación o mediante otros métodos) sin el consentimiento previo de Informatica LLC.

Las bases de datos, el software y los programas de DERECHOS DEL GOBIERNO DE LOS ESTADOS UNIDOS, y la documentación e información técnica relacionadas entregadas a los clientes del Gobierno de los Estados Unidos constituyen "software informático comercial" o "datos técnicos comerciales" de acuerdo con el Reglamento de Adquisición Federal y las regulaciones complementarias específicas del organismo que correspondan. Como tales, el uso, la duplicación, la divulgación, la modificación y la adaptación están sujetos a las restricciones y los términos de licencia establecidos en el contrato gubernamental aplicable, y hasta donde sea aplicable en función de los términos del contrato gubernamental, a los derechos adicionales establecidos en FAR 52.227-19, Licencia de Software Informático Comercial.

Informatica, el logotipo de Informatica y PowerCenter son marcas comerciales o marcas comerciales registradas de Informatica LLC en los Estados Unidos y en muchas otras jurisdicciones de todo el mundo. La lista actual de marcas comerciales de Informatica está disponible en Internet en <https://www.informatica.com/trademarks.html>. Otros nombres de productos y empresas pueden ser nombres o marcas comerciales de sus respectivos titulares.

En relación con los derechos de cancelación de participación, el software transmitirá automáticamente a Informatica en Estados Unidos información sobre el entorno informático y de red en el que se implementa el Software y el uso de los datos y las estadísticas del sistema de la implementación. Esta transmisión se considera parte de los Servicios que se describen en la directiva de privacidad de Informatica, e Informatica usará y procesará esta información según lo descrito en la directiva de privacidad de Informatica que se encuentra en <https://www.informatica.com/in/privacy-policy.html>. Puede deshabilitar la recopilación de usos en la Herramienta del administrador.

Algunas partes de este software o la documentación están sujetas a derechos de autor de terceros. Los avisos obligatorios de terceros se incluyen con el producto.

Consulte las patentes en <https://www.informatica.com/legal/patents.html>.

La información contenida en esta documentación está sujeta a cambios sin previo aviso. Si encuentra algún problema en esta documentación, escríbanos a infa_documentation@informatica.com para notificarnoslo.

Los productos de Informatica gozan de garantía en función de los términos y condiciones de los acuerdos conforme a los cuales se proporcionen. **INFORMATICA PROPORCIONA LA INFORMACIÓN DE ESTE DOCUMENTO "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, EXPRESA O IMPLÍCITA, INCLUIDAS LAS GARANTÍAS DE COMERCIALIZACIÓN, ADAPTACIÓN A UN FIN PARTICULAR Y CUALQUIER GARANTÍA O CONDICIÓN DE NO INCUMPLIMIENTO.**

Fecha de publicación: 2022-07-05

Tabla de contenido

Prefacio	30
Recursos de Informatica	30
Informatica Network.	30
Base de conocimiento de Informatica.	30
Documentación de Informatica	31
Matrices de disponibilidad de producto de Informatica.	31
Informatica Velocity.	31
Catálogo de soluciones de Informatica.	31
Servicio internacional de atención al cliente de Informatica.	31
Capítulo 1: Introducción a las transformaciones.....	32
Introducción al resumen de las transformaciones.	32
Transformaciones activas.	33
Transformaciones pasivas.	33
Transformaciones no conectadas.	33
Transformaciones multiestrategia.	33
Descripciones de transformaciones.	34
Transformaciones en los entornos nativo y no nativo.	36
Administración de los tipos de datos de transformación.	39
Tipo de datos Decimal.	39
Marca de tiempo con zona horaria.	41
Marca de tiempo con zona horaria local.	41
Cómo desarrollar una transformación.	41
Transformaciones de varios grupos.	42
Reglas y directrices para las transformaciones de varios grupos.	42
Expresiones en transformaciones.	43
El editor de expresiones.	44
Nombres de puerto en una expresión.	44
Cómo añadir una expresión a un puerto.	45
Comentarios en una expresión.	45
Validación de expresiones.	45
Probar expresiones.	45
Conversión de tipo de datos.	46
Variables locales.	47
Almacenar datos temporalmente y simplificar expresiones complejas.	47
Almacenar valores en las filas.	48
Capturar valores desde procedimientos almacenados.	49
Directrices para configurar puertos de variables.	49
Inicialización de variables.	50
Valores predeterminados para los puertos.	50

Valores predeterminados definidos por el usuario.	51
Valores de entrada predeterminados definidos por el usuario.	53
Validación de valores predeterminados.	54
Valores de salida predeterminados definidos por el usuario	54
Reglas generales para los valores predeterminados.	56
Validación de valores predeterminados.	57
Niveles de seguimiento.	57
Transformaciones reutilizables.	58
Instancias de transformación reutilizable y cambios heredados.	58
Cómo editar una transformación reutilizable.	58
Vistas de editor para una transformación de reutilizable.	59
Transformaciones no reutilizables.	59
Vistas de editor para una transformación de no reutilizable.	59
Cómo crear una transformación.	60
Capítulo 2: Puertos de transformación.	61
Resumen de puertos de transformación.	61
Crear puertos.	61
Configurar puertos.	62
Cómo vincular puertos.	62
Vínculos "uno a uno".	63
Vínculos "uno a varios".	63
Cómo vincular puertos manualmente.	63
Cómo vincular puertos automáticamente.	63
Reglas y directrices para vincular puertos.	64
Cómo propagar atributos de puerto.	65
Tipos de dependencia.	65
Dependencias de la ruta de acceso al enlace.	65
Dependencias implícitas.	66
Atributos de puerto propagados por transformación.	66
Copiar puertos desde Excel.	68
Edición de transformaciones en Excel.	69
Copiar metadatos a Developer tool.	69
Ejemplo: Editar una transformación en Excel.	70
Reglas y directrices para copiar desde Excel.	71
Capítulo 3: Memorias caché de transformaciones.	72
Resumen de memorias caché de transformaciones.	72
Tipos de memoria caché.	73
Archivos de memoria caché.	73
Directorio de archivos de memoria caché.	74
Tamaño de la memoria caché.	74
Tamaño de la memoria caché automático.	75

Tamaño de la memoria caché específico.	76
Aumento del tamaño de la memoria caché por el Servicio de integración de datos.	77
Tamaño de la memoria caché para memorias caché con particiones.	77
Optimización del tamaño de la memoria caché.	78
Paso 1. Establezca el nivel de seguimiento en inicialización detallada.	78
Paso 2. Ejecute la asignación en modo de memoria caché automática.	78
Paso 3. Analice el rendimiento del almacenamiento en la memoria caché.	78
Paso 4. Configure tamaños de memoria caché específicos.	79

Capítulo 4: Transformación del validador de direcciones. 81

Resumen de la transformación del validador de direcciones.	82
Datos de referencia de direcciones.	82
Tipos de datos de referencia de direcciones.	83
Modos y plantillas.	85
Grupos de puertos y selección de puertos.	85
Grupos de puertos de entrada de la transformación del validador de direcciones.	85
Grupos de puertos de salida de transformación del validador de direcciones.	86
Puertos de varias instancias.	89
Proyectos de validación de direcciones.	90
Direcciones con formato y estándares del proveedor de servicios de correo.	92
Finalización de direcciones parciales.	93
Puertos de estado del validador de direcciones.	93
Definiciones de código de estado de elemento.	94
Valores del puerto de salida del código de resolución de direcciones.	96
Valores del puerto de salida del estado de entrada de elementos.	97
Valores del puerto de salida de relevancia de elementos.	98
Valores del puerto de salida del estado de resultado de elementos.	98
Valores de puerto de salida del estado de resultado de elementos extendido.	100
Valores del puerto de salida de la puntuación de viabilidad de envío de correo.	102
Valores del puerto de salida del código de coincidencia.	103
Valores del puerto de salida del estado de geocodificación.	105
Configuración general de la transformación del validador de direcciones.	106
Propiedades de validación de direcciones en la ventana Preferencias.	107
Propiedades de datos de validación de direcciones.	108
Propiedades de licencia de validación de direcciones.	108
Propiedades del motor de validación de direcciones.	109
Propiedades avanzadas de validación de direcciones.	110
Alias de localidad.	110
Alias de calle.	111
Uso de mayúsculas y minúsculas.	111
País de origen.	112
Tipo de país.	112
País predeterminado.	113

Prioridad de dirección dual.	114
Abreviatura de elemento.	114
Instancias de ejecución.	114
Expansión de rango flexible.	115
Tipo de datos de geocodificación.	116
Longitud de campo máxima global.	116
Descriptor global preferido.	117
Tipo de formato de entrada.	117
Formato de entrada con país	118
Separador de línea.	118
Alternativas coincidentes.	118
Coincidencia de archivo extendido.	119
Ámbito coincidente.	120
Recuento máximo de resultados.	120
Modo	120
Nivel de optimización.	121
Tipo de formato de salida.	121
Formato de salida con país.	122
Idioma preferido.	122
Codificación de caracteres preferida.	129
Rangos para expandir.	129
Estandarizar direcciones no válidas.	130
Nivel de seguimiento.	131
Informes de certificación.	131
Campos de informe AMAS.	132
Campos de informe CASS.	132
Informe SendRight.	133
Campos de informe SERP.	134
Cómo configurar una transformación del validador de direcciones..	134
Cómo añadir puertos a la transformación del validador de direcciones.	135
Cómo crear plantillas definidas por el usuario.	135
Cómo definir modelos del validador de direcciones.	136
Cómo definir un informe de certificación.	136
Transformación del validador de direcciones en un entorno no nativo.	137
Transformación del validador de direcciones en el motor de Blaze.	137
Transformación del validador de direcciones en el motor de Spark.	137
Transformación del validador de direcciones en el motor de Databricks Spark.	137
Capítulo 5: Transformación de agregación.	138
Resumen de la transformación de agregación.	138
Transformaciones de agregación en asignaciones dinámicas.	139
Cómo desarrollar una transformación de agregación.	139
Puertos de transformación de agregación.	139

Expresiones de agregado.	140
Funciones de agregado.	141
Funciones de agregado anidadas.	142
Cláusulas condicionales en expresiones de agregado.	142
Agrupar por puertos.	142
Configurar la agrupación por puertos.	143
Parámetros de la agrupación por puertos.	144
Valores predeterminados de Agrupar por puertos	144
Expresiones de no agregado.	145
Memorias caché de agregación.	145
Entrada ordenada para una transformación de agregación.	145
Condiciones de entrada ordenada.	146
Cómo ordenar datos en una transformación de agregación.	146
Propiedades avanzadas de la transformación de agregación.	147
Cómo crear una transformación de agregación no reutilizable.	148
Cómo crear una transformación de agregación no reutilizable.	148
Consejos para transformaciones de agregación.	149
Solución de problemas de transformaciones de agregación.	149
Transformación de agregación en un entorno no nativo.	150
Transformación de agregación en el motor de Blaze.	150
Transformación de agregación en el motor de Spark.	151
Transformación de agregación en el motor de Databricks Spark.	151
Capítulo 6: Transformación de asociación.	153
Resumen de la transformación de asociación.	153
Asignación de memoria.	154
Propiedades avanzadas de transformación de asociación.	155
Capítulo 7: Transformación de excepción de registros incorrectos.	156
Resumen de la transformación de excepción de registros incorrectos.	156
Tipos de registros de salida de excepción de registros incorrectos.	157
Flujo de proceso de administración de excepciones de registros incorrectos.	158
Asignaciones de excepción de registros incorrectos.	159
Problemas de calidad de excepción de registros incorrectos.	160
Tareas humanas.	160
Puertos de excepción de registros incorrectos	161
Puertos de entrada de transformación de excepción de registros incorrectos.	161
Salida de transformación de excepción de registros incorrectos.	162
Vista de configuración de excepción de registros incorrectos.	162
Cómo generar la tabla de registros incorrectos y la tabla de problemas.	164
Asignación problemática de excepción de registros incorrectos	164
Asignar puertos a problemas de calidad.	165
Propiedades avanzadas de Transformación de excepción.	165

Cómo configurar una transformación de excepción de registros incorrectos.	166
Ejemplo de asignación de excepción de registros incorrectos.	167
Mapplet de excepción de registros incorrectos.	167
Grupos de entrada de ejemplo de excepción de registros incorrectos.	168
Configuración de ejemplo de excepción de registros incorrectos	168
Salida de asignación de ejemplo de excepción de registros incorrectos.	169
Capítulo 8: Transformación de conversión de mayúsculas y minúsculas.	172
Resumen de la transformación de conversión de mayúsculas y minúsculas.	172
Propiedades de la estrategia de distinción de mayúsculas y minúsculas.	173
Cómo configurar una estrategia de conversión de mayúsculas y minúsculas.	173
Propiedades avanzadas de Transformación de conversión de mayúsculas y minúsculas.	174
Transformación de conversión de mayúsculas y minúsculas en un entorno no nativo.	174
Capítulo 9: Transformación de clasificador.	175
Resumen de la transformación de clasificador.	175
Modelos de clasificador.	176
Algoritmos clasificadores.	176
Opciones de la transformación de clasificador.	176
Estrategias de clasificador.	177
Propiedades avanzadas de Transformación de clasificador.	177
Cómo configurar una estrategia de clasificador.	178
Ejemplo de análisis de clasificador.	178
Crear la asignación de clasificador.	179
Muestra de datos de entrada.	180
Configuración de origen de datos.	180
Configuración de transformación de clasificador.	180
Configuración de transformación de enrutador.	180
Configuración de destino de datos.	182
Resultados de la asignación de clasificador.	182
Transformación de clasificador en un entorno no nativo.	183
Capítulo 10: Transformación de comparación.	184
Resumen de la transformación de comparación.	184
Estrategias para la coincidencia de campos.	184
Bigram.	185
Distancia de Hamming.	185
Distancia de edición.	185
Distancia de Jaro.	186
Invertir la distancia de Hamming.	186
Estrategias de coincidencia de identidades.	187
Cómo configurar una estrategia de comparación.	188
Propiedades avanzadas de Transformación de comparación.	188

Transformación de comparación en un entorno no nativo.	189
Capítulo 11: Transformación de consolidación.	190
Resumen de la transformación de consolidación.	190
Asignaciones de consolidación.	191
Puertos de transformación de consolidación.	191
Vistas de transformación de consolidación.	192
Vista Estrategias de la transformación de consolidación.	192
Propiedades avanzadas de Transformación de consolidación.	192
Tamaño de archivo de memoria caché.	193
Estrategias simples.	194
Estrategias basadas en filas.	195
Estrategias avanzadas.	196
Funciones de consolidación simple.	196
CONSOL_AVG.	196
CONSOL_LONGEST.	197
CONSOL_MAX.	198
CONSOL_MIN.	198
CONSOL_MOSTFREQ.	199
CONSOL_MOSTFREQ_NB.	199
CONSOL_SHORTEST.	200
Funciones de consolidación basada en filas.	201
CONSOL_GETROWFIELD.	201
CONSOL_MODAL EXACT.	202
CONSOL_MOSTDATA.	202
CONSOL_MOSTFILLED.	203
Ejemplo de asignación de consolidación.	204
Datos de entrada.	205
Transformación de generador de claves.	205
Transformación de consolidación.	205
Salida de la asignación de consolidación.	205
Cómo configurar una transformación de consolidación.	206
Transformación de consolidación en un entorno no nativo.	206
Transformación de consolidación en el motor de Blaze.	206
Transformación de consolidación en el motor de Spark.	206
Transformación de consolidación en el motor de Databricks Spark.	207
Capítulo 12: Transformación de enmascaramiento de datos.	208
Resumen de transformación de enmascaramiento de datos.	208
Técnicas de enmascaramiento.	209
Enmascaramiento aleatorio.	209
Enmascaramiento de expresiones.	211
Enmascaramiento de claves.	213

Máscara de sustitución.	215
Enmascaramiento dependiente.	218
Enmascaramiento de Tokenization.	220
Cifrado.	220
Reglas de enmascaramiento.	222
Formato de máscara.	222
Caracteres de cadena de origen.	223
Caracteres de reemplazo de cadena de resultado.	224
Intervalo.	225
Desenfoque.	225
Formatos de máscara especiales.	226
Enmascaramiento del número de la tarjeta de crédito.	226
Enmascaramiento de la dirección de correo electrónico.	227
Enmascaramiento de correo electrónico avanzado.	227
Enmascaramiento de direcciones IP.	228
Enmascaramiento del número de teléfono.	229
Enmascaramiento del número de seguridad social.	229
Enmascaramiento de direcciones URL.	230
Enmascaramiento del número de seguro social.	230
Archivo de valores predeterminados.	230
Configuración de la transformación de enmascaramiento de datos.	231
Configuración del Servicio de integración de datos.	231
Cómo crear una transformación de enmascaramiento de datos.	232
Cómo definir los puertos.	232
Cómo configurar un enmascaramiento de datos para cada puerto.	232
Vista previa de los datos enmascarados.	233
Propiedades de tiempo de ejecución de la transformación de enmascaramiento de datos.	233
Ejemplo de enmascaramiento de datos.	234
Read_Customer Data.	235
Transformación de enmascaramiento de datos de clientes.	235
Resultados de datos de prueba de cliente.	236
Propiedades avanzadas de Transformación de enmascaramiento de datos.	237
Transformación de enmascaramiento de datos en un entorno no nativo.	237
Transformación de enmascaramiento de datos en el motor de Blaze.	237
Transformación de enmascaramiento de datos en el motor de Spark.	238
Capítulo 13: Transformación del Procesador de datos.	240
Resumen de la transformación del Procesador de datos.	240
Vistas de la transformación del Procesador de datos.	241
Puertos de la transformación del Procesador de datos.	242
Puertos de entrada de la transformación del Procesador de datos.	242
Puertos de salida de la transformación del Procesador de datos.	243
Puertos de transferencia.	244

Componente de inicio.	244
Referencias.	245
Configuración de la transformación del Procesador de datos.	246
Codificación de caracteres.	246
Reglas y directrices de codificación de caracteres.	248
Configuración de salida.	248
Configuración de procesamiento.	250
Configuración de XMap.	251
Configuración de salida XML.	251
Eventos.	253
Tipos de evento.	253
Vista Eventos del procesador de datos.	254
Registros.	254
Registro de eventos en tiempo de diseño.	255
Registro de eventos de tiempo de ejecución.	255
Visualizar un registro de eventos en la vista Eventos del Procesador de datos.	256
Registro de usuario.	256
Desarrollo de la transformación del Procesador de datos.	257
Crear la transformación del Procesador de datos.	257
Seleccionar los objetos de esquema	258
Crear objetos en una transformación del Procesador de datos en blanco.	258
Crear los puertos.	261
Probar la transformación.	261
Exportación e importación de transformaciones del Procesador de datos.	261
Exportar la transformación del Procesador de datos como un servicio.	262
Importar varios servicios de Data Transformation.	262
Importar un servicio de Data Transformation	263
Exportación de una asignación con una transformación del Procesador de datos a PowerCenter.	263
Transformación del procesador de datos Validación.	264
Utilizar un motor de Data Transformation con velocidad mejorada para las validaciones de VRL.	264
Transformación del procesador de datos en un entorno no nativo.	265
Capítulo 14: Transformación de decisión.	266
Resumen de la transformación de decisión.	266
Funciones de transformación de decisión.	267
Instrucciones condicionales de transformación de decisión	269
Operadores de transformación de decisión.	270
Control de valores NULL en transformaciones de decisión.	271
Cómo configurar una estrategia de decisión	271
Propiedades avanzadas de Transformación de decisión.	272
Transformación de decisión en un entorno no nativo.	272

Transformación de expresión en el motor de Spark.	272
Transformación de decisión en el motor de Databricks Spark.	272
Capítulo 15: Transformación de excepción de registros duplicados.	273
Resumen de la transformación de excepción de registros duplicados.	273
Flujo del proceso de la excepción de registros duplicados.	274
Excepciones de registros duplicados.	274
Vista de configuración de excepción de registros duplicados	275
Cómo generar una tabla de registros duplicados.	276
Puertos.	277
Puertos de entrada de transformación de excepción de registros duplicados.	277
Puertos de salida de transformación de excepción de registros duplicados.	278
Cómo crear puertos	279
Propiedades avanzadas de la transformación de excepción de registros duplicados.	279
Ejemplo de asignación de excepción de registros duplicados.	280
Asignación de excepción de registros duplicados.	280
Transformación de coincidencia.	281
Grupos de entrada de excepción de registros duplicados.	281
Vista de configuración de ejemplo de excepción de registros duplicados.	282
Registros de tabla de salida estándar.	283
Salida del clúster.	284
Creación de una transformación de excepción de registros duplicados.	286
Capítulo 16: Transformación de expresión.	287
Resumen de la transformación de expresión.	287
Puertos de transformación de expresión.	288
Probar expresiones.	289
Cadenas de formato de fecha para datos de ejemplo.	290
Probar una expresión.	290
Selectores de puerto.	291
Configuración del selector de puerto.	291
Reglas de selección.	292
Crear un selector de puerto.	293
Administración de ventanas.	294
Configuración de la administración de ventanas.	294
Expresiones dinámicas.	299
Configuración del puerto de salida.	300
Crear una expresión dinámica	301
Aplanar una estructura dinámica.	303
Propiedades avanzadas de la transformación de expresión.	304
Transformación de expresión en un entorno no nativo.	305
Transformación de expresión en el motor de Blaze.	305
Transformación de expresión en el motor de Spark.	305

Transformación de expresión en el motor de Databricks Spark.	305
Capítulo 17: Transformación de filtro.	306
Resumen de la transformación de filtro.	306
Transformaciones de filtro en asignaciones dinámicas.	307
Condición de filtro.	308
Parametrizar la condición de filtro.	308
Cómo filtrar filas con valores nulos.	310
Propiedades avanzadas de la transformación de filtro.	310
Consejos sobre el rendimiento de la transformación de filtro.	310
Transformación de filtro en un entorno no nativo.	310
Transformación de filtro en el motor de Blaze.	311
Capítulo 18: Transformación de jerárquica a relacional.	312
Resumen de la transformación de jerárquica a relacional.	312
Ejemplo: transformación de jerárquica a relacional.	313
Puertos de salida relacionales y la vista Resumen.	314
Puertos de la transformación de jerárquica a relacional.	315
Referencias a esquemas.	316
Procesamiento de instrucciones de consulta.	316
Desarrollo de la transformación de jerárquica a relacional.	317
Crear la transformación de jerárquica a relacional.	317
Configurar los puertos y la asignación.	317
Probar la transformación.	318
Capítulo 19: Transformación de Java.	319
Resumen de la transformación de Java.	319
Transformaciones de Java reutilizables y no reutilizables.	320
Transformaciones de Java activas y pasivas.	320
Conversión de tipo de datos.	320
Conversión de tipos de datos complejos en el motor de Spark.	322
Cómo diseñar una transformación de Java.	323
Puertos de transformación de Java.	324
Cómo crear puertos.	324
Cómo configurar valores de puerto predeterminados.	324
Propiedades avanzadas de la transformación de Java.	325
Cómo configurar la classpath para el cliente de Developer Tool.	327
Configurar la classpath para el Servicio de integración de datos.	327
Cómo desarrollar código Java.	328
Cómo crear snippets de código Java.	329
Cómo importar paquetes de Java.	331
Cómo definir el código de una aplicación auxiliar.	331
Propiedades de Java de una transformación de Java.	332

Ficha Importaciones.	332
Ficha Aplicaciones auxiliares.	332
Ficha En entrada.	333
Ficha Al final.	333
Ficha Funciones.	334
Ficha Código completo.	334
Optimización de filtro con la transformación de Java.	334
Optimización de primera selección con la transformación de Java.	335
Optimización de inserción con la transformación de Java.	336
Cómo crear una transformación de Java.	337
Cómo crear una transformación de Java reutilizable.	337
Cómo crear una transformación de Java no reutilizable.	338
Cómo compilar una transformación de Java.	338
Solución de problemas de una transformación de Java.	339
Cómo buscar el origen de errores de compilación.	339
Cómo identificar el origen de errores de compilación.	340
Ejemplo de conversión a datos de tipo struct.	340
Transformación Java en un entorno no nativo.	344
Transformación Java en el motor de Blaze.	344
Transformación Java en el motor de Spark.	345
Capítulo 20: Referencia API de transformación de Java.	347
Resumen de los métodos API de transformación de Java.	347
defineJExpression.	348
failSession.	349
generateRow.	350
getInRowType.	350
getMetadata.	351
incrementErrorCount.	351
invokeJExpression.	352
isNull.	353
logError.	354
logInfo.	354
resetNotification.	355
setNull.	355
storeMetadata.	356
Capítulo 21: Expresiones de Java.	358
Resumen de las expresiones de Java.	358
Tipos de función de expresión.	359
Cómo utilizar el cuadro de diálogo Definir función para definir una expresión.	359
Paso 1. Configurar la función.	360
Paso 2. Crear y validar la expresión.	360

Paso 3. Generar código Java para la expresión.	360
Cómo crear una expresión y generar código Java mediante el cuadro de diálogo Definir función.	360
Plantillas de expresiones de Java.	361
Cómo trabajar con la interfaz simple.	361
invokeJExpression.	362
Ejemplo de interfaz simple.	362
Cómo trabajar con la interfaz avanzada.	363
Cómo invocar una expresión con la interfaz avanzada.	363
Reglas y directrices para trabajar con la interfaz avanzada.	364
Clase EDataType.	364
Clase JExprParamMetadata.	365
defineJExpression.	365
Clase JExpression.	366
Ejemplo de interfaz avanzada.	366
Referencia API de la clase JExpression.	367
getBytes.	368
getDouble.	368
getInt.	368
getLong.	368
getResultDataType.	368
getResultMetadata.	369
getStringBuffer.	369
invoke.	369
isResultNull.	370
Capítulo 22: Transformación de incorporación.	371
Resumen de la transformación de combinación.	371
Propiedades avanzadas de la transformación de combinación.	372
Memorias caché de combinación.	373
Puertos de transformación de combinación.	374
Transformaciones de combinación en asignaciones dinámicas.	375
Selectores de puerto en una transformación de combinación.	375
Reglas de selección	376
Crear un selector de puerto.	377
Definir una condición de combinación.	378
Tipo de condición simple.	379
Tipo de condición avanzada.	380
Selectores de puerto en condiciones de combinación.	380
Puertos dinámicos en condiciones de combinación.	381
Parámetro de expresión.	381
Tipos de unión.	382
Unión normal.	382

Outer Join principal.	383
Outer Join de detalles.	384
Outer Join completo.	384
Entrada ordenada para una transformación de incorporación.	385
Cómo configurar el orden de clasificación.	385
Cómo añadir transformaciones a la asignación.	385
Reglas y directrices para las condiciones de combinación.	386
Ejemplo de una condición de combinación y del orden de clasificación.	387
Cómo unir datos de un mismo origen.	389
Cómo unir dos ramificaciones del mismo canal.	389
Cómo unir dos instancias del mismo origen.	390
Pautas para unir datos de un mismo origen.	391
Cómo bloquear los canales de origen.	391
Transformación de incorporación no ordenada.	391
Transformación de incorporación ordenada.	391
Consejos sobre el rendimiento de la transformación de combinación.	392
Reglas y directrices para una transformación de combinación.	393
Transformación de combinación en un entorno no nativo.	393
Transformación de combinación en el motor de Blaze.	393
Transformación de combinación en el motor de Spark.	394
Transformación de combinación en el motor de Databricks Spark.	394
Capítulo 23: Transformación de generador de claves.	395
Resumen de la transformación de generador de claves.	395
Estrategia Soundex.	396
Propiedades de estrategia Soundex.	396
Estrategia de cadena.	397
Propiedades de la estrategia de cadena.	397
Estrategia NYSIIS.	397
Puertos de salida del generador de claves.	398
Cómo configurar una estrategia de agrupación.	398
Propiedades de la creación de claves.	399
Propiedades avanzadas de Transformación de generador de claves.	399
Transformación del generador de claves en un entorno no nativo.	400
Transformación de generador de claves en el motor de Blaze.	400
Transformación de generador de claves en el motor de Spark.	400
Transformación de generador de claves en el motor de Databricks Spark.	400
Capítulo 24: Transformación de etiquetador.	401
Resumen de la transformación de etiquetador.	401
Cuando utilizar una transformación de etiquetador.	402
Uso de los datos de referencia en la transformación de etiquetador.	403
Juegos de caracteres.	404

Modelos probabilísticos.	404
Tablas de referencia.	404
Expresiones regulares.	405
Conjuntos de tokens.	405
Estrategias de transformación de etiquetador.	405
Operaciones de etiquetado de caracteres.	406
Operaciones de etiquetado con tokens.	406
Puertos de transformación de etiquetador.	406
Propiedades de etiquetado de caracteres.	407
Propiedades generales.	407
Propiedades de la tabla de referencia.	407
Propiedades de los conjuntos de caracteres.	408
Propiedades de filtro.	409
Propiedades de etiquetado con tokens.	409
Propiedades generales.	409
Propiedades de conjuntos de tokens.	410
Propiedades de etiqueta personalizada.	411
Propiedades de coincidencia probabilística.	411
Propiedades de la tabla de referencia.	411
Cómo configurar una estrategia de etiquetado de caracteres.	412
Cómo configurar una estrategia de etiquetado de tokens.	413
Propiedades avanzadas de Transformación de etiquetador.	413
Transformación de etiquetador en un entorno no nativo.	414
Capítulo 25: Transformación de búsqueda.	415
Resumen de la transformación de búsqueda.	415
Búsquedas conectadas y no conectadas.	416
Búsquedas conectadas.	417
Búsquedas no conectadas.	418
Desarrollar una transformación de búsqueda.	419
Consulta de búsqueda.	419
Consulta de búsqueda predeterminada.	419
Reemplazo de SQL para una consulta de búsqueda.	419
Parámetros de una consulta de reemplazo de SQL.	420
Palabras reservadas.	421
Directrices para reemplazar la consulta de búsqueda.	421
Reemplazar la consulta de búsqueda.	421
Filtro de origen de búsqueda.	422
Filtrar filas de origen en una búsqueda.	422
Condición de búsqueda.	422
Configurar la condición de búsqueda.	424
Reglas y directrices para las condiciones de una transformación de búsqueda.	424
Memoria caché de búsqueda.	425

Propiedades de consulta.	426
Transformaciones de búsqueda en asignaciones dinámicas.	426
Definir puertos dinámicos.	427
Cambiar el origen de búsqueda.	427
Parametrizar el origen de la búsqueda.	428
Orígenes de búsqueda que contienen parámetros.	430
Configurar parámetros en un objeto de datos duplicado.	430
Selectores de puerto.	432
Configuración del selector de puerto.	433
Reglas de selección.	434
Parametrizar la condición de búsqueda.	435
Crear un selector de puerto.	436
Propiedades de tiempo de ejecución.	438
Propiedades avanzadas.	439
Crear una transformación de búsqueda reutilizable.	441
Crear una transformación de búsqueda no reutilizable.	442
Crear una transformación de búsqueda no conectada.	443
Ejemplo de búsqueda no conectada.	444
Transformación de búsqueda en un entorno no nativo.	446
Transformación de búsqueda en el motor de Blaze.	446
Transformación de búsqueda en el motor de Spark.	447
Transformación de búsqueda en el motor de Databricks Spark.	449
Transformación de búsqueda en una asignación de transmisión.	449
Capítulo 26: Memorias caché de búsqueda.	451
Introducción a las memorias caché de búsqueda.	451
Tipos de memoria caché de búsqueda.	452
Búsqueda no almacenada en caché.	453
Memoria caché de búsqueda estática.	453
Memoria caché de búsqueda persistente.	454
Regeneración de una memoria caché de búsqueda persistente.	454
Memoria caché de búsqueda dinámica.	455
Memoria caché de búsqueda compartida.	455
Reglas y directrices para compartir una memoria caché de búsqueda.	456
Comparación de memorias caché.	457
Partición de la memoria caché para búsquedas.	457
Capítulo 27: Memoria caché de búsqueda dinámica.	458
Introducción a la memoria caché de búsqueda dinámica.	458
Usos de la memoria caché de búsqueda dinámica.	459
Propiedades de la memoria caché de búsqueda dinámica.	459
Memoria caché de búsqueda dinámica y valores de salida.	461
Valores de transformación de búsqueda.	462

Ejemplo de valores de transformación de búsqueda.	462
Reemplazo de SQL y memoria caché de búsqueda dinámica.	464
Configuración de asignación de una memoria caché de búsqueda dinámica.	465
Insertar o actualizar.	465
Actualizar o insertar.	466
Memoria caché de búsqueda dinámica y destino de sincronización.	467
Actualizaciones de memoria caché de búsqueda dinámica condicional.	468
Procesamiento de la memoria caché de búsqueda dinámica condicional.	468
Configurar una memoria caché de búsqueda dinámica condicional.	469
Actualización de memoria caché dinámica con resultados de expresión.	469
Valores de expresión nulos.	469
Procesamiento de expresiones.	470
Configurar una expresión para actualizaciones de memoria caché dinámica.	470
Ejemplo de memoria caché de búsqueda dinámica.	471
Reglas y directrices de las memorias caché de búsqueda dinámica.	472
Capítulo 28: Transformación de macro.	473
Resumen de la transformación de macro.	473
Instrucción macro.	473
Parámetros de la transformación de macro.	474
Transformaciones de macro en asignaciones dinámicas.	474
Ejemplo de la transformación de macro.	475
Transformación de macro en un entorno no nativo.	476
Capítulo 29: Transformación de coincidencia.	477
Resumen de la transformación de coincidencia.	477
Análisis de coincidencia.	478
Análisis de columnas.	478
Análisis de un solo origen y análisis de origen dual.	479
Análisis de coincidencias de campos y análisis de coincidencia de identidades.	479
Grupos en el análisis de coincidencia.	480
Pares de coincidencia y clústeres.	481
Cálculos de puntuación de coincidencia.	482
Puntuaciones ponderadas.	482
Puntuaciones de coincidencia nulas.	482
Opciones de salida del clúster.	483
Puntuaciones de controlador y puntuaciones de enlace en el análisis de clústeres.	484
Análisis de datos principales.	485
Reutilización de la asignación.	487
Datos del análisis de coincidencia de identidad y del índice persistente.	487
Reglas y directrices para datos de índice persistentes.	487
Rendimiento de asignación de coincidencias.	488
Visualización de datos de análisis del clúster de coincidencias.	489

Visualización de los datos del análisis de rendimiento de coincidencia.	489
Rendimiento de coincidencia en el análisis de identidad.	490
Creación de un almacén de datos para datos de índice de identidad.	491
Utilización del almacén de datos de índice en el análisis de un solo origen.	492
Vistas de transformación de coincidencia.	493
Puertos de transformación de coincidencia.	494
Puertos de entrada de la transformación de coincidencia.	495
Puertos de salida de la transformación de coincidencia.	495
Códigos de estado de persistencia y descripciones de estado de persistencia.	496
Selección de puertos de salida y de salida de coincidencia.	499
Mapplets de coincidencia.	499
Creación de un mapplet de coincidencia.	500
Cómo usar un mapplet de coincidencia.	500
Configuración de una operación de análisis de coincidencia.	501
Transformación de coincidencia en un entorno no nativo.	502
Transformación de coincidencia en el motor de Blaze.	502
Transformación de coincidencia en el motor de Spark.	502
Transformación de coincidencia en el motor de Databricks Spark.	502
Capítulo 30: Transformaciones de coincidencia en el análisis de campos.	503
Análisis de coincidencia de campos.	503
Flujo de proceso para el análisis de coincidencia de campos.	504
Opciones del tipo de coincidencia de campos.	504
Estrategias para la coincidencia de campos.	505
Algoritmos de coincidencia de campos.	505
Propiedades de la estrategia para la coincidencia de campos.	507
Opciones de salida de coincidencia de campos.	508
Tipos de salida de coincidencia.	508
Propiedades de salida de coincidencia.	509
Propiedades avanzadas de la coincidencia de campos.	510
Ejemplo de análisis de coincidencia de campos.	511
Crear la asignación.	511
Muestra de datos de entrada.	512
Configuración de la transformación de generador de claves.	512
Configuración de la transformación de coincidencia.	512
Ejecutar el visor de datos.	515
Conclusión.	515
Capítulo 31: Transformaciones de coincidencia en el análisis de identidad.	516
Análisis de coincidencia de identidades.	516
Flujo de proceso para el análisis de coincidencia de identidades.	517
Propiedades de tipo de coincidencia de identidades.	517
Propiedades del directorio de índice y del directorio de la memoria caché.	520

Parámetros del método de persistencia.	521
Estrategias de coincidencia de identidades.	521
Algoritmos de coincidencia de identidades.	522
Propiedades de la estrategia de coincidencia de identidades.	523
Opciones de salida de coincidencia de identidades.	524
Tipos de salida de coincidencia.	524
Propiedades de salida de coincidencia.	525
Propiedades avanzadas de coincidencia de identidades.	527
Caso práctico de índice persistente.	528
Ejemplo de análisis de coincidencia de identidades.	530
Crear la asignación.	531
Muestra de datos de entrada.	532
Configuración de la transformación de expresión.	532
Configuración de la transformación de coincidencia.	533
Ejecutar el visor de datos.	536
Conclusión.	537
Capítulo 32: Transformación de normalizador.	538
Resumen de la transformación de normalizador.	538
Campos de ocurrencia múltiple.	539
ID de columna generados.	539
Registros de ocurrencia múltiple.	540
Definición de la jerarquía de entrada.	541
Puertos de entrada de la transformación de normalizador.	542
Fusionar campos.	543
Aplanar campos.	544
Grupos y puertos de salida de la transformación de normalizador.	549
Crear un grupo de salida.	551
Actualizar un grupo de salida.	552
Generación de claves para grupos de salida.	553
Propiedades avanzadas de transformación de normalizador.	554
Generar grupos de salida de primer nivel.	554
Creación de una transformación de normalizador.	554
Creación de una transformación de normalizador desde un origen de nivel superior.	555
Ejemplo de asignación de normalizador.	556
Ejemplo de asignación de normalizador.	556
Ejemplo de definición del normalizador.	557
Ejemplo de grupos de entrada y salida del normalizador.	557
Salida de asignación de ejemplo de normalizador.	558
Transformación de normalizador en un entorno no nativo.	559
Capítulo 33: Transformación de fusión.	560
Resumen de la transformación de fusión.	560

Cómo configurar una estrategia de fusión.	560
Propiedades avanzadas de Transformación de fusión.	561
Transformación de fusión en un entorno no nativo.	561
Capítulo 34: Transformación de analizador.	562
Resumen de la transformación de analizador.	562
Modos de transformación del analizador.	563
Cuando utilizar una transformación de analizador.	563
Uso de los datos de referencia en la transformación de analizador.	564
Conjuntos de patrones.	565
Modelos probabilísticos.	565
Tablas de referencia.	566
Expresiones regulares.	566
Conjuntos de tokens.	566
Operaciones de análisis de tokens.	566
Puertos del análisis de tokens.	567
Propiedades del análisis de tokens.	568
Propiedades generales.	568
Propiedades del modelo probabilístico.	569
Propiedades de la tabla de referencia.	569
Propiedades de conjuntos de tokens.	570
Modo de análisis basado en patrones.	570
Puertos de análisis basado en patrones.	571
Configurar una estrategia de análisis de tokens.	571
Configurar una estrategia de análisis de patrones.	572
Propiedades avanzadas de Transformación de analizador.	573
Transformación de analizador en un entorno no nativo.	573
Capítulo 35: Transformación de Python.	574
Capítulo 36: Transformación de rango.	575
Resumen de la transformación de rango.	575
Valores de las cadenas de establecimiento de rango.	576
Propiedades de la transformación de rango.	576
Transformaciones de rango en asignaciones dinámicas.	576
Puertos de la transformación de rango.	577
Índice de rango.	577
Puerto de rango.	578
Definir agrupaciones por puertos.	578
Parámetros de la agrupación por puertos.	579
Memorias caché de rango.	580
Propiedades avanzadas de la transformación de rango.	580
Transformación de rango en un entorno no nativo.	581

Transformación de rango en el motor de Blaze	581
Transformación de rango en el motor de Spark.	582
Transformación de rango en el motor de Databricks Spark.	582
Capítulo 37: Transformación de lectura.	583
Resumen de la transformación de lectura.	583
Propiedades de la transformación de lectura	584
Propiedades generales.	585
Propiedades del objeto de datos.	585
Propiedades de consulta.	586
Propiedades de tiempo de ejecución	586
Propiedades de los orígenes	586
Propiedades avanzadas.	587
Sincronizar objetos de datos relacionales.	587
Cambiar el objeto de datos de origen.	588
Parametrizar la transformación de lectura.	589
Parámetros de la transformación de lectura.	590
Restricciones	591
Crear una transformación de lectura.	591
Crear una transformación de lectura en el editor de asignación.	592
Capítulo 38: Transformación de relacional a jerárquica.	594
Resumen de la transformación de relacional a jerárquica.	594
Ejemplo: transformación de relacional a jerárquica.	595
Puertos relacionales de entrada y vista Resumen.	597
Puertos de transformación de relacional a jerárquica.	598
Referencias a esquemas.	598
Desarrollo de la transformación de relacional a jerárquica.	598
Crear la transformación de relacional a jerárquica.	598
Crear los puertos.	599
Capítulo 39: Transformación de consumidor de servicio web REST.	600
Resumen de la transformación de consumidor de servicio web REST.	600
Proceso de transformación de consumidor de servicio web REST.	602
Configuración de la transformación de consumidor de servicio web REST.	602
Configuración de mensaje.	602
Identificación de recurso.	603
Métodos HTTP.	604
Método HTTP Get.	604
Método HTTP Post.	605
Método HTTP Put.	605
Método HTTP Delete.	606
Puertos de la transformación de consumidor de servicio web REST.	607

Puertos de entrada.	607
Puertos de salida.	607
Puertos de transferencia.	607
Puertos de argumento.	608
Puertos URL.	608
Puertos de encabezado HTTP.	608
Puertos de cookies.	609
Puertos de salida XML.	609
Puertos de código de respuesta.	609
Asignación de entrada de la transformación de consumidor de servicio web REST.	610
Reglas y directrices para asignar puertos de entrada a elementos.	610
Asignación de puertos de entrada a la entrada del método.	611
Asignación de salida de la transformación de consumidor de servicio web REST.	612
Reglas y directrices para asignar elementos a puertos de salida.	613
Personalizar opciones de vista.	613
Asignación de la salida del método a puertos de salida.	613
Propiedades avanzadas de la transformación de consumidor de servicio web REST.	614
Creación de la transformación de consumidor de servicio web REST.	615
Crear una transformación de consumidor de servicio web REST.	615
Análisis de un mensaje de respuesta JSON que contiene matrices.	616
Ejemplo de mensaje de respuesta JSON.	616
Matrices sin nombre en un mensaje de respuesta.	617
Capítulo 40: Transformación de enrutador.	618
Resumen de la transformación de enrutador.	618
Transformaciones de enrutador en asignaciones dinámicas.	619
Cómo trabajar con grupos.	620
Grupo de entrada.	620
Grupos de salida.	620
Utilizar las condiciones del filtro de grupo.	620
Puertos dinámicos en condiciones de filtro de grupo.	623
Parametrizar el filtro de grupo.	623
Cómo añadir grupos.	623
Cómo trabajar con puertos.	624
Cómo conectar transformaciones de enrutador de una asignación.	624
Propiedades avanzadas de la transformación de enrutador.	625
Transformación de enrutador en un entorno no nativo.	625
Capítulo 41: Transformación de generador de secuencia.	626
Introducción a la transformación de generador de secuencia.	626
Puertos de generador de secuencia.	626
Puertos de transferencia.	627
Puerto NEXTVAL.	627

Propiedades de transformación de generador de secuencia.	629
Valor inicial.	629
Valor final.	629
Valor de incremento.	630
Realizar ciclos a través de un intervalo de valores.	630
Propiedades avanzadas del generador de secuencia.	631
Restablecer.	631
Mantener orden de filas.	631
Objeto de datos de secuencia.	632
Crear un objeto de datos de secuencia.	632
Cómo crear una transformación de generador de secuencia.	635
Preguntas frecuentes.	636
Transformación de generador de secuencia en un entorno no nativo.	637
Transformación de generador de secuencia en el motor de Blaze.	637
Transformación de generador de secuencia en el motor de Spark.	637
Capítulo 42: Transformación de ordenación.	638
Resumen de la transformación de ordenación.	638
Transformaciones de ordenación en asignaciones dinámicas.	639
Cómo desarrollar una transformación de incorporación.	639
Puertos de transformación de ordenación.	640
Ficha Ordenar.	640
Configurar claves de ordenación.	640
Parametrizar las claves de ordenación.	641
Propiedades avanzadas de la transformación de ordenación.	643
Memoria caché de ordenación.	644
Optimizar la memoria caché de ordenación.	644
Cómo crear una transformación de ordenación.	644
Cómo crear una transformación de ordenación reutilizable.	645
Cómo crear una transformación de ordenación no reutilizable.	645
Ejemplo de transformación de ordenación.	645
Transformación de ordenación en un entorno no nativo.	647
Transformación de ordenación en el motor de Blaze.	647
Transformación de ordenación en el motor de Spark.	648
Transformación de ordenación en el motor de Databricks Spark.	648
Capítulo 43: Transformación de SQL.	650
Resumen de la transformación de SQL.	650
Puertos de transformación de SQL.	651
Puertos de entrada.	652
Puertos de salida.	652
Puertos de transferencia.	653
Puerto SQLException.	654

Número de filas afectadas.	655
Propiedades avanzadas de Transformación de SQL.	655
Consulta de transformación de SQL.	657
Definir la consulta SQL.	658
Cardinalidad entre fila de entrada y fila de salida.	660
Cómo procesar instrucciones de consultas.	660
Configuración de puertos.	661
Recuento máximo de filas de salida.	661
Filas de error.	661
Continuar ante error de SQL.	663
Optimización del filtro con la transformación de SQL.	663
Optimización de primera selección con la transformación de SQL.	664
Optimización de inserción con la transformación de SQL.	664
Ejemplo de transformación de SQL con una consulta SQL.	665
Asignación de objeto de datos lógicos.	665
Tabla Salario.	666
Tabla Empleado	666
Ejemplo de transformación de SQL.	666
Salida.	668
Procedimientos almacenados.	669
Puertos de transformación de SQL para procedimientos almacenados.	669
Conjuntos de resultados de un procedimiento almacenado.	671
Ejemplo de procedimiento almacenado.	673
Conexión de transformación de SQL.	674
Creación de un parámetro de nombre de conexión.	674
Cómo crear manualmente una transformación de SQL.	675
Creación de una transformación de SQL desde un procedimiento almacenado.	676
Capítulo 44: Transformación de estandarizador.	678
Resumen de la transformación de estandarizador.	678
Estrategias de estandarización.	678
Propiedades de estandarización.	679
Cómo configurar una estrategia de estandarización.	680
Propiedades avanzadas de Transformación de estandarizador.	680
Transformación de estandarizador en un entorno no nativo.	681
Capítulo 45: Transformación de unión.	682
Resumen de la transformación de unión.	682
Grupos y puertos.	683
Propiedades avanzadas de la transformación de unión.	683
Cómo procesar una transformación de unión.	684
Cómo crear una transformación de unión.	684
Cómo crear una transformación de unión reutilizable.	684

Cómo crear una transformación de unión no reutilizable.	685
Transformación de unión en un entorno no nativo.	685
Transformación de unión en una asignación de transmisión.	685
Transformación de unión en el motor de Databricks Spark.	685
Capítulo 46: Transformación de estrategia de actualización.	686
Resumen de la transformación de estrategia de actualización.	686
Cómo configurar la estrategia de actualización.	687
Actualizar transformaciones de estrategia en asignaciones dinámicas.	687
Cómo marcar filas dentro de una asignación.	687
Expresiones de estrategia de actualización.	688
Propiedades avanzadas de la transformación de estrategia de actualización.	688
Transformaciones de estrategia de agregación y actualización.	689
Cómo especificar opciones de actualización para destinos individuales.	689
Transformación de estrategia de actualización en un entorno no nativo.	690
Transformación de estrategia de actualización en el motor de Blaze.	690
Transformación de estrategia de actualización en el motor de Spark.	691
Transformación de estrategia de actualización en el motor de Databricks Spark.	693
Capítulo 47: Transformación de consumidor de servicio web.	695
Resumen de la transformación de consumidor de servicio web.	695
Mensajes SOAP.	696
Archivos WSDL.	696
Operaciones.	697
Seguridad de servicio web.	697
Selección WSDL.	698
Puertos de transformación de consumidor de servicio web.	699
Puertos de entrada de encabezado HTTP.	700
Otros puertos de entrada.	700
Asignación de entrada de transformación de consumidor de servicio web.	701
Reglas y directrices para asignar puertos de entrada a nodos.	702
Personalizar opciones de vista.	702
Cómo asignar puertos de entrada a la entrada de operación.	702
Asignación de salida de transformación de consumidor de servicio web.	704
Reglas y directrices para asignar nodos a puertos de salida.	705
Cómo asignar el mensaje SOAP como XML.	705
Personalizar opciones de vista.	705
Cómo asignar la salida de operación a puertos de salida.	706
Propiedades avanzadas de la transformación de consumidor de servicio web.	707
Control de errores de servicio web.	710
Compresión de mensajes	710
Simultaneidad.	711
Optimizaciones de filtro.	712

Habilitar la optimización de primera selección con la transformación del consumidor de servicio web.	712
Optimización de inserción con la transformación del consumidor de servicio web.	712
Crear una transformación de consumidor de servicio web.	714
Ejemplo de transformación de consumidor de servicio web.	716
Archivo de entrada.	716
Modelo de objeto de datos lógicos.	716
Asignación de objetos de datos lógicos.	717
Transformación de consumidor de servicio web.	717

Capítulo 48: Análisis de mensajes SOAP del servicio web. 720

Resumen del análisis de mensajes SOAP del servicio web.	720
Interfaz de usuario de transformación.	721
Configuración de salidas de ocurrencia múltiple.	722
Salida relacional normalizada.	722
Claves generadas.	722
Salida relacional desnormalizada.	723
Salida relacional pivotada.	724
Cómo analizar elementos anyType.	724
Cómo analizar tipos derivados.	725
Análisis de elementos QName.	726
Cómo analizar grupos de sustitución.	727
Cómo analizar construcciones XML en mensajes SOAP.	727
Elemento de selección.	727
Elemento de lista.	727
Elemento de unión.	728

Capítulo 49: Generación de mensajes SOAP del servicio web. 729

Resumen de la generación de mensajes SOAP del servicio web.	729
Interfaz de usuario de transformación.	730
Área Puertos de entrada.	730
Área Operación.	731
Relaciones entre puerto y nivel de jerarquía.	731
Claves.	732
Asignar puertos.	734
Asignación de un puerto.	735
Asignación de un grupo.	735
Asignación de varios puertos.	736
Cómo pivotar puertos de ocurrencia múltiple.	736
Asignación de datos desnormalizados.	737
Tipos derivados y sustitución de elementos.	739
Cómo generar tipos derivados.	739
Cómo generar elementos anyType y atributos.	739

Cómo generar grupos de sustitución.	740
Cómo generar construcciones XML en mensajes SOAP.	740
Elemento de selección.	740
Elemento de lista.	741
Elemento de unión.	742
Capítulo 50: Transformación de media ponderada.	743
Resumen de la transformación de media ponderada.	743
Cómo configurar una transformación de media ponderada.	743
Ejemplo de puntuaciones de coincidencia ponderadas.	744
Propiedades avanzadas de Transformación de media ponderada.	744
Transformación de media ponderada en un entorno no nativo.	745
Capítulo 51: Transformación de ventana.	746
Capítulo 52: Transformación de escritura.	747
Resumen de la transformación de escritura.	747
Propiedades de la transformación de escritura	747
Propiedades generales.	748
Propiedades del objeto de datos.	749
Propiedades de puertos.	749
Propiedades de tiempo de ejecución.	750
Propiedades de la vinculación en tiempo de ejecución.	751
Propiedades avanzadas.	751
Crear una transformación de escritura.	754
Crear una transformación de escritura a partir de un objeto de datos.	754
Crear una transformación de escritura a partir de un flujo de asignación.	755
Crear una transformación de escritura a partir de un parámetro.	755
Crear una transformación de escritura a partir de una transformación existente	756
Apéndice A: Delimitadores de transformación.	758
Resumen de delimitadores de transformación.	758
Índice.	759

Prefacio

Lea la *Guía de transformaciones de Informatica® Developer* para conocer la configuración, las directrices, el uso y el comportamiento en tiempo de ejecución de las transformaciones de Informatica. Aprenda a usar las transformaciones con casos de uso adecuados en función del entorno y el motor de tiempo de ejecución. Vea la compatibilidad de cada transformación en función de dónde se ejecute la asignación y cómo tenga pensado ejecutarla.

Recursos de Informatica

Informatica proporciona una variedad de recursos de productos a través de Informatica Network y otros portales en línea. Use los recursos para sacar el mayor provecho de los productos y las soluciones de Informatica y aprender de otros expertos en la materia y usuarios de Informatica.

Informatica Network

Informatica Network es la puerta de entrada a muchos recursos, entre ellos, la base de conocimientos de Informatica y el servicio internacional de atención al cliente de Informatica. Para entrar en Informatica Network, visite <https://network.informatica.com>.

Como miembro de Informatica Network, tiene las siguientes opciones:

- Buscar recursos de productos en la base de conocimientos
- Vea la información de disponibilidad del producto.
- Crear y revisar casos de soporte
- Buscar su red de grupos de usuarios de Informatica locales y colaborar con sus iguales.

Base de conocimiento de Informatica

Use la base de conocimientos de Informatica para encontrar recursos de productos como artículos prácticos, procedimientos recomendados, tutoriales de video y respuestas a preguntas frecuentes.

Para buscar en la base de conocimiento, visite <https://search.informatica.com>. Si tiene preguntas, comentarios o ideas relacionadas con la base de conocimiento de Informatica, póngase en contacto con el equipo de la base de conocimiento de Informatica en KB_Feedback@informatica.com.

Documentación de Informatica

Use el portal de documentación de Informatica para recorrer una extensa biblioteca de documentación para las versiones de productos actuales y recientes. Para recorrer el portal de documentación, visite <https://docs.informatica.com>.

Si tiene preguntas, comentarios o ideas acerca de la documentación de los productos, póngase en contacto con el equipo de la documentación de Informatica en infa_documentation@informatica.com.

Matrices de disponibilidad de producto de Informatica

Las matrices de disponibilidad de producto (PAM, Product Availability Matrixes) indican las versiones de sistemas operativos, bases de datos y otros tipos de orígenes y destinos de datos admitidos por la versión de un producto. Puede recorrer las PAM de Informatica en <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity es una colección de consejos y procedimientos recomendados desarrollados por los servicios profesionales de Informatica que se basan en experiencias reales de cientos de proyectos de administración de datos. Informatica Velocity representa el conocimiento colectivo de los consultores de Informatica que trabajan con organizaciones de todo el mundo para planificar, desarrollar, implementar y dar mantenimiento a soluciones de administración de datos exitosas.

Puede encontrar recursos de Informatica Velocity en <http://velocity.informatica.com>. Si tiene alguna pregunta, comentario o idea acerca de Informatica Velocity, póngase en contacto con los servicios profesionales de Informatica en ips@informatica.com.

Catálogo de soluciones de Informatica

El catálogo de soluciones de Informatica es un foro donde puede buscar soluciones que aumenten, amplíen o mejoren sus implementaciones de Informatica. Aproveche cualquiera de los cientos de soluciones de socios y desarrolladores de Informatica que se encuentran en el catálogo para mejorar su productividad y acelerar la implementación de los proyectos. Puede encontrar el catálogo de soluciones de Informatica en <https://marketplace.informatica.com>.

Servicio internacional de atención al cliente de Informatica

Puede ponerse en contacto con un centro de atención global por teléfono o a través de Informatica Network.

Para encontrar el número de teléfono local del servicio internacional de atención al cliente de Informatica, visite el sitio web de Informatica en el siguiente vínculo:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

Para buscar recursos de asistencia en línea en Informatica Network, visite <https://network.informatica.com> y seleccione la opción eSupport.

CAPÍTULO 1

Introducción a las transformaciones

Este capítulo incluye los siguientes temas:

- [Introducción al resumen de las transformaciones, 32](#)
- [Transformaciones en los entornos nativo y no nativo, 36](#)
- [Administración de los tipos de datos de transformación, 39](#)
- [Cómo desarrollar una transformación, 41](#)
- [Transformaciones de varios grupos, 42](#)
- [Expresiones en transformaciones, 43](#)
- [Variables locales, 47](#)
- [Valores predeterminados para los puertos, 50](#)
- [Niveles de seguimiento, 57](#)
- [Transformaciones reutilizables, 58](#)
- [Transformaciones no reutilizables, 59](#)
- [Cómo crear una transformación, 60](#)

Introducción al resumen de las transformaciones

Una transformación es un objeto que genera, modifica o pasa datos.

Informatica Developer proporciona un conjunto de transformaciones que realizan funciones específicas. Por ejemplo, una transformación de agregación realiza cálculos en grupos de datos.

Las transformaciones de una asignación representan las operaciones que realiza el servicio de integración de datos en los datos. Los datos pasan por los puertos de transformación que se enlazan en una asignación o un mapplet.

Las transformaciones pueden ser activas o pasivas. Las transformaciones pueden conectarse al flujo de datos o pueden estar desconectadas.

Transformaciones activas

Una transformación activa cambia el número de filas que pasan por una transformación. También puede cambiar el tipo de fila.

Por ejemplo, la transformación de filtro es activa porque quita filas que no cumplen la condición del filtro. La transformación de estrategia de actualización es activa porque marca filas para su inserción, eliminación, actualización o rechazo.

No se pueden conectar varias transformaciones activas o una transformación activa y otra pasiva a la misma transformación o grupo de entrada de transformación de nivel inferior porque es posible que el servicio de integración de datos no pueda concatenar las filas pasadas por las transformaciones activas.

Por ejemplo, una ramificación de una asignación contiene una transformación de estrategia de actualización que marca una fila para su eliminación. Otra ramificación contiene una transformación de estrategia de actualización que marca una fila para su inserción. Si conecta estas transformaciones a un grupo de entrada de transformación individual, el servicio de integración de datos no puede combinar las operaciones de eliminación y de inserción para la fila.

Transformaciones pasivas

Una transformación pasiva no cambia el número de filas que pasan por la transformación, mantiene el límite de transacciones y mantiene el tipo de fila.

Puede conectar varias transformaciones a la misma transformación de nivel inferior o al mismo grupo de entrada de transformación si todas las ramificaciones de un nivel superior son pasivas. La transformación que origina la ramificación puede ser activa o pasiva.

Transformaciones no conectadas

Las transformaciones pueden conectarse al flujo de datos o pueden estar desconectadas. Una transformación no conectada no está conectada a otras transformaciones en la asignación. Se llama a una transformación no conectada dentro de otra transformación y devuelve un valor a esa transformación.

Transformaciones multiestrategia

Una estrategia es el conjunto de una o más operaciones que una transformación puede realizar en los datos. Puede asignar un conjunto de puertos de entrada y de salida diferentes a cada estrategia de una transformación. La transformación almacena las estrategias que se definen en un objeto de transformación individual.

Utilice la vista **Dependencias** para ver los puertos que utiliza cada estrategia.

Puede definir varias estrategias de transformación en las siguientes transformaciones:

- Conversión de mayúsculas y minúsculas
- Clasificador
- Decisión
- Generador de claves
- Etiquetador
- Coincidencia
- Fusión
- Analizador

Para utilizar varias estrategias en una transformación del analizador, configure la transformación para que analice tokens.

- Estandarizador

Descripciones de transformaciones

Developer Tool contiene transformaciones comunes y de calidad de datos. En Data Engineering Quality hay disponibles transformaciones de calidad de datos.

En la siguiente tabla se describen cada una de las transformaciones:

Transformación	Tipo	Descripción
Validador de direcciones	Activa o pasiva	Comprueba y mejora la precisión de los registros de direcciones de correo postal y agrega información que ayuda a los usuarios a seleccionar los destinatarios y a enviar el correo.
Asociación	Activa	Crea vínculos entre los registros duplicados que una transformación de coincidencia asigna a diferentes clústeres.
Agregación	Activa	Realiza cálculos de agregado.
Excepción de registros incorrectos	Activa	Identifica los registros que pueden contener datos con errores y carga los registros en tablas que una Herramienta del analista puede revisar y actualizar.
Conversión de mayúsculas y minúsculas	Pasiva	Estandariza el uso de mayúsculas y minúsculas de las cadenas.
Clasificador	Pasiva	Escribe etiquetas que resumen la información de los campos de puertos de entrada. Se utiliza cuando los campos contienen una gran cantidad de texto.
Comparación	Pasiva	Genera puntuaciones numéricas que indican el grado de similitud entre pares de cadenas de entrada.
Consolidación	Activa	Crea un registro consolidado a partir de registros que la transformación de coincidencia ha identificado como duplicados.
Enmascaramiento de datos	Pasiva	Reemplaza los datos de producción confidenciales con datos de prueba realistas en entornos que no son de producción. Esta transformación se puede conectar o desconectar dentro del canal.
Procesador de datos	Activa	Procesa formatos de archivo sin estructura y semiestructurados en una asignación.
Decisión	Pasiva	Evalúa las condiciones de los datos de entrada y crea la salida en función de los resultados de dichas condiciones.
Excepción de registros duplicados	Activa	Identifica los registros que pueden contener información duplicada y carga los registros en tablas que una Herramienta del analista puede revisar y actualizar.
Expresión	Pasiva	Calcula un valor

Transformación	Tipo	Descripción
Filtro	Activa	Filtra datos.
De jerárquica a relacional	Activa	Procesa entradas jerárquicas y las transforma en salidas relacionales.
Java	Activa o pasiva	Ejecuta la lógica de usuario codificada en Java. El repositorio almacena el código de byte para la lógica de usuario.
Unión	Activa	Une datos de diferentes bases de datos o sistemas de archivos sin formato.
Generador de claves	Activa	Asigna registros a grupos basados en los valores de datos de una columna que seleccione.
Etiquetador	Pasiva	Escribe etiquetas que describen los caracteres o las cadenas en un campo de puerto de entrada.
Búsqueda	Activa o pasiva	Busca y devuelve datos de un archivo sin formato, un objeto de datos lógicos, una tabla de referencia, una tabla relacional, una vista o un sinónimo. Esta transformación se puede conectar o desconectar dentro del canal.
Macro	Pasiva	Habilita la función dinámica de un mapplet.
Coincidencia	Activa	Genera puntuaciones que indican el grado de similitud entre registros de entrada.
Fusión	Pasiva	Lee los valores de datos de varias columnas de entrada y crea una columna de salida.
Normalizador	Activa	Procesa filas de origen que contienen datos de ocurrencia múltiple y devuelve una fila de destino para cada instancia de los datos de ocurrencia múltiple.
Salida	Pasiva	Define filas de salida de mapplets.
Analizador	Pasiva	Analiza los valores en un puerto de entrada en diferentes puertos de salida en función de los tipos de información que contienen los valores.
Rango	Activa	Limita los registros a un intervalo superior o inferior.
Lectura	Pasiva	Lee los datos de un origen.
De relacional a jerárquica	Activa	Procesa la entrada relacional y la transforma en salida jerárquica.
Consumidor de servicio web REST	Activa	Se conecta a un servicio web REST como cliente de servicio web para acceder a datos o transformarlos.
Enrutador	Activa	Distribuye datos a varias transformaciones con base en las condiciones de grupo.
Generador de secuencia	Pasiva	Genera una secuencia de valores numéricos.
Ordenación	Activa	Ordena datos según una clave de ordenación.

Transformación	Tipo	Descripción
SQL	Activa o pasiva	Ejecuta consultas SQL en una base de datos.
Estandarizador	Pasiva	Genera versiones estandarizadas de cadenas de entrada.
Unión	Activa	Fusiona datos de diferentes bases de datos o sistemas de archivos sin formato.
Estrategia de actualización	Activa	Determina si las filas se deben insertar, eliminar, actualizar o rechazar.
Consumidor de servicio web	Activa	Se conecta a un servicio web como cliente de servicio web para acceder a datos o transformarlos.
Media ponderada	Pasiva	Lee las puntuaciones de coincidencia que genera una transformación de coincidencia para los registros en un conjunto de datos y calcula una puntuación media para cada par de registros. Puede aplicar distintos niveles de importancia a las puntuaciones que la transformación genera para cada par de registros.
Escritura	Pasiva	Representa un destino en el que la asignación escribe datos.

Transformaciones en los entornos nativo y no nativo

Las asignaciones que se ejecutan en el entorno no nativo pueden devolver resultados distintos de los de las asignaciones que se ejecutan en el entorno nativo.

Considere las siguientes diferencias de procesamiento:

- El entorno no nativo usa el procesamiento distribuido y procesa datos en diferentes nodos. Los nodos carecen de acceso a los datos que se están procesando en otros nodos. Como resultado, el motor en tiempo de ejecución tal vez no pueda determinar el orden en el que se originaron los datos. Por lo tanto, al ejecutar una asignación en un entorno no nativo y al ejecutar la misma asignación en el entorno nativo, ambas asignaciones devuelven resultados correctos, pero puede que no sean idénticos.
- Cada uno de los motores en tiempo de ejecución del entorno no nativo puede procesar la lógica de asignación de forma diferente. En el entorno no nativo, las transformaciones de Informatica podrían no admitirse por completo, con restricciones o no admitirse de ningún modo. De igual manera, en el entorno nativo, es posible que algunas transformaciones de Informatica y comportamientos de transformación no se admitan.

En la siguiente tabla se enumeran las transformaciones y la compatibilidad con diferentes motores en un entorno no nativo:

Transformación	Motores compatibles
<i>Las transformaciones que no se enumeran en esta tabla solo se admiten en el entorno nativo.</i>	
Validador de direcciones	- Blaze - Spark* - Databricks Spark
Agregación	- Blaze - Spark* - Databricks Spark
Conversión de mayúsculas y minúsculas	- Blaze - Spark** - Databricks Spark
Clasificador	- Blaze - Spark* - Databricks Spark
Comparación	- Blaze - Spark** - Databricks Spark
Consolidación	- Blaze - Spark** - Databricks Spark
Enmascaramiento de datos	- Blaze - Spark*
Procesador de datos	- Blaze - Spark**
Decisión	- Blaze - Spark** - Databricks Spark
Expresión	- Blaze - Spark* - Databricks Spark
Filtro	- Blaze - Spark* - Databricks Spark
Java	- Blaze - Spark*
Combinación	- Blaze - Spark* - Databricks Spark
Generador de claves	- Blaze - Spark** - Databricks Spark

Transformación	Motores compatibles
Etiquetador	- Blaze - Spark** - Databricks Spark
Búsqueda	- Blaze - Spark* - Databricks Spark
Macro	- Blaze - Spark* - Databricks Spark
Coincidencia	- Blaze - Spark** - Databricks Spark
Fusión	- Blaze - Spark** - Databricks Spark
Normalizador	- Blaze - Spark* - Databricks Spark
Analizador	- Blaze - Spark* - Databricks Spark
Python	- Spark* - Databricks Spark
Rango	- Blaze - Spark* - Databricks Spark
Enrutador	- Blaze - Spark* - Databricks Spark
Especificación de regla	- Databricks Spark
Generador de secuencia	- Blaze - Spark
Ordenación	- Blaze - Spark* - Databricks Spark
Estandarizador	- Blaze - Spark* - Databricks Spark
Unión	- Blaze - Spark* - Databricks Spark

Transformación	Motores compatibles
Estrategia de actualización	- Blaze - Spark - Databricks Spark
Media ponderada	- Blaze - Spark** - Databricks Spark
Ventana	- Spark***
<p>* Compatible tanto para asignaciones por lotes como de transmisión.</p> <p>** Compatible con restricciones en las asignaciones por lotes. No compatible en asignaciones de transmisión. Para obtener información sobre la compatibilidad de la transformación del procesador de datos en el motor de Spark, vea el KB article.</p> <p>*** Compatible únicamente en asignaciones de transmisión. Para obtener más información, consulte la Guía del usuario de Data Engineering Streaming.</p>	

Administración de los tipos de datos de transformación

Las transformaciones pueden procesar funciones específicas de tipos de datos o permitir transferencias de datos sin procesar los datos. El Servicio de integración de datos procesa algunos tipos de datos (como Decimal, Fecha y hora con zona horaria y Marca de tiempo con zona horaria local) en función de las transformaciones.

Tipo de datos Decimal

Puede utilizar el tipo de datos Decimal para leer y escribir datos en archivos sin formatos y en las bases de datos admitidas, como Oracle, Microsoft SQL Server, IBM DB2 y ODBC.

En las transformaciones que admiten una precisión de hasta 38 dígitos, la precisión es de 1 a 38 dígitos y, la escala, de 0 a 38.

Transformaciones que admiten el tipo de datos Decimal

Las siguientes transformaciones admiten el tipo de datos Decimal con una precisión de hasta 38 dígitos y pueden realizar cálculos en los datos:

- Agregación
- Enmascaramiento de datos
- Expresión
- Filtro
- Java
- Combinación
- Búsqueda

- Normalizador
- Rango
- Enrutador
- Generador de secuencia
- Ordenación
- Unión
- Estrategia de actualización

Transformaciones con compatibilidad de transferencia para el tipo de datos Decimal

Algunas transformaciones solo pueden pasar datos decimales con una precisión de hasta 38 dígitos a través de la transformación. La transformación no puede realizar ningún cálculo en los datos. Cuando utilice un tipo de datos Decimal con una precisión de hasta 38 dígitos para realizar cálculos en las transformaciones, el servicio de integración de datos de procesa el tipo de datos como Doble.

Las siguientes transformaciones admiten transferencias en el tipo de datos Decimal con una precisión de hasta 38 dígitos:

- Procesador de datos
- De jerárquica a relacional
- Macro
- Transformación de consumidor de servicio web REST
- SQL
- Consumidor de servicios web

Transformaciones que no admiten el tipo de datos Decimal

Algunas transformaciones no admiten el tipo de datos Decimal, como las transformaciones de Data Quality.

La siguiente lista de transformaciones de Data Quality no admiten el tipo de datos Decimal con una precisión de hasta 38 dígitos:

- Validador de direcciones
- Asociación
- Conversión de mayúsculas y minúsculas
- Clasificador
- Comparación
- Consolidación
- Decisión
- Generador de claves
- Etiquetador
- Coincidencia
- Fusión
- Analizador
- Estandarizador
- Media ponderada

En las transformaciones que no admiten el tipo de datos Decimal 38, cuando un tipo de datos Decimal tiene una precisión superior a 28 dígitos, el servicio de integración de datos convierte los valores decimal en doble en el modo de alta precisión.

Marca de tiempo con zona horaria

Marca de tiempo con zona horaria es una variante del tipo de datos Marca de tiempo que incluye una variación o región de zona horaria.

Las siguientes transformaciones admiten el tipo de datos Marca de tiempo con zona horaria:

- Agregación
- Expresión
- Filtro
- Java
- Combinación
- Búsqueda
- Normalizador
- Rango
- Enrutador
- Generador de secuencia
- Ordenación
- Unión
- Estrategia de actualización

La compatibilidad con transferencia significa que se pueden pasar datos a través de las transformaciones, pero no se pueden realizar funciones en el tipo de datos Marca de tiempo con zona horaria.

Las siguientes transformaciones admiten la transferencia en el tipo de datos Marca de tiempo con zona horaria:

- Enmascaramiento de datos
- Procesador de datos
- De jerárquica a relacional
- Macro
- SQL

Marca de tiempo con zona horaria local

Marca de tiempo con zona horaria local está implícitamente admitido por las transformaciones, ya que la funcionalidad es equivalente a marca de tiempo.

Cómo desarrollar una transformación

Cuando se crea una asignación, puede añadir transformaciones y configurarlas para manejar datos según un propósito empresarial.

Lleve a cabo las siguientes tareas para desarrollar una transformación e incorporarla a una asignación:

1. Añada una transformación no reutilizable a una asignación o mapplet. También puede crear una transformación reutilizable que puede añadir a varias asignaciones o mapplets.

2. Configure la transformación. Cada tipo de transformación tiene un conjunto de opciones único que puede configurarse.
3. Si la transformación es reutilizable, añádala a la asignación o al mapplet.
4. Vincule la transformación a otros objetos en la asignación o mapplet.

Puede arrastrar puertos desde objetos de nivel superior a los puertos de entrada de transformación.
Puede arrastrar puertos de salida desde la transformación a los puertos en objetos de nivel inferior.
Algunas transformaciones utilizan puertos predefinidos que puede seleccionar.

Nota: Si crea una transformación reutilizable, debe añadir los puertos de entrada y salida necesarios antes de enlazar la transformación a otros objetos. No puede añadir puertos a la instancia de transformación en el mapplet o en el lienzo de la asignación. Para actualizar los puertos en una transformación reutilizable, abra el objeto de transformación desde el proyecto del repositorio y añada los puertos.

Transformaciones de varios grupos

Una transformación puede tener varios grupos de entrada y salida. Un grupo es un conjunto de puertos que definen una fila de datos de entrantes o salientes.

Un grupo es equivalente a una tabla en un origen relacional o una definición de destino. La mayoría de las transformaciones tienen un grupo de entrada y un grupo de salida. Sin embargo, algunas tienen varios grupos de entrada, varios grupos de salida, o ambos. Un grupo es la representación de una fila de datos que entran o salen de una transformación.

Todas las transformaciones de varios grupos son transformaciones activas. No se pueden conectar varias transformaciones activas o una transformación activa y una transformación pasiva a la misma transformación de nivel inferior o grupo de entrada de transformación.

Algunas transformaciones de varios grupos de entrada requieren que el Servicio de integración bloquee los datos en un grupo de entrada mientras el Servicio de integración espera una fila de un grupo de entrada diferente. Una transformación de bloqueo es una transformación de varios grupos de entrada que bloquea los datos entrantes. Las siguientes transformaciones son transformaciones de bloqueo:

- Transformación personalizada con la propiedad Las entradas deben bloquear habilitada
- Transformación de combinación configurada para la entrada no ordenada

Cuando guarde o valide una asignación, algunas asignaciones que contienen transformaciones activas o de bloqueo podrían no ser válidas.

Reglas y directrices para las transformaciones de varios grupos

Cuando conecte transformaciones en una asignación, deberá tener en cuenta ciertas reglas y directrices para conectar transformaciones de varios grupos.

Tenga en cuenta las siguientes reglas y directrices para las transformaciones de varios grupos:

- Puede conectar un grupo a una transformación o destino.
- Puede conectar uno o más puertos de salida de un grupo a varias transformaciones o destinos.
- No puede conectar campos de varios grupos de salida de una transformación con el mismo grupo de entrada de otra transformación.

- No puede conectar campos de varios grupos de salida de diferentes transformaciones con el mismo grupo de entrada de otra transformación a menos que cada transformación entre el origen y la transformación sean transformaciones pasivas.
- No puede conectar campos de varios grupos de salida de una transformación con el mismo grupo de entrada de otra transformación, a menos que la otra transformación sea una transformación de bloqueo.
- No puede conectar un campo de salida con varios campos de entrada en el mismo grupo de entrada, a menos que el grupo esté en una transformación de normalizador.

Expresiones en transformaciones

En algunas expresiones puede especificar expresiones en el **Editor de expresiones**. Las expresiones modifican datos o comprueban si los datos cumplen ciertas condiciones.

Cree expresiones que utilicen funciones del lenguaje de transformación. Las funciones del lenguaje de transformación son funciones, similares a las de SQL, que transforman datos.

Especifique una expresión en un puerto que utilice el valor de los datos de un puerto de entrada o de entrada/salida. Por ejemplo, tiene una transformación con un puerto de entrada IN_SALARY que contiene los salarios de todos los empleados. Puede utilizar los valores de la columna IN_SALARY más adelante en la asignación. Al mismo tiempo, también puede usar la transformación para calcular el salario medio y el total de los salarios. Developer Tool requiere que cree un puerto de salida independiente para cada valor calculado.

La siguiente tabla enumera las transformaciones en las que se pueden especificar expresiones:

Transformación	Expresión	Valor de devolución
Agregación	Realiza un cálculo de agregación basado en todos los datos pasados por la transformación. También puede especificar un filtro para los registros en el cálculo de agregación para excluir ciertos tipos de registros. Por ejemplo, con esta transformación se puede buscar el número total y el salario promedio de todos los empleados de una sucursal.	Resultado de un cálculo de agregación de un puerto.
Expresión	Realiza un cálculo basado en los valores de una fila individual. Por ejemplo, según el precio y la cantidad de un artículo concreto, puede calcular el precio total de compra de ese artículo en un pedido.	Resultado de un cálculo de nivel de fila de un puerto.
Filtro	Especifica una condición que se usa para filtrar las filas que se pasan por esta transformación. Por ejemplo, si desea escribir datos de clientes en la tabla BAD_DEBT para los clientes con saldos pendientes de pago, puede utilizar la transformación de filtro para filtrar los datos de clientes.	TRUE o FALSE, en función de si una fila cumple la condición especificada. El Servicio de integración de datos pasa filas que devuelven TRUE mediante esta transformación. La transformación aplica este valor a cada fila que pasa por ella.

Transformación	Expresión	Valor de devolución
Combinación	Especifica una condición avanzada que se usa para combinar datos de origen sin ordenar. Por ejemplo, puede concatenar los puertos principales de nombre y apellido y, a continuación, hacerlos coincidir con el puerto detallado de nombre completo.	TRUE o FALSE según si la fila cumple la condición especificada. En función del tipo de combinación seleccionado, el Servicio de integración de datos añade la fila al conjunto de resultados o la descarta.
Rango	Define las condiciones de las filas incluidas en un rango. Por ejemplo, puede averiguar quiénes son los 10 mejores vendedores de la organización.	Resultado de una condición o de un cálculo de un puerto.
Enrutador	Distribuye datos por varias transformaciones según una expresión de grupo. Utilice esta transformación, por ejemplo, para comparar los salarios de los empleados a tres niveles de remuneración diferentes. Puede hacer esto creando tres grupos en la transformación de enrutador. Por ejemplo, cree una expresión de grupo para cada rango salarial.	TRUE o FALSE, en función de si una fila cumple la expresión de grupo especificada. El Servicio de integración de datos pasa filas que devuelven TRUE mediante cada grupo definido por el usuario en esta transformación. Las filas que devuelven FALSE pasan por el grupo predeterminado.
Estrategia de actualización	Marca una fila para su actualización, inserción, eliminación o rechazo. Utilice esta transformación para controlar las actualizaciones de un destino según las condiciones que se apliquen. Por ejemplo, puede usar la transformación de estrategia de actualización para marcar todas las filas de cliente para que se actualicen después de cambiar la dirección de envío de correo. También puede marcar para rechazar todas las filas de empleados que ya no trabajan para la organización.	Código numérico de actualización, inserción, eliminación o rechazo. La transformación aplica este valor a cada fila que pasa por ella.

El editor de expresiones

Utilice el **Editor de expresiones** para crear instrucciones al estilo SQL.

Puede introducir una expresión manualmente o utilizar el método de apuntar y hacer clic. Seleccione funciones, puertos, variables y operadores desde la interfaz de apuntar y hacer clic para reducir al mínimo el número de errores en la construcción de expresiones. El número máximo de caracteres que se pueden incluir en una expresión es de 32.767.

Nombres de puerto en una expresión

Se pueden introducir nombres de puerto de transformación en una expresión.

En el caso de transformaciones conectadas, si se utilizan nombres de puerto en una expresión, Developer Tool actualiza esa expresión cuando se cambian los nombres de puerto en la transformación. Por ejemplo, se escribe una expresión que determina la diferencia entre dos fechas: Date_Promised y Date_Delivered. Si se cambia el nombre de puerto Date_Promised a Due_Date, Developer Tool cambia el nombre de puerto Date_Promised a Due_Date en la expresión.

Nota: Se puede propagar el nombre Due_Date a otras transformaciones no reutilizables que dependen de este puerto en la asignación.

Cómo añadir una expresión a un puerto

Se puede añadir una expresión a un puerto de salida.

1. En la transformación, seleccione el puerto y abra el **Editor de expresiones**.
2. Introduzca la expresión. Utilice las fichas Funciones y Puertos y las claves del operador.
Nota: No se puede utilizar un carácter de escape en una expresión. Si incluye un carácter de escape en la expresión, Developer tool podría mostrar un error de análisis.
3. Si lo desea, añada comentarios a la expresión.
Utilice los indicadores de comentario -- o //.
4. Haga clic en el botón Validar para validar la expresión.
5. Haga clic en **Aceptar**.
6. Si la expresión no es válida, corrija los errores de validación y valide nuevamente la expresión.
7. Cuando la expresión sea válida, haga clic en **Aceptar** para cerrar el **Editor de expresiones**.

Comentarios en una expresión

Se pueden añadir comentarios a una expresión para describirla o para especificar una URL válida para acceder a documentación empresarial sobre la expresión.

Para añadir comentarios dentro de la expresión, utilice los indicadores de comentario -- o //.

Validación de expresiones

Es necesario validar una expresión para ejecutar una asignación u obtener una vista previa de una salida de mapplet.

Utilice el botón Validar del **Editor de expresiones** para validar una expresión. Si no se valida una expresión, Developer Tool la valida cuando se cierra el **Editor de expresiones**. Si la expresión no es válida, Developer Tool muestra una advertencia. Puede guardar la expresión no válida o modificarla.

Probar expresiones

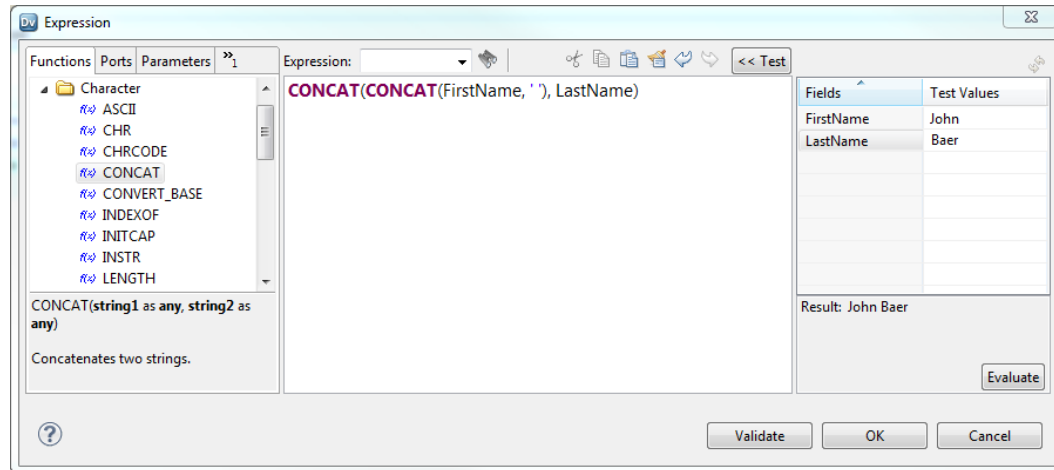
Puede probar algunas expresiones que tenga configuradas en el editor de expresiones. Para probar una expresión, hay que introducir datos de ejemplo y, a continuación, evaluar la expresión.

Al configurar expresiones, las puede probar de las siguientes formas:

- En un puerto de salida o de variable en la transformación de expresión
- En la vista Salidas de asignación de una transformación de expresión, después de añadir la transformación a una asignación

Por ejemplo, después de configurar una expresión que concatena el nombre, un espacio y el apellido, puede introducir datos de ejemplo de los puertos y, a continuación, evaluar la expresión para comprobar el resultado.

La siguiente imagen muestra los resultados de una expresión que concatena un nombre y un apellido de ejemplo:



Cadenas de formato de fecha para datos de ejemplo

Al probar una expresión que utiliza un puerto con el tipo de datos Fecha/Hora o Marca de tiempo con zona horaria, debe introducir datos de ejemplo para el puerto utilizando la cadena de formato de fecha correspondiente.

Para introducir datos de ejemplo para un puerto con el tipo de datos Fecha/Hora, utilice el formato MM/DD/YYYY HH24:MI:SS. Cuando la expresión se evalúe, el editor de expresiones mostrará el resultado con el formato que especifique en la expresión. Si omite la cadena de formato en la expresión, el editor de expresiones mostrará el resultado con el mismo formato, MM/DD/YYYY HH24:MI:SS.

Para introducir datos de ejemplo para un puerto con el tipo de datos Marca de tiempo con zona horaria, utilice el formato MM/DD/YYYY HH24:MI:SS TZR. Cuando la expresión se evalúe, el editor de expresiones mostrará el resultado con el formato YYYY-MM-DD HH24:MI:SS.NS TZR.

Conversión de tipo de datos

Varias funciones de agregación y expresión pueden generar datos pertenecientes a un tipo de datos distinto al de los datos de entrada.

Por ejemplo, cuando se multiplican dos números decimales con una precisión de 18 dígitos, el tipo de datos resultante podría ser un decimal con una precisión de 28 dígitos.

Para un tipo de datos de entrada Decimal con una precisión de 38 dígitos, el resultado de determinadas operaciones podría producir datos que no encajen en el tipo de datos resultante. Esto provocaría que el usuario obtenga una excepción por desbordamiento.

Las siguientes funciones pueden requerir la conversión del tipo de datos para dar cabida a un aumento del tamaño de los datos en comparación con los tipos de datos de entrada:

- avg
- cume
- divide
- median
- movingavg
- movingsum

- multiply
- Percentile
- Sum

Por ejemplo, si los datos de entrada son del tipo de datos entero y usa una operación de multiplicación, los datos resultantes podrían ser del tipo de datos bigint. De igual modo, cuando el tipo de datos de entrada es Decimal con una precisión de 18 dígitos, el resultado de la operación de multiplicación podría ser grande y dentro del tipo de datos Decimal con una precisión de 28 dígitos.

Variables locales

Utilice variables locales en transformaciones de agregación, expresión y rango para mejorar el rendimiento. Puede hacer referencia a las variables en una expresión o usarlas para almacenar los datos temporalmente.

Puede utilizar variables para completar las siguientes tareas:

- Almacenar datos temporalmente.
- Simplificar expresiones complejas.
- Almacenar valores de filas anteriores.
- Capturar varios valores de retorno de un procedimiento almacenado.
- Comparar valores.
- Almacenar los resultados de una transformación de búsqueda no conectada.

Almacenar datos temporalmente y simplificar expresiones complejas

Las variables aumentan el rendimiento cuando se especifican varias expresiones relacionadas en la misma transformación. Puede definir componentes como variables en lugar de analizar y validar los mismos componentes de expresiones varias veces en la transformación.

Por ejemplo, si una transformación de agregación utiliza la misma condición de filtro antes de calcular la suma y los promedios, podrá definir esta condición como una variable y después reutilizarla en los dos cálculos de agregado.

Puede simplificar expresiones complejas. Si una agregación incluye el mismo cálculo en varias expresiones, puede aumentar el rendimiento creando una variable para almacenar los resultados del cálculo.

Por ejemplo, podría crear las siguientes expresiones para encontrar tanto el salario medio como el salario total con los mismos datos:

```
AVG( SALARY, ( ( JOB_STATUS = 'Full-time' ) AND ( OFFICE_ID = 1000 ) ) )
SUM( SALARY, ( ( JOB_STATUS = 'Full-time' ) AND ( OFFICE_ID = 1000 ) ) )
```

En lugar de introducir los mismos argumentos para ambos cálculos, puede crear un puerto variable para cada condición de este cálculo y después cambiar la expresión para que use las variables.

La siguiente tabla muestra cómo utilizar variables para simplificar expresiones complejas y almacenar los datos temporalmente:

Puerto	Valor
V_CONDITION1	JOB_STATUS = 'Full-time'
V_CONDITION2	OFFICE_ID = 1000
AVG_SALARY	AVG(SALARY, (V_CONDITION1 AND V_CONDITION2))
SUM_SALARY	SUM(SALARY, (V_CONDITION1 AND V_CONDITION2))

Almacenar valores en las filas

Puede configurar las variables en las transformaciones para almacenar datos de las filas de origen. Puede utilizar las variables en expresiones de transformación.

Por ejemplo, un archivo de origen contiene las siguientes filas:

```
California
California
California
Hawaii
Hawaii
New Mexico
New Mexico
New Mexico
```

Cada fila contiene un estado. Deberá contar el número de filas y devolver el recuento de filas para cada estado:

```
California,3
Hawaii ,2
New Mexico,3
```

Puede configurar una transformación de agregación para agrupar las filas de origen por estado y contar el número de filas de cada grupo. Configure una variable en la transformación de agregación para almacenar el recuento de filas. Defina otra variable para almacenar el nombre de estado de la fila anterior.

La transformación de agregación tiene los siguientes puertos:

Puerto	Tipo de puerto	Expresión	Descripción
Estado	Transferencia	n/a	El nombre de un estado. Las filas de origen se agrupan por el nombre de estado. La transformación de agregación devuelve una fila por cada estado.
State_Count	Variable	IIF (PREVIOUS_STATE = STATE, STATE_COUNT +1, 1)	El recuento de filas del estado actual. Cuando el valor de la columna State actual es el mismo que el de la columna Previous_State, el Servicio de integración incrementa State_Count. De lo contrario, restablece el State_Count en 1.

Puerto	Tipo de puerto	Expresión	Descripción
Previous_State	Variable	State	El valor de la columna State en la fila anterior. Cuando el Servicio de integración procesa una fila, mueve el valor State a Previous_State.
State_Counter	Salida	State_Count	El número de filas de la transformación de agregación procesadas para un estado. El Servicio de integración devuelve State_Counter una vez por cada estado.

Capturar valores desde procedimientos almacenados

Las variables proporcionan una forma para capturar varias columnas de valores de devolución de procedimientos almacenados.

Directrices para configurar puertos de variables

Tenga en cuenta los siguientes factores cuando configure puertos de variables en una transformación:

- **Orden de puertos.** El Servicio de integración evalúa los puertos por dependencia. El orden de los puertos en una transformación debe coincidir con el orden de evaluación: puertos de entrada, puertos de variables, puertos de salida.
- **Tipo de datos.** El tipo de datos que elija refleja el valor de devolución de la expresión que especifique.
- **Inicialización de variables.** El Servicio de integración establece valores iniciales en puertos de variables, donde puede crear los contadores.

Orden de puertos

El Servicio de integración evalúa primero los puertos de entrada. El Servicio de integración evalúa después los puertos de variables y por último los puertos de salida.

El Servicio de integración evalúa los puertos en el orden siguiente:

1. **Puertos de entrada.** El Servicio de integración evalúa todos los puertos de entrada primero porque no dependen de los otros puertos. Por tanto, puede crear puertos de entrada en cualquier orden. El Servicio de integración no ordena los puertos de entrada porque estos puertos no hacen referencia a otros puertos.
2. **Puertos de variables.** Los puertos de variables puede hacer referencia a puertos de entrada y puertos de variable, pero no a los puertos de salida. Como los puertos de variables pueden hacer referencia a los puertos de entrada, el Servicio de integración evalúa los puertos de variables después de los puertos de entrada. Las variables pueden hacer referencia a otras variables, de modo que el orden de visualización para los puertos de variables es el mismo que el orden en que el Servicio de integración evalúa cada variable.

Por ejemplo, si desea calcular el valor original de un edificio y después ajusta la depreciación del valor, puede crear el cálculo del valor original como un puerto de variables. Este puerto de variables debe aparecer antes que el puerto ajustado para la depreciación.
3. **Puertos de salida.** El Servicio de integración evalúa por último los puertos de salida, ya que estos puertos pueden hacer referencia a puertos de entrada y puertos de variables. El orden de visualización de los puertos de salida es indiferente porque los puertos de salida no pueden hacer referencia otros puertos de salida. Asegúrese de que los puertos de salida se muestren al final de la lista de puertos.

Tipo de datos

Cuando configure un puerto como una variable, podrá especificar cualquier expresión o condición en él. El tipo de datos que elija para este puerto refleja el valor de devolución de la expresión que especifique. Si especifica una condición mediante el puerto variable, cualquier tipo de datos numérico devuelve los valores para TRUE (distinto de cero) y FALSE (cero).

Inicialización de variables

El Servicio de integración no establece como NULL el valor inicial de las variables.

El Servicio de integración utiliza las siguientes directrices para establecer los valores iniciales de las variables:

- Cero para los puertos numéricos
- Cadenas vacías para los puertos de cadena
- 01/01/0001 para los puertos de fecha y hora

Por tanto, utilice variables como contadores, que necesitan un valor inicial. Por ejemplo, puede crear una variable numérica con la siguiente expresión:

```
VAR1 + 1
```

Esta expresión cuenta el número de filas en el puerto VAR1. Si el valor inicial de la variable se estableciera como NULL, la expresión siempre se evaluaría como NULL. Esta es la razón por la que el valor inicial se establece como cero.

Valores predeterminados para los puertos

Todas las transformaciones utilizan valores predeterminados que determinan la forma en que el Servicio de integración gestiona los valores nulos de entrada y los errores de transformación de salida.

Los puertos de entrada, salida y entrada/salida tienen un valor predeterminado del sistema que en ocasiones puede reemplazar con un valor predeterminado definido por el usuario. Los valores predeterminados tienen diferentes funciones en distintos tipos de puertos:

- **Puerto de entrada.** El valor predeterminado del sistema para los puertos de entrada nulos es NULL. El valor predeterminado aparece en blanco en la transformación. Si un valor de entrada es NULL, el Servicio de integración lo deja como NULL.
- **Puerto de salida.** El valor predeterminado del sistema para los errores de transformación de salida es ERROR. El valor predeterminado aparece en la transformación como ERROR('error de transformación'). Si se produce un error de transformación, el Servicio de integración omite la fila. El Servicio de integración anota todas las filas de entrada que la función ERROR omite en el archivo de registro.

Los siguientes errores son errores de transformación:

- Errores de conversión de datos, como pasar un número a una función de fecha.
- Errores de evaluación de expresiones, como dividir por cero.
- Llama a una función ERROR.

- **Puerto de transferencia** El valor predeterminado del sistema para una entrada null es como el de los puertos de entrada, NULL. El valor predeterminado del sistema aparece en blanco en la transformación. El valor predeterminado para los errores de transformación de salida es el mismo que el de los puertos de salida. El valor predeterminado para los errores de transformación de salida no se muestra en la transformación.

Nota: La transformación de Java convierte los tipos de datos de PowerCenter® en tipos de datos de Java, según el tipo de puerto de la transformación de Java. Los valores predeterminados para la entrada nula varían según el tipo de datos de Java.

La siguiente tabla muestra los valores predeterminados del sistema para los puertos en transformaciones conectadas:

Tipo de puerto	Valor predeterminado	Comportamiento del servicio de integración	Valor predeterminado definido por el usuario compatible
Entrada, transferencia	NULL	El servicio de integración pasa todos los valores nulos de entrada como NULL.	Entrada, entrada/salida
Salida, transferencia	ERROR	El servicio de integración llama a la función ERROR para los errores de transformación del puerto de salida. El servicio de integración omite las filas con errores y escribe los datos de entrada y el mensaje de error en el archivo de registro.	Salida

Los puertos de variables no admiten valores predeterminados. El Servicio de integración inicializa los puertos de variables según el tipo de datos.

Puede reemplazar algunos de los valores predeterminados para cambiar el comportamiento del Servicio de integración cuando detecta valores de entrada nulos y errores de transformación de salida.

Valores predeterminados definidos por el usuario

Puede reemplazar los valores predeterminados del sistema con valores predeterminados definidos por el usuario para los puertos compatibles de entrada, transferencia y salida dentro de una transformación conectada.

Utilice las siguientes reglas y directrices para reemplazar los valores predeterminados del sistema para los puertos:

- **Puertos de entrada.** Puede especificar valores predeterminados definidos por el usuario para puertos de entrada si no desea que el servicio de integración trate los valores nulos como NULL. Si se pasa NULL al puerto de entrada, el servicio de integración reemplaza NULL con el valor predeterminado.
- **Puertos de salida.** Puede especificar valores predeterminados definidos por el usuario para puertos de salida si no desea que el servicio de integración omita la fila o si desea que el servicio de integración escriba un mensaje específico con la fila omitida en el registro. Si define un valor predeterminado en el puerto de salida, el servicio de integración reemplaza la fila con el valor predeterminado cuando el puerto de salida tiene un error de transformación.
- **Puertos de transferencia.** Puede especificar valores predeterminados definidos por el usuario para puertos de transferencia si no desea que el servicio de integración trate los valores nulos como NULL. No puede especificar valores predeterminados definidos por el usuario para errores de transformación de salida en un puerto de transferencia.

Nota: El Servicio de integración omite los valores predeterminados definidos por el usuario para las transformaciones no conectadas. Por ejemplo, si llama a una transformación de búsqueda o de

procedimiento almacenado a través de una expresión, el Servicio de integración omite cualquier valor predeterminado definido por el usuario y aplica el valor predeterminado del sistema.

Utilice las siguientes opciones para especificar valores predeterminados definidos por el usuario:

- **Valor constante.** Utilizar cualquier constante (numérica o de texto), incluido NULL.
- **Expresión de constante:** "{ Puede incluir una función de transformación con parámetros de constante.
- **ERROR.** Generar un error de transformación. Escribir la fila y un mensaje en el registro de la asignación o en el registro de error de fila.
- **ABORT** Anular la asignación.

Valores constantes

Puede especificar cualquier valor constante como valor predeterminado. El valor constante debe coincidir con el tipo de datos del puerto.

Por ejemplo, un valor predeterminado para un puerto numérico debe ser una constante numérica. Algunos valores de constante son:

```
0
9999
NULL
'Unknown Value'
'Null input data'
```

Expresiones constantes

Una expresión de constante es cualquier expresión que utiliza las funciones de transformación (excepto las funciones de agregado) para escribir expresiones de constante. No se puede utilizar valores de los puertos de entrada, entrada/salida o variable en una expresión de constante.

Algunas expresiones de constante válidas serían:

```
500 * 1.75
TO_DATE('January 1, 1998, 12:05 AM','MONTH DD, YYYY, HH:MI AM')
ERROR ('Null not allowed')
ABORT('Null not allowed')
SESSSTARTTIME
```

No puede utilizar valores de los puertos de la expresión porque el Servicio de integración asigna valores predeterminados a toda la asignación cuando inicializa la asignación.

Los ejemplos siguientes no son válidos porque utilizan valores de puertos:

```
AVG(IN_SALARY)
IN_PRICE * IN_QUANTITY
:LKP(LKP_DATES, DATE_SHIPPED)
```

Nota: No puede llamar a un procedimiento almacenado o a una tabla de búsqueda desde una expresión de valor predeterminado.

Funciones ERROR y ABORT

Utilice las funciones ERROR y ABORT para los valores predeterminados de los puertos de entrada y de salida, y valores de entrada para los puertos de entrada/salida. El Servicio de integración omite la fila cuando detecta la función ERROR. La asignación se anula cuando detecta la función ABORT.

Valores de entrada predeterminados definidos por el usuario

Puede introducir un valor de entrada predeterminado definido por el usuario si no desea que el Servicio de integración trate los valores nulos como NULL.

Para reemplazar los valores nulos, realice una de las tareas siguientes:

- Reemplace el valor nulo con un valor constante o una expresión de constante.
- Omite el valor nulo con una función ERROR.
- Anule la asignación con la función ABORT.

La siguiente tabla resume la forma en que el Servicio de integración gestiona la entrada de valores nulos para los puertos de entrada y de entrada/salida:

Valor predeterminado	Tipo de valor predeterminado	Descripción
NULL (se muestra en blanco)	Sistema	El Servicio de integración pasa NULL.
Constante o expresión de constante	Definida por el usuario	El Servicio de integración reemplaza el valor nulo con el valor de la constante o de la expresión de constante.
ERROR	Definida por el usuario	El Servicio de integración lo trata como un error de transformación: - Aumenta el recuento de errores de transformación en 1. - Omite la fila y escribe el mensaje de error en el archivo de registro o en el registro de errores de fila. El Servicio de integración no escribe filas en el archivo de rechazo.
ABORT	Definida por el usuario	La asignación se anula cuando el Servicio de integración detecta un valor de entrada nulo. El Servicio de integración no aumenta el recuento de errores ni escribe filas en el archivo de rechazo.

Reemplazar valores nulos

Utilice un valor o una expresión de constante para sustituir un valor especificado para los valores nulos en un puerto.

Por ejemplo, si un puerto de cadena de entrada se llama DEPT_NAME y desea reemplazar los valores nulos con la cadena 'UNKNOWN DEPT', puede establecer el valor predeterminado como 'UNKNOWN DEPT'. En función de la transformación, el Servicio de integración pasa 'UNKNOWN DEPT' a una expresión o variable de la transformación o a la siguiente transformación del flujo de datos.

Por ejemplo, el Servicio de integración reemplaza todos los valores nulos de un puerto con la cadena 'UNKNOWN DEPT'.

DEPT_NAME	REPLACED VALUE
Housewares	Housewares
NULL	UNKNOWN DEPT
Produce	Produce

Omitir registros nulos

Utilice la función ERROR como valor predeterminado cuando no desee que los valores nulos pasen a una transformación. Por ejemplo, sería conveniente omitir una fila cuando el valor de entrada de DEPT_NAME es NULL. Puede utilizar la siguiente expresión como valor predeterminado:

```
ERROR('Error. DEPT is NULL')
```

Cuando se utiliza la función ERROR como valor predeterminado, el Servicio de integración omite la fila con el valor nulo. El Servicio de integración escribe todas las filas omitidas por la función ERROR en el archivo de registro. No escribe estas filas en el archivo de rechazo.

DEPT_NAME	RETURN VALUE
Housewares	Housewares
NULL	'Error. DEPT is NULL' (Row is skipped)
Produce	Produce

El siguiente registro muestra el lugar donde el Servicio de integración omite la fila con el valor nulo:

```
TE_11019 Port [DEPT_NAME]: Default value is: ERROR(<<Transformation Error>> [error]:
Error. DEPT is NULL
... error('Error. DEPT is NULL')
).
CMN_1053 EXPTRANS: : ERROR: NULL input column DEPT_NAME: Current Input data:
CMN_1053 Input row from SRCTRANS: Rowdata: ( RowType=4 Src Rowid=2 Targ Rowid=2
  DEPT_ID (DEPT_ID:Int:): "2"
  DEPT_NAME (DEPT_NAME:Char.25:): "NULL"
  MANAGER_ID (MANAGER_ID:Int:): "1"
)
```

Anular la asignación

Utilice la función ABORT para anular una asignación cuando el Servicio de integración detecte valores de entrada nulos.

Validación de valores predeterminados

La herramienta del desarrollador valida los valores predeterminados a medida que los introduce.

La herramienta del desarrollador también valida los valores predeterminados cuando se guarda una asignación. Si especifica un valor predeterminado que no es válido, la herramienta del desarrollador marca la asignación como no válida.

Valores de salida predeterminados definidos por el usuario

Puede crear valores predeterminados definidos por el usuario para reemplazar los valores predeterminados del sistema para los puertos de salida.

Puede especificar valores predeterminados definidos por el usuario para puertos de salida si no desea que el Servicio de integración omita filas con errores o si desea que el Servicio de integración escriba un mensaje específico con la fila omitida en el registro. Puede especificar valores predeterminados para completar las siguientes funciones cuando el Servicio de integración detecte errores de transformación de salida:

- Reemplazar el error con un valor constante o una expresión de constante. El Servicio de integración no omite la fila.
- Anule la asignación con la función ABORT.
- Escribir mensajes específicos en el registro para los errores de transformación.

No puede especificar valores de salida predeterminados definidos por el usuario para los puertos de entrada/salida.

La siguiente tabla resume la forma en que el Servicio de integración gestiona los errores de transformación del puerto de salida y los valores predeterminados en las transformaciones:

Valor predeterminado	Tipo de valor predeterminado	Descripción
Error de transformación	Sistema	Cuando se produce un error de transformación y no ha reemplazado el valor predeterminado, el Servicio de integración realiza las siguientes tareas: <ul style="list-style-type: none">- Aumenta el recuento de errores de transformación en 1.- Omite la fila y escribe el error y la fila de entrada en el archivo de registro de la sesión o en el registro de error de las filas, según la configuración de la sesión. El Servicio de integración no escribe la fila en el archivo de rechazo.
Constante o expresión de constante	Definida por el usuario	El Servicio de integración reemplaza el error con el valor predeterminado. El Servicio de integración no aumenta el recuento de errores ni escribe un mensaje en el registro.
ABORT	Definida por el usuario	La asignación se anula y el Servicio de integración escribe un mensaje en el registro. El Servicio de integración no aumenta el recuento de errores ni escribe filas en el archivo de rechazo.

Errores de reemplazo

Si no desea que el Servicio de integración omita una fila cuando se produzca un error de transformación, utilice una constante o expresión de constante como el valor predeterminado de un puerto de salida.

Por ejemplo, si tiene un puerto de salida numérico llamado NET_SALARY y desea utilizar el valor constante '9999' cuando se produce un error una transformación, asigne el valor predeterminado 9999 al puerto NET_SALARY. Si no hay ningún error de transformación (como dividir por cero) al calcular el valor de NET_SALARY, el Servicio de integración utiliza el valor predeterminado de 9999.

Anular la asignación

Utilice la función ABORT para anular una sesión cuando el Servicio de integración detecte valores de entrada nulos.

Escribir mensajes en el registro de asignación

Puede configurar un valor predeterminado definido por el usuario en un puerto de salida si desea que el Servicio de integración escriba un mensaje en el registro de asignación para una fila que se ha omitido. El sistema predeterminado es ERROR ('error de transformación') y el Servicio de integración escribe el mensaje 'error de transformación' en el registro junto con la fila que se ha omitido. Puede reemplazar 'error de transformación' si desea escribir un mensaje diferente.

Funciones ERROR en expresiones del puerto de salida

Si especifica una expresión que utiliza la función ERROR, el valor predeterminado definido por el usuario del puerto de salida podría reemplazar la función ERROR en la expresión.

Por ejemplo, introduzca la siguiente expresión que indica al Servicio de integración que utilice el valor 'Negative Sale' cuando detecte un error:

```
IIF( TOTAL_SALES>0, TOTAL_SALES, ERROR ('Negative Sale'))
```

Los siguientes ejemplos muestran cómo los valores predeterminados definidos por el usuario pueden reemplazar la función ERROR en la expresión:

- **Valor o expresión de constante.** El valor o la expresión de constante reemplaza la función ERROR en la expresión del puerto de salida.

Por ejemplo, si especifica '0' como valor predeterminado, el Servicio de integración reemplaza la función ERROR de la expresión del puerto de salida. Pasa el valor 0 cuando detecta un error. No omite la fila ni escribe 'Negative Sale' en el registro.

- **ABORT.** La función ABORT reemplaza la función ERROR en la expresión del puerto de salida.

Si utiliza la función ABORT como valor predeterminado, el Servicio de integración anula la fila cuando se produce un error de transformación. La función ABORT reemplaza la función ERROR en la expresión del puerto de salida.

- **ERROR.** Si utiliza la función ERROR como valor predeterminado, el Servicio de integración incluye la siguiente información en el registro:

- Mensaje de error del valor predeterminado
- Mensaje de error indicado en la función ERROR en la expresión del puerto de salida
- Fila omitida

Por ejemplo, puede reemplazar el valor predeterminado con la siguiente función ERROR:

```
ERROR('No default value')
```

El Servicio de integración omite la fila e incluye los dos mensajes de error en el registro.

```
TE_7007 Transformation Evaluation Error; current row skipped...
TE_7007 [<<Transformation Error>> [error]: Negative Sale
... error('Negative Sale')
]
Sun Sep 20 13:57:28 1998
TE_11019 Port [OUT_SALES]: Default value is: ERROR(<<Transformation Error>> [error]:
No default value
... error('No default value')
```

Reglas generales para los valores predeterminados

Utilice las siguientes reglas y directrices cuando cree los valores predeterminados:

- El valor predeterminado debe ser un valor NULL, un valor constante, una expresión de constante, una función ERROR o una función ABORT.
- Para los puertos de entrada/salida, el Servicio de integración utiliza valores predeterminados para gestionar valores de entrada nulos. El valor predeterminado de salida de los puertos de entrada/salida siempre es ERROR('Error de transformación').
- Los puertos de variables no utilizan valores predeterminados.
- Puede asignar valores predeterminados para agrupar por puertos en las transformaciones de agregación y de rango.
- No todos los tipos de puerto de todas las transformaciones permiten valores predeterminados definidos por el usuario. Si un puerto no permite valores predeterminados definidos por el usuario, el campo de valor predeterminado estará deshabilitado.
- No todas las transformaciones permiten valores predeterminados definidos por el usuario.
- Si una transformación no está conectada al flujo de datos de la asignación, el Servicio de integración omite los valores predeterminados definidos por el usuario.

- Si un puerto de entrada no está conectado, se supondrá que su valor es NULL y el Servicio de integración utilizará el valor predeterminado para ese puerto de entrada.
- Si un valor predeterminado de puerto de entrada contiene la función ABORT y el valor de entrada es NULL, el Servicio de integración detendrá inmediatamente la asignación. Utilice la función ABORT como valor predeterminado para restringir valores de entrada nulos. El primer valor nulo de un puerto de entrada detendrá la asignación.
- Si un valor predeterminado de puerto de salida contiene la función ABORT y se produce cualquier error de transformación para ese puerto, la asignación se detendrá inmediatamente. Utilice la función ABORT como valor predeterminado para aplicar reglas estrictas para los errores de transformación. El primer error de transformación de este puerto detendrá la asignación.
- La función ABORT, los valores constantes y las expresiones de constantes reemplazan las funciones ERROR configuradas en las expresiones del puerto de salida.

Validación de valores predeterminados

La herramienta del desarrollador valida los valores predeterminados a medida que los introduce.

La herramienta del desarrollador también valida los valores predeterminados cuando se guarda una asignación. Si especifica un valor predeterminado que no es válido, la herramienta del desarrollador marca la asignación como no válida.

Niveles de seguimiento

Cuando configure una transformación, puede establecer la cantidad de detalles que el servicio de integración de datos escribe en el registro.

De manera predeterminada, el nivel de seguimiento para cada transformación es Normal. Cambie el nivel de seguimiento a una configuración de detalle si necesita solucionar problemas en una transformación que no tiene el comportamiento esperado. Establezca el nivel de seguimiento en Simple cuando desee que aparezca la cantidad mínima de detalle en el registro.

Configure la siguiente propiedad en la ficha Avanzadas:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro para una transformación.

En la siguiente tabla se describen los niveles de seguimiento:

Nivel de seguimiento	Descripción
Simplificado	Registra información sobre la inicialización y mensajes de error y notificaciones de datos rechazados.
Normal	Registra información sobre la inicialización y el estado, los errores que aparecen y las filas que se omitieron debido a errores en filas de transformación. Resume los resultados de asignación, aunque no en el nivel de filas individuales. El valor predeterminado es Normal.

Nivel de seguimiento	Descripción
Inicialización detallada	Además del seguimiento normal, registra detalles de inicialización adicionales, nombres de los archivos de datos e índices utilizados, y estadísticas de transformación detalladas.
Datos detallados	Además del seguimiento de inicialización detallada, registra cada fila que pasa por la asignación. También indica dónde se truncan los datos de la cadena para ajustarse a la precisión de una columna, y proporciona estadísticas detalladas de transformación. Cuando se configura este nivel de seguimiento, los datos de fila de todas las filas de un bloque se escriben en el registro cuando se procesa una transformación.

Transformaciones reutilizables

Las transformaciones reutilizables son transformaciones que se pueden usar en varias asignaciones o mapplets.

Por ejemplo, puede crear una transformación de expresión que calcule el impuesto de valor añadido en Canadá para analizar el coste de operar en ese país. En lugar de hacer el mismo trabajo cada vez, puede crear una transformación reutilizable. Cuando se desea incorporar esta transformación a una asignación, se añade una instancia de la misma a la asignación. Si se cambia la definición de la transformación, todas las instancias de la misma heredan los cambios.

Developer Tool almacena cada transformación reutilizable como metadatos aparte de cualquier asignación o mapplet que utilizan la transformación. Almacena las transformaciones reutilizables en un proyecto o una carpeta.

Cuando se añaden instancias de una transformación reutilizable a asignaciones, los cambios efectuados en la transformación pueden invalidar la asignación o generar datos inesperados.

Instancias de transformación reutilizable y cambios heredados

Cuando se añade una transformación reutilizable a una asignación o un mapplet, se añade una instancia de la transformación. La definición de la transformación todavía existe fuera de la asignación o del mapplet, mientras que una instancia de la transformación aparece dentro de la asignación o el mapplet.

Cuando se cambia la transformación, las instancias de la transformación reflejan los cambios efectuados. En lugar de actualizar la misma transformación en todas las asignaciones que la utilicen, puede actualizar la transformación reutilizable una vez, con lo que todas las instancias de la transformación heredarán los cambios. Las instancias heredan los cambios realizados en puertos, expresiones, propiedades y el nombre de la transformación.

Cómo editar una transformación reutilizable

Cuando se edita una transformación reutilizable, todas las instancias de esa transformación heredan los cambios. Algunos cambios pueden invalidar las asignaciones que utilizan la transformación reutilizable.

Puede abrir la transformación en el editor para editar una transformación reutilizable. No se puede editar una instancia de la transformación en una asignación. Sin embargo, puede editar las propiedades de tiempo de ejecución de la transformación.

Si efectúa alguno de los siguientes cambios en una transformación reutilizable, las asignaciones que utilizan instancias de la misma pueden no ser válidas:

- Cuando se elimina uno o más puertos de una transformación se desconecta la instancia de parte o todos los datos de la asignación.
- Cuando se cambia el tipo de datos de los datos de un puerto hace que sea imposible asignar los datos de ese puerto a otro puerto que utiliza un tipo de datos incompatible.
- Cuando se cambia el nombre de un puerto, las expresiones que se refieren al puerto dejan de ser válidas.
- Cuando se especifica una expresión que no es válida en la transformación reutilizable, las asignaciones que utilizan la transformación dejan de ser válidas. El servicio de integración de datos no puede ejecutar asignaciones que no son válidas.

Vistas de editor para una transformación de reutilizable

Se definen propiedades y se crea código Java para una transformación de Java en vistas del editor.

Para las transformaciones de reutilizables, están disponibles las siguientes vistas:

Resumen

Escriba el nombre y la descripción de la transformación y cree y configure los puertos de entrada y de salida.

Avanzadas

Defina propiedades avanzadas para la transformación.

Transformaciones no reutilizables

Una transformación no reutilizable es una transformación que se crea en una asignación específica. No podrá utilizar la transformación en ninguna otra asignación.

Por ejemplo, podría crear una asignación que contenga varias transformaciones. Cada transformación realizará cálculos en los datos de origen. Cree una transformación de agregación no reutilizable al final de la asignación para procesar los resultados. Cuando cree una transformación no reutilizable, podrá arrastrar los puertos de una transformación a otra transformación de la asignación y crear los puertos de entrada.

La herramienta del desarrollador almacena la transformación de agregación no reutilizable como metadatos que guarda con la asignación.

Vistas de editor para una transformación de no reutilizable

Defina las propiedades para una transformación no reutilizable en las vistas del editor.

Las siguientes vistas aparecen para las transformaciones no reutilizables:

General

Proporcione el nombre y la descripción de la transformación.

Puertos

Cree y configure puertos de entrada y de salida.

Avanzadas

Defina propiedades avanzadas para la transformación.

Cómo crear una transformación

Puede crear una transformación reutilizable para reutilizarla en varias asignaciones o mapplets. También puede crear una transformación no reutilizable para usarla una vez en una asignación o en un mapplet.

Para crear una transformación reutilizable, haga clic en **Archivo > Nueva > Transformación**, y complete el asistente.

Para crear una transformación no reutilizable en una asignación o en un mapplet, seleccione una transformación en la paleta Transformación y arrástrela hasta el editor.

Algunas transformaciones requieren que se elija un modo o que se lleve a cabo una configuración adicional cuando se crea la transformación. Por ejemplo, cuando se crea una transformación de búsqueda, se debe elegir un objeto de datos para utilizarlo como origen de búsqueda.

Después de crear una transformación, ésta aparece en el editor. Algunas transformaciones contienen puertos y grupos predefinidos. Otras transformaciones están vacías.

CAPÍTULO 2

Puertos de transformación

Este capítulo incluye los siguientes temas:

- [Resumen de puertos de transformación, 61](#)
- [Crear puertos, 61](#)
- [Configurar puertos, 62](#)
- [Cómo vincular puertos, 62](#)
- [Cómo propagar atributos de puerto, 65](#)
- [Copiar puertos desde Excel, 68](#)

Resumen de puertos de transformación

Los puertos de transformación son las columnas individuales de datos que se conectan en una asignación o un mapplet. Las transformaciones reciben los datos de puertos de entrada y los envían utilizando puertos de salida. Los puertos de entrada/salida reciben datos y los pasan sin cambios.

Cada objeto de entrada, objeto de salida, mapplet y transformación contiene una colección de puertos. Los objetos de entrada proporcionan datos, de modo que solo contienen puertos de salida. Los objetos de salida reciben datos, por lo que solo contienen puertos de entrada. Los mapplets solo contienen puertos de entrada y de salida. Las transformaciones contienen una combinación de puertos de entrada, puertos de salida y puertos de entrada/salida, en función de la transformación y su aplicación.

Un puerto dinámico puede recibir una o más columnas de una transformación de nivel superior. Los puertos dinámicos pueden recibir columnas nuevas o modificadas basadas en el flujo de datos.

Después de crear una transformación en una asignación, cree los puertos y defina las propiedades del puerto. Complete la asignación vinculándola a sus destinos y a otras transformaciones a través de los puertos. Propague los atributos de los puertos para pasar los atributos modificados a un puerto a través de una asignación.

Crear puertos

Cuando se crean algunas transformaciones, no es necesario crear todos los puertos manualmente. Por ejemplo, podría crear una transformación de búsqueda y hacer referencia a una tabla de búsqueda. Si ve los puertos de transformación, podrá observar que la transformación tiene un puerto de salida para cada columna de la tabla a la que ha hecho referencia. No es necesario definir esos puertos.

Cree un puerto de las siguientes maneras:

- **Arrastre un puerto desde otra transformación.** Cuando se arrastra un puerto desde otra transformación, Designer crea un puerto con las mismas propiedades y vincula los dos puertos.
- **Haga clic en el botón Añadir en la ficha Puertos.** Designer crea un puerto vacío que puede configurar.

Configurar puertos

Cuando defina los puertos de transformación, defina las propiedades de los puertos. Las propiedades de los puertos incluyen el nombre de puerto, el tipo de datos, el tipo de puerto y el valor predeterminado.

Configure las siguientes propiedades de los puertos:

- **Nombre de puerto.** El nombre del puerto. Utilice las siguientes convenciones para dar nombre a los puertos:
 - Empiece por una letra de byte único o doble byte o por un guión bajo de byte único o doble byte (_).
 - Los nombres de los puertos pueden contener los siguientes caracteres de byte único o de doble byte: letras, números, guiones bajos (_), \$, # o @.
- **Tipo de datos, precisión y escala.** Si tiene pensado especificar una expresión o una condición, compruebe que el tipo de datos coincide con el valor de devolución de la expresión.
- **Tipo de puerto.** Las transformaciones pueden contener una combinación de tipos de puerto de entrada, salida, entrada/salida y variables.
- **Valor predeterminado.** Asigne un valor predeterminado para un puerto que contiene valores null o un error de la transformación de salida. Puede reemplazar el valor predeterminado en algunos puertos.
- **Descripción.** Una descripción del puerto.
- **Otras propiedades.** Algunas transformaciones tienen propiedades específicas de esa transformación, como expresiones o las propiedades Agrupar por.

Cómo vincular puertos

Después de añadir y configurar objetos de entrada, salida, transformación y mapplet en una asignación, finalice la asignación vinculando los puertos entre los objetos de la misma.

Developer Tool solamente crea la conexión cuando ésta cumple los requisitos de validación y concatenación de vínculos.

Puede dejar puertos sin conectar. El servicio de integración de datos omite los puertos no conectados.

Cuando se vinculan puertos entre objetos de entrada, transformaciones, mapplets y objetos de salida, puede crear los siguientes tipos de vínculo:

- Uno a uno
- Uno a muchos

Puede vincular los puertos tanto manual como automáticamente.

Vínculos "uno a uno"

Vincule un puerto de un objeto o transformación de entrada con un puerto de un objeto o transformación de salida.

Vínculos "uno a varios"

Cuando desee utilizar los mismos datos para propósitos diferentes, puede vincular el puerto que suministra los datos con varios puertos de la asignación.

Puede crear un vínculo "uno a varios" de las siguientes maneras:

- Vinculando un puerto con varias transformaciones u objetos de salida.
- Vinculando varios puertos de una transformación con varias transformaciones u objetos de salida.

Por ejemplo, desea utilizar la información de los salarios para calcular el salario medio en una sucursal bancaria mediante la transformación de agregación. Puede utilizar la misma información en una transformación de expresión configurada para calcular la paga mensual de cada empleado.

Cómo vincular puertos manualmente

Puede vincular manualmente uno o varios puertos.

Arrastre un puerto desde un objeto o transformación de entrada hasta el puerto de un objeto o transformación de salida.

Utilice la tecla Ctrl o Mayús para seleccionar varios puertos y vincularlos a otra transformación u objeto de salida. Developer Tool vincula los puertos, comenzando por el par superior. Vincula todos los puertos que cumplen los requisitos de validación.

Cuando se arrastra un puerto hasta un puerto vacío, Developer Tool copia el puerto y crea un vínculo.

Cómo vincular puertos automáticamente

Cuando se vinculan puertos automáticamente, se pueden vincular por posición o por nombre.

Cuando se vinculan puertos automáticamente por nombre, puede especificar un prefijo o sufijo según los cuales vincular los puertos. Utilice prefijos o sufijos para indicar en qué lugares de una asignación ocurren los puertos.

Cómo vincular puertos por nombre

Cuando se vinculan puertos por nombre, Developer Tool añade enlaces entre los puertos de entrada y salida que tienen el mismo nombre. Vincule por nombre cuando utilice los mismos nombres de puerto en distintas transformaciones.

Puede vincular puertos según los prefijos y sufijos que defina. Utilice prefijos o sufijos para indicar en qué lugares de una asignación ocurren los puertos. Enlace por nombre y prefijo o sufijo si utiliza prefijos o sufijos en puertos de nombres para distinguir dónde ocurren en la asignación o el mapplet.

El vínculo por nombre no distingue entre mayúsculas y minúsculas.

1. Haga clic en **Asignación > Vínculo automático**.
A continuación, aparece el cuadro de diálogo **Vínculo automático**.
2. Seleccione un objeto en la ventana **Desde** desde el cual efectuar el vínculo.

3. Seleccione un objeto en la ventana **Hasta** hasta el cual efectuar el vínculo.
4. Seleccione **Nombre**.
5. Si lo desea, haga clic en **Mostrar opciones avanzadas** para vincular puertos según prefijos o sufijos.
6. Haga clic en **Aceptar**.

Cómo vincular puertos por posición

Cuando se vincula por posición, Developer Tool vincula cada puerto de salida con el puerto de entrada correspondiente. Por ejemplo, el primer puerto de salida se vincula con el primer puerto de entrada y el segundo puerto de salida con el segundo puerto de entrada. Vincule por posición cuando cree transformaciones con puertos relacionados en el mismo orden.

1. Haga clic en **Asignación > Vínculo automático**.
A continuación, aparece el cuadro de diálogo **Vínculo automático**.
2. Seleccione un objeto en la ventana **Desde** desde el cual efectuar el vínculo.
3. Seleccione un objeto en la ventana **Hasta** hasta el cual efectuar el vínculo.
4. Seleccione **Posición** y haga clic en **Aceptar**.
Developer Tool vincula cada puerto de salida con el puerto de entrada correspondiente. Por ejemplo, el primer puerto de salida se vincula con el primer puerto de entrada y el segundo puerto de salida con el segundo puerto de entrada.

Reglas y directrices para vincular puertos

Hay ciertas reglas y directrices que se aplican cuando se vinculan puertos.

Tenga en cuenta las siguientes reglas y directrices al conectar objetos de asignación:

- Si Developer Tool detecta un error cuando se intenta vincular puertos entre dos objetos de asignación, muestra un símbolo que indica que no se pueden vincular los puertos.
- Siga la lógica del flujo de datos de la asignación. Puede vincular los siguientes tipos de puertos:
 - El puerto receptor debe ser un puerto de entrada o entrada/salida.
 - El puerto de origen debe ser un puerto de entrada o entrada/salida.
 - No se pueden vincular puertos de entrada con puertos de entrada o puertos de salida con puertos de salida.
- Debe vincular como mínimo un puerto de un grupo de entrada con una transformación de nivel superior.
- Debe vincular como mínimo un puerto de un grupo de entrada con una transformación de nivel inferior.
- Se pueden vincular puertos desde una transformación activa o un grupo de salida de una transformación activa con un grupo de entrada de otra transformación.
- No puede conectar una transformación activa y una transformación pasiva a la misma transformación de nivel inferior o grupo de entrada de transformación.
- No puede conectar más de una transformación activa a la misma transformación de nivel inferior o grupo de entrada de transformación.
- Puede conectar cualquier número de transformaciones pasivas a la misma transformación de nivel inferior, grupo de entrada de transformación o destino.
- Puede vincular puertos desde dos grupos de salida de la misma transformación con una transformación de incorporación configurada para datos ordenados en el caso de que los datos de ambos grupos de salida estén ordenados.

- Solamente puede vincular puertos con tipos de datos compatibles. Developer Tool comprueba si puede realizar la asignación entre dos tipos de datos antes de vincularlos. El servicio de integración de datos no puede transformar datos entre puertos con tipos de datos incompatibles..
- Developer Tool marca ciertas asignaciones como no válidas si la asignación infringe la validación del flujo de datos.

Cómo propagar atributos de puerto

Propague atributos de puerto para pasar atributos cambiados a un puerto a través de una asignación.

1. En el editor, seleccione un puerto de una transformación.
2. Haga clic en **Asignación > Propagar atributos**.
A continuación, aparece el cuadro de diálogo **Propagar atributos**.
3. Seleccione una dirección a la que desee que se propaguen los atributos.
4. Seleccione los atributos que desee propagar.
5. Si lo desea, puede previsualizar los resultados.
6. Haga clic en **Aplicar**.
Developer Tool propaga los atributos de puerto.

Tipos de dependencia

Cuando se propagan atributos de puerto, Developer Tool actualiza las dependencias.

Developer Tool puede actualizar las siguientes dependencias:

- Dependencias de la ruta de acceso al enlace
- Dependencias implícitas

Dependencias de la ruta de acceso al enlace

Una dependencia de la ruta de acceso al enlace es una dependencia entre un puerto propagado y los puertos en su ruta de acceso al enlace.

Cuando se propagan dependencias en una ruta a un enlace, Developer Tool actualiza todos los puertos de entrada y de entrada/salida en su ruta al enlace posterior y todos los puertos de salida y de salida/entrada en su ruta al enlace anterior. Developer Tool realiza las siguientes actualizaciones:

- Actualiza el nombre de puerto, el tipo de datos, la precisión, la escala y la descripción para todos los puertos en la ruta al enlace del puerto propagado.
- Actualiza todas las expresiones y condiciones que hacen referencia al puerto propagado con el nombre de puerto cambiado.
- Actualiza la propiedad de puerto asociado en una transformación de búsqueda en el caso de que cambie el nombre del puerto asociado.

Dependencias implícitas

Dentro de una transformación, una dependencia implícita es una dependencia basada en una expresión o condición que hay entre dos puertos.

Con dependencias implícitas, puede propagar el tipo de datos, la precisión, la escala y la descripción a puertos. También puede analizar condiciones y expresiones para identificar las dependencias implícitas del puerto propagado. Todos los puertos con dependencias implícitas son puertos de salida o de entrada/salida.

Cuando se incluyen condiciones, Developer Tool actualiza las siguientes dependencias:

- Dependencias de la ruta de acceso al enlace
- Puertos de salida utilizados en la misma condición de búsqueda que el puerto propagado
- Puertos asociados en transformaciones de búsqueda dinámicas que están asociadas con el puerto propagado
- Puertos maestros utilizados en la misma condición de unión que el puerto de detalle

Cuando se incluyen expresiones, Developer Tool actualiza las siguientes dependencias:

- Dependencias de la ruta de acceso al enlace
- Puertos de salida que contienen una expresión que utiliza el puerto propagado

Developer Tool no se propaga a dependencias implícitas dentro de una misma transformación. Debe propagar los atributos cambiados desde otra transformación. Por ejemplo, cuando se cambia el tipo de datos de un puerto que se utiliza en una condición de búsqueda y se propaga ese cambio desde la transformación de búsqueda, Developer Tool no propaga el cambio al otro puerto dependiente de la condición en la misma transformación de búsqueda.

Atributos de puerto propagados por transformación

Developer Tool propaga las dependencias y atributos para cada transformación.

La siguiente tabla describe las dependencias y atributos que Developer Tool propaga para cada transformación.

Transformación	Dependencia	Atributos propagados
Validador de direcciones	Ninguno.	Ninguno. Esta transformación tiene nombres de puerto y tipos de datos predefinidos.
Agregación	<ul style="list-style-type: none">- Puertos en la ruta de enlace- Expresión- Dependencias implícitas	<ul style="list-style-type: none">- Nombre de puerto, tipo de datos, precisión, escala, descripción- Nombre de puerto- Tipo de datos, precisión, escala
Asociación	<ul style="list-style-type: none">- Puertos en la ruta de enlace	<ul style="list-style-type: none">- Nombre de puerto, tipo de datos, precisión, escala, descripción
Conversión de mayúsculas y minúsculas	<ul style="list-style-type: none">- Puertos en la ruta de enlace	<ul style="list-style-type: none">- Nombre de puerto, tipo de datos, precisión, escala, descripción
Clasificador	<ul style="list-style-type: none">- Puertos en la ruta de enlace	<ul style="list-style-type: none">- Nombre de puerto, tipo de datos, precisión, escala, descripción
Comparación	<ul style="list-style-type: none">- Puertos en la ruta de enlace	<ul style="list-style-type: none">- Nombre de puerto, tipo de datos, precisión, escala, descripción

Transformación	Dependencia	Atributos propagados
Consolidador	Ninguno.	Ninguno. Esta transformación tiene nombres de puerto y tipos de datos predefinidos.
Enmascaramiento de datos	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Procesador de datos	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Decisión	- Puertos de nivel inferior en la ruta al enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Expresión	- Puertos en la ruta de enlace - Expresión - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto - Tipo de datos, precisión, escala
Filtro	- Puertos en la ruta de enlace - Condición	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto
De jerárquica a relacional	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Combinación	- Puertos en la ruta de enlace - Condición - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto - Tipo de datos, precisión, escala
Generador de claves	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Etiquetador	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Búsqueda	- Puertos en la ruta de enlace - Condición - Puertos asociados (búsqueda dinámica) - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto - Nombre de puerto - Tipo de datos, precisión, escala
Coincidencia	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Fusión	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Normalizador	- Puertos en la ruta de enlace	- Nombre de puerto
Analizador	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Rango	- Puertos en la ruta de enlace - Expresión - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto - Tipo de datos, precisión, escala

Transformación	Dependencia	Atributos propagados
Lectura		
Consumidor de servicio web REST	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Enrutador	- Puertos en la ruta de enlace - Condición	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto
Generador de secuencia	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Ordenación	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
SQL	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Estandarizador	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Unión	- Puertos en la ruta de enlace - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Tipo de datos, precisión, escala
Estrategia de actualización	- Puertos en la ruta de enlace - Expresión - Dependencias implícitas	- Nombre de puerto, tipo de datos, precisión, escala, descripción - Nombre de puerto - Tipo de datos, precisión, escala
Media ponderada	- Puertos en la ruta de enlace	- Nombre de puerto, tipo de datos, precisión, escala, descripción
Escritura		

Copiar puertos desde Excel

Puede configurar los puertos y las propiedades de puerto en Excel y copiarlos en los puertos de transformación en Developer tool. Las propiedades de puerto incluyen el nombre de columna, el tipo de datos, la precisión y la escala. Esto puede resultar útil cuando necesite desarrollar o editar una transformación con muchos puertos.

Puede copiar metadatos a los siguientes tipos de transformación:

- Agregación
- Expresión
- Filtro
- Java
- Combinación
- Búsqueda

- Normalizador
- Rango
- Lectura
- Enrutador
- Secuencia
- Ordenación
- SQL
- Unión
- Estrategia de actualización
- Consumidor de servicio web
- Ventana
- Escritura

Edición de transformaciones en Excel

Cuando necesita editar una gran parte de una transformación, no es necesario cambiar cada valor en Developer tool. En vez de ello, puede copiar los puertos de transformación en Excel, cambiar todos los valores de forma simultánea con Autorrellenar y, a continuación, debe **Pegar (reemplazar)** los puertos de transformación de nuevo en Developer tool.

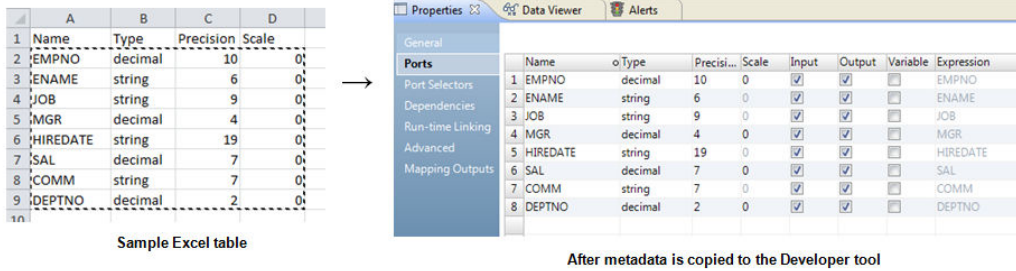
1. En el editor de asignaciones de Developer tool, seleccione la transformación de la que desea copiar los puertos.
2. Para copiar la transformación original de Developer tool, haga clic con el botón derecho en Puertos y haga clic en **Seleccionar todo**.
3. Copie los puertos a una hoja de cálculo de Excel.
4. Realice cambios en la hoja de cálculo de Excel. Si va a cambiar gran parte de los metadatos, puede utilizar la función Autorrellenar en Excel. Esto le permite rellenar datos en función de las celdas adyacentes arrastrando el controlador de relleno. Consulte [“Ejemplo: Editar una transformación en Excel” en la página 70](#) para obtener más información.
5. Copie los metadatos desde Excel.
6. Para actualizar la transformación con los cambios, haga clic con el botón derecho en Puertos y haga clic en **Pegar (reemplazar)**.

Copiar metadatos a Developer tool

Puede crear puertos de transformación en Excel y, a continuación, copiarlos a Developer tool.

1. Cree una asignación en Developer tool con las transformaciones necesarias.
2. Defina metadatos para una transformación en Excel.
3. Copie los metadatos desde Excel.
4. Para mover los metadatos a la transformación en Developer tool, haga clic con el botón derecho en Puertos y haga clic en **Pegar (reemplazar)**.

La siguiente imagen muestra una tabla de Excel de ejemplo y la transformación resultante después de copiar los metadatos a Developer tool:



Nota: Debe confirmar que los valores en cada celda son válidos antes de copiarlos a una transformación. Por ejemplo, un tipo de cadena no puede tener un valor de escala distinto de "0". Los valores de precisión no pueden ser palabras y los valores de tipo no pueden ser números. Si los metadatos son incorrectos, aparecerá un mensaje de error.

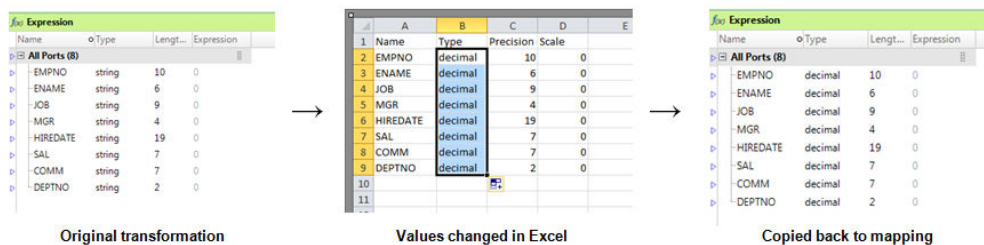
Ejemplo: Editar una transformación en Excel

Está desarrollando una transformación y tiene que cambiar todos los tipos de datos de cadena a decimal. En vez de cambiar cada campo individualmente, puede realizar cambios globales en Excel y copiarlos a Developer tool.

1. Haga clic con el botón derecho en Puertos, haga clic en **Seleccionar todo** y pegue los metadatos en Excel.
2. Puede cambiar el primer valor de tipo de datos de «cadena» a «decimal» y, a continuación, usar el controlador de relleno para cambiar automáticamente las celdas restantes de la columna.
3. Para actualizar la transformación con los cambios, copie los metadatos de Excel, haga clic con el botón derecho en Puertos y haga clic en **Pegar (reemplazar)**.

Con Excel, evita tener que cambiar cada campo de forma individual.

La siguiente imagen muestra el proceso de mover una transformación a Excel, usar Autorrellenar para cambiar algunos valores y, a continuación, copiar la transformación de nuevo en Developer tool:



Reglas y directrices para copiar desde Excel

Tenga en cuenta las siguientes reglas y directrices al copiar metadatos de Excel a Developer tool:

- Debe confirmar que los valores en cada celda son válidos antes de copiarlos a una transformación. Por ejemplo, un tipo de cadena no puede tener un valor de escala distinto de "0". Los valores de precisión no pueden ser palabras y los valores de tipo no pueden ser números. Si los metadatos son incorrectos, aparecerá un mensaje de error.
- La ubicación a la que pueda copiar los metadatos dependerá del tipo de transformación que esté actualizando. Por ejemplo, la opción Pegar (reemplazar) podría no estar disponible al hacer clic con el botón derecho en la vista Propiedades de una transformación. Sin embargo, la opción seguirá estando disponible si hace clic con el botón derecho directamente en los puertos de transformación en el editor.

CAPÍTULO 3

Memorias caché de transformaciones

Este capítulo incluye los siguientes temas:

- [Resumen de memorias caché de transformaciones, 72](#)
- [Tipos de memoria caché, 73](#)
- [Archivos de memoria caché, 73](#)
- [Tamaño de la memoria caché, 74](#)
- [Aumento del tamaño de la memoria caché por el Servicio de integración de datos, 77](#)
- [Tamaño de la memoria caché para memorias caché con particiones, 77](#)
- [Optimización del tamaño de la memoria caché, 78](#)

Resumen de memorias caché de transformaciones

El Servicio de integración de datos asigna memoria caché a las transformaciones de agregación, combinación, búsqueda, rango y ordenación en una asignación. El Servicio de integración de datos crea memorias caché de índice y de datos para las transformaciones de agregación, combinación, búsqueda y rango. El Servicio de integración de datos crea una memoria caché para la transformación de ordenación.

Puede configurar el tamaño de las memorias caché para estas transformaciones. El tamaño de la memoria caché determina cuánta memoria asigna el Servicio de integración de datos a cada memoria caché de una transformación al inicio de la ejecución de una asignación.

Si el tamaño de la memoria caché es mayor que la memoria disponible en el equipo, el Servicio de integración de datos no puede asignar suficiente memoria y la asignación no se ejecutará correctamente.

Si el tamaño de la memoria caché es menor que la cantidad de memoria necesaria para ejecutar la transformación, el Servicio de integración de datos procesa parte de la transformación en la memoria y almacena los datos desbordados en archivos de memoria caché. Cuando el Servicio de integración de datos pagina archivos de memoria caché en el disco, el tiempo de procesamiento aumenta. Para obtener un rendimiento óptimo, configure el tamaño de la memoria caché de modo que el Servicio de integración de datos pueda procesar la transformación completa en la memoria.

De forma predeterminada, el Servicio de integración de datos calcula automáticamente los requisitos de memoria en tiempo de ejecución, según la cantidad máxima de memoria que el servicio pueda asignar. Después de ejecutar una asignación en modo de memoria caché automática, puede ajustar los tamaños de la memoria caché para las transformaciones. Analice las estadísticas de la transformación en el registro de

la asignación para determinar los tamaños de memoria caché necesarios para obtener un rendimiento óptimo y, a continuación, configure tamaños de memoria caché específicos para las transformaciones.

Tipos de memoria caché

Las transformaciones de agregación, combinación, búsqueda y rango requieren una memoria caché de índice y una memoria caché de datos. El Servicio de integración de datos almacena los valores clave en la memoria caché de índice y los valores de salida en la memoria caché de datos. Las transformaciones de ordenación requieren una única memoria caché. El Servicio de integración de datos almacena claves y datos para ordenar en la memoria caché de ordenación.

En la siguiente tabla se describe el tipo de información que el Servicio de integración de datos almacena en cada memoria caché:

Objeto de asignación	Tipos de memoria caché y descripciones
Agregación	<ul style="list-style-type: none">- Índice. Almacena valores de grupo tal y como están configurados en los puertos de agrupación.- Datos. Almacena cálculos basados en los puertos de agrupación.
Combinación	<ul style="list-style-type: none">- Índice. Almacena todas las filas principales en la condición de combinación que tienen claves únicas.- Datos. Almacena las filas de origen maestras.
Búsqueda	<ul style="list-style-type: none">- Índice. Almacena la información de condición búsqueda.- Datos. Almacena los datos de búsqueda que no se almacenan en la memoria caché de índice.
Rango	<ul style="list-style-type: none">- Índice. Almacena valores de grupo tal y como están configurados en los puertos de agrupación.- Datos. Almacena la información de clasificación basada en los puertos de agrupación.
Ordenación	<ul style="list-style-type: none">- Ordenación. Almacena datos y claves de ordenación.

Archivos de memoria caché

Al ejecutar una asignación, el Servicio de integración de datos crea al menos un archivo de memoria caché para cada transformación de agregación, combinación, búsqueda, rango y ordenación. Si el Servicio de

integración de datos no puede ejecutar una transformación en la memoria, escribe los datos desbordados en los archivos de memoria caché.

En la siguiente tabla se describen los tipos de archivos de memoria caché que el Servicio de integración de datos crea para los diferentes objetos de asignación:

Objeto de asignación	Archivo de memoria caché
Transformaciones de agregación, combinación, búsqueda y rango	El Servicio de integración de datos crea los siguientes tipos de archivos de memoria caché: <ul style="list-style-type: none">- Un archivo de encabezado para cada memoria caché de índice y de datos- Un archivo de datos para cada memoria caché de índice y de datos
Transformación de ordenación	El Servicio de integración de datos crea un archivo de memoria caché de ordenación.

Al ejecutar una asignación, el Servicio de integración de datos escribe un mensaje en el registro de la asignación indicando el nombre del archivo de memoria caché y el nombre de la transformación. Cuando una asignación finaliza, el Servicio de integración de datos libera la memoria caché y normalmente elimina los archivos de memoria caché. Es posible encontrar archivos de memoria caché de índice y de datos en el directorio de la memoria caché en las siguientes circunstancias:

- Puede configurar la transformación de búsqueda para que utilice una memoria caché constante.
- La asignación no se completa correctamente. La próxima vez que ejecute la asignación, el Servicio de integración de datos eliminará los archivos de memoria caché existentes y creará otros nuevos.

Dado que escribir en los archivos de memoria caché afecta al rendimiento de las asignaciones, configure el tamaño de las memorias caché para ejecutar la transformación en la memoria.

Directorio de archivos de memoria caché

Para las transformaciones de agregación, combinación, búsqueda y rango, el Servicio de integración de datos crea los archivos de memoria caché en el directorio especificado en la propiedad Directorio de la memoria caché. Para las transformaciones de ordenación, el servicio crea los archivos de memoria caché en el directorio especificado en la propiedad Directorio de trabajo.

Si el proceso del Servicio de integración de datos no encuentra el directorio, se producirá un error en la asignación y se escribirá un mensaje en el registro de asignaciones indicando que no se pudo crear o abrir el archivo de memoria caché. El Servicio de integración de datos puede crear varios archivos de memoria caché. El número de archivos de memoria caché está limitado a la cantidad de espacio en disco disponible en el directorio de la memoria caché.

Tamaño de la memoria caché

El tamaño de la memoria caché determina cuánta memoria asigna el Servicio de integración de datos a cada memoria caché de una transformación al inicio de la ejecución de una asignación. Puede configurar el

tamaño de la memoria caché de una transformación para que use el modo de memoria caché automático o especificar un valor.

Tamaño de la memoria caché automático

De forma predeterminada, el tamaño de la memoria caché de una transformación se establece como Auto. El Servicio de integración de datos calcula automáticamente los requisitos de la memoria caché en tiempo de ejecución. Es posible definir la cantidad máxima de memoria que puede asignar el servicio.

El Servicio de integración de datos utiliza las siguientes directrices para asignar automáticamente la memoria:

Asigna más memoria a las transformaciones con un mayor tiempo de procesamiento.

El Servicio de integración de datos asigna más memoria a las transformaciones que suelen requerir un mayor tiempo de procesamiento. Por ejemplo, el Servicio de integración de datos asigna más memoria a la transformación de ordenación porque esta suele tardar más en ejecutarse.

Asigna más memoria a la memoria caché de datos que a la memoria caché de índice.

Las transformaciones de agregación, combinación, búsqueda y rango requieren una memoria caché de índice y una memoria caché de datos. Cuando el Servicio de integración de datos divide la memoria asignada para la transformación entre la memoria caché de índice y la de datos, asigna más memoria a la memoria caché de datos.

Las transformaciones de ordenación requieren una única memoria caché. El servicio asigna toda la memoria asignada para la transformación a la memoria caché de ordenación.

Memoria máxima para tamaño de la memoria caché automático

Defina la cantidad máxima de memoria que el Servicio de integración de datos puede asignar a las memorias caché de las transformaciones en la propiedad Memoria máxima por solicitud para los módulos del Servicio de integración de datos en la Herramienta del administrador.

Cada módulo ejecuta diferentes tipos de solicitudes que tienen distintos requisitos de memoria. Por ejemplo, las solicitudes de asignación y de perfil normalmente requieren más memoria caché que las solicitudes de servicio SQL o servicio web. Puede configurar la propiedad Memoria máxima por solicitud para los siguientes módulos del Servicio de integración de datos:

- Módulo del servicio de asignación
- Módulo del servicio de creación de perfiles
- Módulo del servicio de SQL
- Módulo de servicios web

Nota: Las solicitudes del Módulo del servicio de asignación incluyen asignaciones y asignaciones ejecutadas desde tareas de asignación dentro de un flujo de trabajo.

Para el Módulo del servicio de creación de perfiles, Memoria máxima por solicitud define la cantidad máxima de memoria que el Servicio de integración de datos puede asignar para cada ejecución de asignación para una única solicitud de perfil.

Para los módulos restantes, el comportamiento de Memoria máxima por solicitud depende de la configuración del Servicio de integración de datos. El comportamiento depende de las propiedades Iniciar opciones de trabajos y Tamaño máximo de memoria del Servicio de integración de datos.

En la siguiente tabla se describe el comportamiento de Memoria máxima por solicitud para los módulos de asignación, servicio de SQL y servicios web basados en la configuración del Servicio de integración de datos:

Configuración del Servicio de integración de datos	Comportamiento de Memoria máxima por solicitud
Ejecuta tareas en procesos independientes del sistema local o remoto, o bien Tamaño máximo de memoria es 0 (predeterminado)	<p>Cantidad máxima de memoria, en bytes, que el Servicio de integración de datos puede asignar para todas las transformaciones que utilizan el modo de memoria caché automática en una única solicitud.</p> <p>El valor que defina para Memoria máxima por solicitud solo afectará a las transformaciones que utilicen el modo de memoria caché automática. El Servicio de integración de datos asigna memoria por separado a las transformaciones para las que configure un tamaño de memoria caché específico. La memoria total utilizada por la solicitud puede superar el valor de Memoria máxima por solicitud.</p> <p>Por ejemplo, Memoria máxima por solicitud está establecida en 800 MB. Una asignación tiene tres transformaciones que requieren almacenamiento en caché. Puede configurar dos transformaciones para que utilicen el modo de memoria caché automática y configurar la tercera transformación para que utilice un total de 500 MB para los tamaños de las memorias caché. El Servicio de integración de datos asigna un total de 1 300 MB de memoria para todas las memorias cache de las transformaciones.</p>
Ejecuta tareas en el proceso del Servicio de integración de datos y Tamaño máximo de memoria es mayor que 0	<p>Cantidad máxima de memoria, en bytes, que el Servicio de integración de datos puede asignar para una única solicitud.</p> <p>El valor que defina para la propiedad Memoria máxima por solicitud afecta a todas las transformaciones. La memoria total utilizada por la solicitud no puede superar el valor de Memoria máxima por solicitud.</p>

Al aumentar la cantidad máxima de memoria utilizada para el modo de memoria caché automática, incrementa el tamaño máximo de la memoria caché que se puede utilizar para todas las solicitudes que recibe el módulo. Puede aumentar la cantidad máxima de memoria para asegurarse de que ningún archivo de memoria caché se pague en el disco. Sin embargo, como este valor se utiliza para todas las solicitudes, el Servicio de integración de datos podría asignar más memoria de la necesaria para algunas solicitudes.

Tamaño de la memoria caché específico

Puede configurar un tamaño de memoria caché específico para una transformación. El Servicio de integración de datos asigna la cantidad de memoria especificada a la memoria caché de la transformación al inicio de la ejecución de la asignación. Configure un valor específico en bytes al ajustar el tamaño de la memoria caché.

La primera vez que configure un tamaño de memoria caché, utilice el modo de memoria caché automática. Después de ejecutar la asignación, analice las estadísticas de transformación en el registro de la asignación para determinar los tamaños de las memorias caché necesarios para ejecutar las transformaciones en la memoria. Al configurar el tamaño de memoria caché para que use el valor especificado en el registro de la asignación, se asegurará de que no se desperdicie nada de la memoria asignada. Sin embargo, el tamaño de memoria caché óptimo varía según el tamaño de los datos de origen. Revise los registros de la asignación después de ejecutar la asignación posteriormente para supervisar los cambios en el tamaño de la memoria caché. Si configura un tamaño de memoria caché específico para una transformación reutilizable, compruebe que el tamaño de la memoria caché sea óptimo para cada uso de la transformación en una asignación.

Para definir tamaños de memoria caché específicos, configure los valores de memoria caché en las propiedades de la transformación en Developer tool.

Aumento del tamaño de la memoria caché por el Servicio de integración de datos

El Servicio de integración de datos crea cada memoria caché según el tamaño de la memoria caché configurado. En algunos casos, el Servicio de integración de datos podría aumentar el tamaño de la memoria caché configurado si se necesita más memoria caché.

El Servicio de integración de datos podría aumentar el tamaño de la memoria caché configurado por uno de estos motivos:

El tamaño de la memoria caché configurado es menor que el tamaño de la memoria caché mínimo necesario para procesar la operación.

El Servicio de integración de datos requiere una cantidad mínima de memoria para inicializar cada asignación. Si el tamaño de la memoria caché configurado es menor que el tamaño de la memoria caché mínimo necesario, el Servicio de integración de datos aumenta el tamaño de la memoria caché configurado para satisfacer los requisitos mínimos. Si el Servicio de integración de datos no puede asignar la cantidad mínima de memoria necesaria, la asignación no se llevará a cabo.

El tamaño de la memoria caché configurado no es un múltiplo del tamaño de página de la memoria caché.

El Servicio de integración de datos almacena los datos de memoria caché en páginas de memoria caché. Las páginas de memoria caché deben caber de forma uniforme en la memoria caché. Por ejemplo, si configura 10 MB (1.048.576 bytes) como el tamaño de la memoria caché y el tamaño de la página de memoria caché es de 10.000 bytes, el Servicio de integración de datos aumenta el tamaño de la memoria caché configurado hasta 1.050.000 bytes para que sea un múltiplo del tamaño de página de 10.000 bytes.

Cuando el Servicio de integración de datos aumenta el tamaño de la memoria caché configurado, continúa ejecutando la asignación y escribe los siguientes mensajes en el registro de la asignación:

```
INFO: MAPPING,      TE_7212,      Increasing [Index Cache] size for transformation
<transformation name> from <configured cache size> to <new cache size>.
INFO: MAPPING,      TE_7212,      Increasing [Data Cache] size for transformation
<transformation name> from <configured cache size> to <new cache size>.
```

Tamaño de la memoria caché para memorias caché con particiones

Si tiene la opción Partición, crear particiones de la memoria caché crea una memoria caché independiente para cada partición que ejecute una transformación de agregación, combinación, rango, búsqueda u ordenación. Durante la partición de la memoria caché, cada partición almacena datos diferentes en una memoria caché distinta. Cuando el Servicio de integración de datos utiliza particiones de la memoria caché para estas transformaciones, divide el tamaño de la memoria caché asignado entre las particiones.

Por ejemplo, configura el tamaño de la memoria caché de la transformación en 100 MB. El Servicio de integración de datos usa cuatro particiones para ejecutar la transformación. El servicio divide el valor de tamaño de la memoria caché de modo que cada partición utiliza un máximo de 25 MB para el tamaño de la memoria caché.

Optimización del tamaño de la memoria caché

Para un rendimiento óptimo de la asignación, configure los tamaños de memoria caché de modo que el Servicio de integración de datos pueda ejecutar la transformación completa en la memoria.

Para configurar los tamaños de memoria caché óptimos, realice las siguientes tareas:

1. Establezca el nivel de seguimiento en inicialización detallada.
2. Ejecute la asignación en modo de memoria caché automática.
3. Analice el rendimiento del almacenamiento en la memoria caché en el registro de la asignación.
4. Configure valores específicos para los tamaños de memoria caché.

Paso 1. Establezca el nivel de seguimiento en inicialización detallada

En Developer tool, establezca el nivel de seguimiento en inicialización detallada para permitir que el Servicio de integración de datos escriba estadísticas de la transformación en el registro de la asignación. Las estadísticas de la transformación enumeran los tamaños de memoria caché necesarios para obtener un rendimiento óptimo. De forma predeterminada, el nivel de seguimiento es normal.

Establezca el nivel de seguimiento en inicialización detallada de una de las siguientes maneras:

- Modifique las propiedades avanzadas para cada transformación que utilice una memoria caché.
- Modifique las propiedades de configuración de la asignación predeterminadas si va a ejecutar la asignación por primera vez desde Developer tool. Para obtener más información, consulte la *Guía de Informática Developer tool*.
- Modifique las propiedades avanzadas para una aplicación que contenga la asignación si va a ejecutar la asignación implementada por primera vez desde la línea de comandos. Para obtener más información, consulte la *Guía de Informática Developer tool*.

Paso 2. Ejecute la asignación en modo de memoria caché automática

La primera vez que ejecute la asignación, utilice el modo de memoria caché automática para los tamaños de memoria caché de la transformación.

Puede ejecutar la asignación desde Developer tool. O bien, puede agregar la asignación a una aplicación y, a continuación, implementar la aplicación en el Servicio de integración de datos de modo que pueda ejecutar la asignación desde la línea de comandos.

Paso 3. Analice el rendimiento del almacenamiento en la memoria caché

Después de ejecutar la asignación en modo de memoria caché automática, analice las estadísticas de la transformación en el registro de la asignación para determinar los tamaños de memoria caché necesarios para obtener un rendimiento de asignación óptimo.

Cuando una transformación de agregación, combinación, búsqueda o rango se paginan en el disco, el registro de la asignación especifica el tamaño de las memorias caché de índice y de datos necesario para

ejecutar la transformación en la memoria. Por ejemplo, ejecuta una transformación de agregación llamada AGG_TRANS. El registro de la asignación contiene el siguiente texto:

```
CMN 1791, The index cache size that would hold [1098] aggregate groups of input rows for [AGG_TRANS], in memory, is [286720] bytes
CMN 1790, The data cache size that would hold [1098] aggregate groups of input rows for [AGG_TRANS], in memory, is [1774368] bytes
```

El registro muestra que la memoria caché de índice requiere 286 720 bytes y que la memoria caché de datos requiere 1 774 368 bytes para ejecutar la transformación en la memoria sin paginar en el disco.

Cuando una transformación de ordenación pagina en el disco, el registro de la asignación indica que el Servicio de integración de datos efectuó varios pases por los datos de origen. El Servicio de integración de datos realiza varias pases por los datos cuando debe paginar en el disco para completar la ordenación. El mensaje especifica el número de bytes necesarios para un solo pase, que es cuando el Servicio de integración de datos lee los datos una vez y realiza la ordenación en la memoria sin paginar en el disco.

Por ejemplo, ejecuta una transformación de ordenación llamada SRT_TRANS. El registro de la asignación contiene el siguiente texto:

```
SORT 40427, Sorter Transformation [SRT_TRANS] required 2-pass sort (1-pass temp I/O: 13126221824 bytes). You may try to set the cache size to 14128 MB or higher for 1-pass in-memory sort.
```

El registro muestra que la memoria caché de ordenación requiere 14,128 MB, de modo que el Servicio de integración de datos realice un pase por los datos.

Paso 4. Configure tamaños de memoria caché específicos

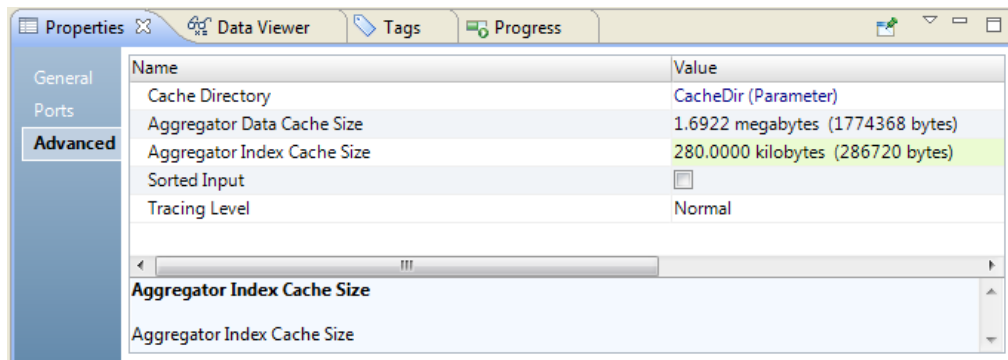
Para obtener un rendimiento óptimo, configure los tamaños de memoria caché para que usen los valores especificados en el registro de la asignación. Actualice las propiedades de la transformación de tamaño de la memoria caché de índice y de datos en Developer tool

1. En Developer tool, abra la transformación reutilizable o no reutilizable.
2. Localice las propiedades de tamaño de la memoria caché según los siguientes tipos de transformación:

Opción	Descripción
Transformación reutilizable de agregación, combinación, rango u ordenación	Haga clic en la vista Opciones avanzadas .
Transformación no reutilizable de agregación, combinación, rango u ordenación	Haga clic en la ficha Avanzado en la vista Propiedades .
Transformación reutilizable de búsqueda	Haga clic en la vista Tiempo de ejecución .
Transformación no reutilizable de búsqueda	Haga clic en la ficha Tiempo de ejecución en la vista Propiedades .

3. Introduzca los valores en bytes que ha recomendado el registro de la asignación para los tamaños de memoria caché de índice y de datos.

La siguiente imagen muestra una transformación de agregación no reutilizable con valores específicos configurados para el tamaño de las memorias caché de índice y de datos:



4. Haga clic en **Archivo > Guardar**.

CAPÍTULO 4

Transformación del validador de direcciones

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación del validador de direcciones, 82](#)
- [Datos de referencia de direcciones, 82](#)
- [Modos y plantillas, 85](#)
- [Grupos de puertos y selección de puertos, 85](#)
- [Grupos de puertos de entrada de la transformación del validador de direcciones, 85](#)
- [Grupos de puertos de salida de transformación del validador de direcciones, 86](#)
- [Puertos de varias instancias, 89](#)
- [Proyectos de validación de direcciones, 90](#)
- [Direcciones con formato y estándares del proveedor de servicios de correo, 92](#)
- [Finalización de direcciones parciales , 93](#)
- [Puertos de estado del validador de direcciones, 93](#)
- [Configuración general de la transformación del validador de direcciones, 106](#)
- [Propiedades de validación de direcciones en la ventana Preferencias, 107](#)
- [Propiedades avanzadas de validación de direcciones, 110](#)
- [Informes de certificación, 131](#)
- [Cómo configurar una transformación del validador de direcciones., 134](#)
- [Cómo añadir puertos a la transformación del validador de direcciones, 135](#)
- [Cómo crear plantillas definidas por el usuario, 135](#)
- [Cómo definir modelos del validador de direcciones, 136](#)
- [Cómo definir un informe de certificación, 136](#)
- [Transformación del validador de direcciones en un entorno no nativo, 137](#)

Resumen de la transformación del validador de direcciones

La transformación del validador de direcciones es una transformación de varios grupos que compara datos de direcciones de entrada con datos de referencia de direcciones. Determina la exactitud de las direcciones y corrige los errores en estas. La transformación crea registros que cumplen los estándares de envío de correo para la estructura y el contenido de datos. La transformación también añade información de estado a cada dirección.

La transformación del validador de direcciones realiza las siguientes operaciones en datos de direcciones:

- La transformación compara los registros de direcciones en los datos de origen con las definiciones de direcciones en los datos de referencia de direcciones.
- Genera informes de estado detallados sobre la validez de cada dirección de entrada, su estado de entrega y la naturaleza de los errores o ambigüedades que contiene.
- Corrige errores y completa registros de direcciones parciales. Para corregir una dirección, la transformación debe encontrar una coincidencia positiva con una dirección de los datos de referencia. La transformación copia los elementos de datos necesarios de los datos de referencia de direcciones en los registros de direcciones.
- Añade información que no aparece en la dirección estándar pero que ayuda a la entrega postal, como la información de punto de entrega y la información de geocodificación.
- Escribe las direcciones de salida en el formato que el proyecto de datos y el proveedor de servicios de correo requieran. Defina el formato cuando seleccione los puertos de salida en la transformación.

Datos de referencia de direcciones

Un conjunto de datos de referencia de direcciones describe las direcciones que un proveedor de servicios de correo nacional reconoce en un país. Antes de realizar la validación de direcciones con la transformación del validador de direcciones, debe instalar los datos de referencia de direcciones en el equipo de servicios de Informática en el dominio. Adquiera y descargue los datos de referencia de direcciones de Informática.

Instale un archivo de datos de referencia de direcciones para cada país que los datos de direcciones de origen identifiquen. Un país que tiene una gran población puede necesitar varios archivos. Además, puede instalar archivos de datos que complementan o enriquecen los datos de direcciones. El proveedor de servicios de correo puede utilizar los datos enriquecidos para certificar la exactitud de las direcciones y para acelerar la entrega de correo.

Al realizar la validación de direcciones, la transformación del validador de direcciones compara cada registro de entrada con los datos de referencia de direcciones. Si la transformación encuentra la dirección de entrada en los datos de referencia de direcciones, la transformación puede actualizar el registro con los datos de direcciones correctos y completos. Si adquirió conjuntos de datos de referencia adicionales, la transformación también puede enriquecer los datos de dirección.

Utilice la ventana **Preferencias** en Developer Tool para ver información sobre los archivos de datos de referencia de direcciones en el equipo de servicios de Informática en el dominio.

Tipos de datos de referencia de direcciones

El modo de validación que usted selecciona determina cómo compara la transformación la dirección de entrada con los datos de referencia de direcciones.

La transformación del validador de direcciones puede leer los siguientes tipos de datos de referencia de direcciones:

Datos de búsqueda de código de dirección

Instale datos de búsqueda del código de dirección para recuperar una dirección parcial o completa de un valor de código en un puerto de entrada. La integridad de la dirección depende del nivel de compatibilidad del código de dirección en el país al que pertenece la dirección. Para leer el código de dirección de una dirección de entrada, seleccione los puertos específicos del país en el grupo de puertos Discreto.

Puede seleccionar puertos para los siguientes países:

- Austria. Devuelve una dirección a nivel de edificio.
- Alemania. Devuelve una dirección a nivel de localidad, municipio o calle.
- Japón. Devuelve una dirección a nivel de buzón único.
- Sudáfrica. Devuelve una dirección a nivel de calle.
- Corea del sur. Devuelve una dirección a nivel de buzón único.
- Serbia. Devuelve una dirección a nivel de calle.
- Reino Unido. Devuelve una dirección a nivel de buzón único.

La transformación del validador de direcciones lee datos de búsqueda del código de dirección cuando configura la transformación para que se ejecute en modo de búsqueda del código de dirección.

Datos interactivos y por lotes

Instale datos por lotes e interactivos para realizar la validación de direcciones en un conjunto de registros de direcciones. Utilice los datos por lotes e interactivos para comprobar que las direcciones de entrada son completamente válidas para la entrega y están completas basándose en los datos postales actuales del proveedor de servicios de correo nacional.

Cuando configure la transformación para que se ejecute en modo por lotes, la transformación del validador de direcciones devolverá una única dirección por cada dirección de entrada. Cuando configure la transformación para que se ejecute en modo interactivo, la transformación del validador de direcciones devolverá una o varias direcciones por cada dirección de entrada.

Datos de CAMEO

Instale datos de CAMEO para añadir datos de segmentación de clientes en registros de direcciones residenciales. Los datos de segmentación de clientes indican el nivel de renta y las preferencias de estilo de vida probables de los residentes en cada dirección.

La transformación del validador de direcciones lee datos de CAMEO cuando configura la transformación para que se ejecute en el modo por lotes o en el modo de certificación.

Datos de certificación

Instale datos de certificación para comprobar que los registros de direcciones cumplen los estándares de certificación que un proveedor de servicios de correo define. Una dirección cumple un estándar de certificación si contiene elementos de datos que pueden identificar un buzón único, como los elementos de datos de punto de entrega. Si una dirección cumple un estándar de certificación, el proveedor de servicios de correo cobra una tarifa de entrega reducida.

Los siguientes países definen estándares de certificación:

- Australia. Certifica el correo de acuerdo con el estándar del sistema de aprobación de coincidencia de direcciones (AMAS).
- Canadá. Certifica el correo de acuerdo con el estándar del programa evaluación y reconocimiento de software (SERP).
- Francia. Certifica el correo de acuerdo con el estándar del servicio de administración de direcciones nacional (SNA).
- Nueva Zelanda. Certifica el correo de acuerdo con el estándar SendRight.
- Estados Unidos. Certifica el correo de acuerdo con el estándar del sistema de soporte de precisión de codificación (CASS).

La transformación del validador de direcciones lee datos certificados cuando configura la transformación para que se ejecute en el modo de certificación.

Datos de geocodificación

Instale datos de geocodificación para añadir geocodificación a los registros de direcciones. Los datos de geocodificación son las coordenadas de latitud y longitud.

La transformación del validador de direcciones lee datos de geocodificación cuando configura la transformación para que se ejecute en el modo por lotes o en el modo de certificación.

Datos de la lista de sugerencias

Instale datos de lista de sugerencias para buscar versiones válidas alternativas de un registro de direcciones parciales. Use los datos de la lista de sugerencias cuando configure una asignación de validación de direcciones para procesar registros de direcciones de uno en uno en tiempo real. La transformación del validador de direcciones usa los elementos de datos de la dirección parcial para realizar una comprobación de duplicados en los datos de la lista de sugerencias. La transformación devolverá cualquier dirección válida que incluya la información en la dirección parcial.

La transformación del validador de direcciones lee datos de la lista de sugerencias cuando configura la transformación para que se ejecute en el modo de lista de sugerencias.

Datos complementarios

Instale datos complementarios para añadir datos a un registro de direcciones que faciliten la entrega de correo al proveedor de servicios de correo. Utilice los datos complementarios para añadir detalles sobre la zona geográfica o postal que contiene la dirección. En algunos países, los datos complementarios pueden proporcionar un identificador único para un buzón en el sistema postal.

La transformación del validador de direcciones lee datos complementarios cuando configura la transformación para que se ejecute en el modo por lotes o en el modo de certificación.

Nota: La transformación no lee datos de referencia de direcciones en el modo de reconocimiento del país o en el modo de análisis.

TEMAS RELACIONADOS

- [“ Configuración general de la transformación del validador de direcciones” en la página 106](#)

Modos y plantillas

Cuando configure la transformación del validador de direcciones, seleccione el tipo de validación que realizará la transformación. La transformación define el tipo de validación como el modo. Seleccione el modo en la ficha **Configuración general** o como una propiedad avanzada de la transformación.

Puede crear una plantilla de puertos en la transformación. Una plantilla es un subconjunto de puertos de uno o varios grupos de puertos. Utilice una plantilla para organizar los puertos que va a utilizar en un proyecto.

Grupos de puertos y selección de puertos

La transformación del validador de direcciones contiene grupos de puertos predefinidos que contienen los puertos de entrada y de salida que se pueden utilizar. Cuando configure una transformación del validador de direcciones, busque los grupos y seleccione los puertos que necesite.

Seleccione los puertos de entrada que correspondan a la estructura de los datos de entrada de las direcciones. Seleccione los puertos de salida que contengan los datos de direcciones que requiera el proyecto.

Puede añadir directamente a la transformación los puertos de entrada y salida, o bien puede crear un modelo predeterminado que contenga puertos de entrada y salida. Si añade los puertos directamente a la transformación, los puertos que seleccione se aplicarán únicamente a dicha transformación. Si añade los puertos al modelo predeterminado, los puertos que seleccione se aplicarán a las futuras transformaciones del validador de direcciones que cree.

Además, también puede añadir a la transformación puertos de transferencia para las columnas que no desee que procese la transformación del validador de direcciones.

Grupos de puertos de entrada de la transformación del validador de direcciones

Antes de conectar los datos de direcciones con los puertos de entrada de la transformación, busque los grupos de entrada y seleccione los puertos que se corresponden con la estructura y el contenido de los datos de entrada. Busque los grupos de salida y seleccione los puertos que cumplan sus requisitos de datos.

La transformación del validador de direcciones muestra los grupos de puertos en un modelo básico y en un modelo avanzado. Puede definir la mayoría de direcciones utilizando un grupo de puertos en el modelo básico. Si las direcciones son complejas, utilice los puertos adicionales que están disponibles en el modelo avanzado.

Nota: Seleccione puertos solamente de un grupo de puertos de entrada.

La transformación tiene los siguientes grupos de puerto de entrada:

Discreto

Utilice los puertos Discreto para leer columnas de datos que contengan información completa sobre un único elemento de datos, como un número de casa, una calle o un código postal. Busque el grupo Discreto en los modelos básico y avanzado.

Híbrido

Utilice los puertos Híbrido para leer columnas de datos que contengan información sobre uno o varios elementos de datos. El grupo híbrido combina puertos de los grupos discreto y de varias líneas. Use puertos híbridos para crear registros de direcciones que puede enviar a un proveedor de servicios de correo. Los puertos híbridos permiten estructurar una dirección según los estándares del proveedor de servicios de correo e identificar los tipos de datos de cada línea. Busque el grupo Híbrido en los modelos básico y avanzado.

Varias líneas

Utilice los puertos Varias líneas para leer columnas de datos que contengan varios elementos de datos. Cada columna de entrada corresponde a una línea de una dirección. Para obtener mejores resultados, defina los datos de entrada en el formato que requiera el proveedor de servicios de correo. Seleccione los puertos Varias líneas para crear un conjunto imprimible de registros de direcciones.

Cada puerto Varias líneas representa una línea de la dirección impresa, como la siguiente línea de datos de calle:

```
123 Main Street, Apartment 2
```

Los puertos de varias líneas no especifican el tipo de datos que se muestra en cada línea de dirección. Busque el grupo Varias líneas en los modelos básico y avanzado.

Una sola línea

Utilice los puertos Una sola línea para leer una única columna de datos que contenga todos los elementos de dirección a nivel de provincia y que no incluya un separador entre los elementos. Utilice el puerto Dirección completa del grupo de puertos para enviar los elementos de dirección. El grupo de puertos también incluye un puerto País que puede utilizar para leer la información relacionada con el país de la dirección. Busque el grupo Una sola línea en los modelos básico y avanzado.

Grupos de puertos de salida de transformación del validador de direcciones

Antes de conectar la transformación del validador de direcciones con otras transformaciones u objetos de datos, determine los tipos de información que necesite y la estructura que tendrán las direcciones de salida.

Busque los grupos de salida y seleccione los puertos que cumplan sus requisitos de datos.

Nota: Puede seleccionar puertos de varios grupos de salida y puede seleccionar puertos que tienen funciones comunes.

La transformación tiene los siguientes grupos de salida predefinidos:

Elementos de dirección

Escribe elementos de datos de dirección, como el número de domicilio, el número de apartamento y el nombre de calle en puertos independientes. Busque el grupo de elementos de dirección en los modelos básico y avanzado.

Complementario de AT

Escribe datos en las direcciones de Austria que facilitan la entrega postal, como datos de código postal a nivel de edificio. Busque el grupo complementario de AT en el modelo básico.

Complementario de AU

Escribe datos en direcciones de Australia que identifican las regiones geográficas a las que la Australian Bureau of Statistics asigna las direcciones. Busque el grupo complementario de AU en el modelo básico.

Específico de Australia

Escribe datos en las direcciones de Australia que permiten que las direcciones cumplan los estándares del Sistema de aprobación de coincidencia de direcciones (AMAS) del servicio postal de Australia. Busque el grupo específico de Australia en los modelos básico y avanzado.

Complementario de BE

Escribe datos en direcciones de Bélgica que facilitan la entrega postal. Los datos incluyen códigos de identificación de localidades y barrios del Instituto Nacional de Estadísticas de Bélgica. Busque el grupo complementario de BE en el modelo básico.

Complementario de BR

Escribe datos en las direcciones de Brasil que facilitan la entrega postal, como códigos de identificación de distrito del Instituto de Geografía y Estadística (IBGE). Busque el grupo complementario de BR en el modelo básico.

CAMEO

Genera datos de resumen demográficos y de ingresos que puede usar en los análisis de segmentación de clientes. Busque el grupo CAMEO en el modelo básico.

Específico de Canadá

Escribe datos en direcciones de Canadá que permiten adaptar las direcciones a los estándares del Programa de evaluación y reconocimiento de software (SERP) del servicio postal de Canadá. Busque el grupo específico de Canadá en el modelo básico.

Complementario de CH

Escribe datos en las direcciones de Suiza que facilitan la entrega postal, como datos de código postal ampliados. Busque el grupo complementario de CH en el modelo básico.

Complementario de CZ

Escribe datos en las direcciones de la República Checa que facilitan la entrega postal, como datos de código postal ampliados. Busque el grupo complementario de CZ en el modelo básico.

Elementos de contacto

Escribe datos personales o de contacto, como nombres, tratamientos y puestos. Busque el grupo de elementos de contacto en el modelo avanzado.

País

Escribe el nombre o el código de país definido por la Organización Internacional de Normalización (ISO). Busque el grupo de país en los modelos básico y avanzado.

Complementario de DE

Escribe datos en las direcciones de Alemania que facilitan la entrega postal, como datos de código de municipio y distrito. Busque el grupo complementario de DE en el modelo básico.

Complementario de ES

Escribe datos en las direcciones de España que facilitan la entrega postal. Busque el grupo complementario de ES en el modelo básico.

Línea de dirección formateada

Escribe direcciones con el formato correspondiente para la impresión y el envío de correo. Busque el grupo de línea de dirección formateada en los modelos básico y avanzado.

Complementario de FR

Escribe datos en las direcciones de Francia que facilitan la entrega postal, como códigos de identificación del Instituto Nacional de Estadística y Estudios Económicos (INSEE). Busque el grupo complementario de FR en el modelo básico.

Específico de Francia

Escribe datos en direcciones de Francia que permiten que las direcciones cumplan con los estándares del Servicio nacional de administración de direcciones (SNA) del servicio postal de Francia. Busque el grupo específico de Francia en el modelo básico.

Geocodificación

Genera datos de geocodificación, como las coordenadas de latitud y longitud, para una dirección. Busque el grupo de geocodificación en el modelo básico.

Elementos de ID

Escribe datos de ID de registro y clave de transacción. Busque el grupo de elementos de ID en el modelo avanzado.

IT complementario

Escribe datos en direcciones en Italia que facilitan la entrega postal. Busque el grupo complementario de IT en el modelo básico.

complementario de JP

Escribe datos en las direcciones de Japón que facilitan la entrega postal, como los códigos Choumei Aza. Busque el grupo complementario de JP en el modelo básico.

Complementario para KR

Escribe datos en las direcciones de Corea del Sur que facilitan la entrega postal, como los identificadores únicos que pueden especificar la versión actual y la antigua de una determinada dirección. Busque el grupo complementario de KR en el modelo básico.

Elementos de última línea

Escribe datos que pueden aparecer en la última línea de una dirección nacional. Busque el grupo de elementos de última línea en los modelos básico y avanzado.

Específicos de Nueva Zelanda

Escribe datos en direcciones de Nueva Zelanda que permiten que las direcciones cumplan con los estándares SendRight del servicio postal de Nueva Zelanda. Busque el grupo específico de Nueva Zelanda en el modelo básico.

Complementario de PL

Escribe datos en las direcciones de Polonia que facilitan la entrega postal, como datos de División territorial (TERYT). Busque el grupo complementario de PL en el modelo básico.

Residuos

Escribe elementos de datos que la transformación no puede analizar para otros puertos. Busque el grupo de residuos en los modelos básico y avanzado.

Complementario de RS

Escribe datos en las direcciones de Serbia que facilitan la entrega postal, como datos de sufijo de código postal. Busque el grupo Complementario de RS en el modelo básico.

Complementario de RU

Escribe datos en las direcciones de Rusia que facilitan la entrega postal, como el identificador del Sistema Federal de Información de Direcciones para la dirección. Busque el grupo Complementario de RU en el modelo básico.

Información de estado

Genera datos detallados sobre la calidad de cada dirección de entrada y salida. Busque el grupo de información de estado en el modelo básico.

Complementario del Reino Unido

Escribe datos en direcciones del Reino Unido que facilitan la entrega postal, como datos del punto de entrega y datos de Ordnance Survey. Busque el grupo Complementario del Reino Unido en el modelo básico.

Específico de EE. UU.

Escribe datos en las direcciones de Estados Unidos que permiten que las direcciones cumplan los estándares del Sistema de apoyo para la exactitud de codificación (CASS) del servicio postal de Estados Unidos. Busque el grupo específico de EE.UU. en el modelo básico.

Complementario de EE.UU.

Escribe datos geográficos y demográficos, como los códigos FIPS (Federal Information Processing Standards) para las direcciones de Estados Unidos. Busque el grupo Complementario de EE.UU. en el modelo básico.

XML

Escribe datos de registros de direcciones con una estructura XML definida por la biblioteca de software de Address Verification. Busque el grupo XML en el modelo avanzado.

Complementario de ZA

Escribe datos en direcciones de Sudáfrica que facilitan la entrega postal, como datos de la Base de datos Nacional de Direcciones. Busque el grupo complementario de ZA en el modelo básico.

Puertos de varias instancias

Muchos tipos de datos de direcciones se pueden incluir varias veces en una dirección. Puede seleccionar varias instancias de un puerto si la dirección contiene varios formatos de mayúsculas y minúsculas de un elemento de datos.

Un puerto de varias instancias puede contener hasta seis instancias. Muchas direcciones usan una única instancia de un puerto para cada elemento de datos que contienen. Algunas direcciones usan una segunda instancia de un puerto. Un conjunto de direcciones reducido usa varias instancias de un puerto.

A menudo, la primera instancia de un puerto es el nombre principal o el área más amplia que el puerto identifica. Debe comprobar la relación entre las instancias de puerto de cualquier puerto que seleccione.

Ejemplo de puertos Calle completa

Un registro de direcciones del Reino Unido puede contener dos nombres de calle si uno de ellos forma parte de un esquema de calle más amplio.

En la siguiente tabla se muestra una dirección que utiliza dos puertos Calle completa:

Puerto	Datos
Número de calle completo 1	1A
Calle completa 1	THE PHYGTLE
Calle completa 2	SOUTH STREET
Nombre de localidad 1	NORFOLK
Código postal 1	NR25 7QE

En el ejemplo que se muestra más arriba, los datos de calle del puerto Calle completa 1 dependen de los datos de calle del puerto Calle completa 2. Los datos del número de calle completo 1 hacen referencia a los datos de la calle completa 1.

Nota: Aunque el puerto Calle completa 1 especifica la ubicación del buzón, es posible que el puerto Calle completa 2 corresponda a la calle más grande.

Ejemplos de puertos Contacto

Un registro de direcciones puede contener varios contactos si cada uno forma parte de la misma familia.

En la siguiente tabla se muestra una dirección que utiliza dos puertos Nombre de contacto:

Puerto	Datos
Nombre de contacto 1	SR. JOHN DOE
Nombre del contacto 2	SRA. JANE DOE
Línea de dirección formateada 1	2 MCGRATH PLACE EAST
Línea de dirección formateada 2	ST. JOHN'S NL A1B 3V4
Línea de dirección formateada 3	CANADÁ

En el ejemplo que se muestra más arriba, la organización puede decidir la prioridad que se debe asignar a Nombre de contacto 1 o a Nombre de contacto 2. La transformación del validador de direcciones no establece la prioridad de los datos de contacto.

En caso de que decida aplicar un formato a las direcciones para imprimirlas, puede usar varias instancias de puertos Línea de dirección formateada. Puede seleccionar hasta 12 puertos Línea de dirección formateada.

Proyectos de validación de direcciones

Puede usar la transformación del validador de direcciones en diversos tipos de proyectos. Puede crear una plantilla de dirección con distintos puertos para cada tipo de proyecto.

Puede definir un proyecto de validación de direcciones con uno o varios de los siguientes objetivos:

Crear direcciones con formato conforme a los estándares del proveedor de servicios de correo

Puede preparar un conjunto de registros de direcciones de gran tamaño para una campaña de correo. Si crea las direcciones con el formato preferido por el proveedor de servicios de correo, los costos de correo se reducen considerablemente. Cuando prepare las direcciones para el envío de correo, seleccione los puertos de salida que escriben cada línea de la dirección con formato en un solo puerto. Puede seleccionar otro puerto para el nombre de contacto, las líneas de dirección y las líneas de localidad y código postal.

Organizar las direcciones por indicadores de renta y estilo de vida

Puede añadir datos de segmentación de clientes a los registros de direcciones particulares. Los datos de segmentación de clientes indican el nivel de renta y las preferencias de estilo de vida probables de los residentes en cada dirección. Seleccione puertos del grupo de salida CAMEO para añadir datos de segmentación de clientes a registros de direcciones. Puede utilizar datos de segmentación de clientes en campañas de correo dirigidas a varios mercados de consumidores.

Crear direcciones certificadas por el proveedor de servicios de correo

Si prepara un conjunto de registros para Australia Post, Canada Post o United States Postal Service (USPS), puede añadir datos para confirmar la garantía de entrega de cada dirección.

La transformación del validador de direcciones puede generar informes para certificar que los registros de direcciones son completos y exactos de conformidad con los estándares de cada proveedor de servicios de correo.

Crear direcciones que cumplan los requisitos para el cumplimiento de la normativa

Puede comprobar si los registros de direcciones de su organización son exactos de conformidad con la normativa gubernamental o del sector. Seleccione puertos de salida que escriban cada elemento de datos de dirección en un campo independiente. Además, seleccione los puertos de estado de validación de direcciones que proporcionan información detallada acerca de la exactitud e integridad de los datos de salida.

Completar direcciones parciales

Puede especificar una dirección parcial y recuperar las direcciones completas válidas que coincidan con la dirección parcial de los datos de referencia. Para completar las direcciones parciales, configure la transformación para que se ejecute en modo de lista de sugerencias o en modo interactivo. Puede especificar la dirección de entrada como una sola línea en el puerto de Dirección completa.

Mejorar la calidad de los datos de las direcciones

En conjunción con otros proyectos de datos, puede mejorar la estructura y la calidad general de los datos del conjunto de datos de direcciones. Por ejemplo, el conjunto de datos puede contener más columnas de las necesarias o puede contener el mismo tipo de datos en varias columnas. Puede reducir el número de columnas del conjunto de datos y simplificar las columnas usadas para los distintos tipos de datos.

Direcciones con formato y estándares del proveedor de servicios de correo

Si prepara registros de direcciones para una campaña de correo, debe crear una estructura de direcciones imprimible que coincida con los estándares de formato del proveedor de servicios de correo.

Por ejemplo, USPS mantiene el siguiente formato de dirección para las direcciones nacionales de Estados Unidos:

Line 1	Person/Contact Data	JOHN DOE
Line 2	Street Number, Street, Sub-Building	123 MAIN ST NW STE 12
Line 3	Locality, State, ZIP Code	ANYTOWN NY 12345

Puede definir un formato de dirección imprimible que escriba cada línea de la dirección en un solo puerto. Puede usar puertos que reconozcan los tipos de datos de cada línea o puertos que rellenen la estructura de direcciones con independencia de los datos de cada línea.

En la siguiente tabla, se muestran los distintos métodos para aplicar un formato a una dirección de Estados Unidos para su impresión:

Para esta dirección	Usar estos puertos	0 usar estos puertos
JOHN DOE	Línea de destinatario 1	Línea de dirección formateada 1
123 MAIN ST NW STE 12	Línea de dirección de entrega 1	Línea de dirección formateada 2
ANYTOWN NY 12345	Última línea específica de país 1	Línea de dirección formateada 3

Use los puertos de línea de dirección formateada si el conjunto de datos contiene distintos tipos de dirección, como direcciones comerciales o direcciones particulares. Es posible que para una dirección comercial se necesiten tres líneas de dirección para los datos de contacto y organización. La transformación del validador de direcciones garantiza la aplicación del formato correcto a cada dirección comercial o particular mediante el uso de puertos de línea de dirección formateada solamente si son necesarios. No obstante, los puertos de línea de dirección formateada no identifican el tipo de datos que contienen.

Use los puertos de línea de destinatario, línea de dirección de entrega y última línea específica de país si todas las direcciones tienen el mismo formato. Los puertos de línea de destinatario, línea de dirección de entrega y última línea específica de país separan los elementos de datos de direcciones según el tipo de información y facilitan la comprensión del conjunto de datos.

Nota: Puede seleccionar otros puertos para procesar esta dirección. En este ejemplo, se incluyen los puertos que permiten aplicar el formato correspondiente a las direcciones para la impresión y el envío.

Datos demográficos y geográficos

Al crear un conjunto de registros para una campaña de correo, puede añadir varios tipos de datos que, en caso contrario, podrían no aparecer en la dirección. Use estos datos para revisar la extensión demográfica y geográfica de los elementos de correo.

Por ejemplo, puede identificar el distrito electoral al que pertenece una dirección de Estados Unidos. Además, puede generar las coordenadas de latitud y longitud si el país de destino incluye estas coordenadas en los datos de referencia de su sistema de correo.

Finalización de direcciones parciales

Cuando utiliza el modo de lista de sugerencias o el modo interactivo, puede especificar una dirección incompleta para recuperar direcciones completas y válidas de los datos de referencia.

Seleccione el modo de lista de sugerencias cuando no esté seguro de la validez de una dirección y desee consultar una lista de posibles direcciones válidas. Seleccione el modo interactivo cuando esté seguro de la validez de una dirección y desee comprobar su forma completa. En cada caso, la transformación del validador de direcciones busca los datos de referencia de direcciones y devuelve todas las direcciones que contienen los datos de entrada.

Tenga en cuenta las siguientes reglas y directrices al configurar la transformación que se ejecutará en modo de lista de sugerencias o en modo interactivo:

- Puede definir una dirección de entrada en varios puertos o puede especificar todos los elementos de dirección en el puerto de entrada Dirección completa.
- Cuando configure la transformación en el modo de lista de sugerencias, seleccione puertos del grupo de entrada Discreto. Por otro lado, también puede seleccionar un puerto Dirección completa y, opcionalmente, un puerto Nombre de país del grupo de Varias líneas.
- El modo de lista de sugerencias y el modo interactivo pueden devolver varias direcciones para cada dirección de entrada. La propiedad Recuento máximo de resultados especifica el límite superior del número de direcciones devueltas. Si el número de coincidencias de dirección es mayor que el valor de Recuento máximo de resultados, el puerto Desbordamiento de recuento devolverá el número de direcciones adicionales.
- En Informatica Address Verification, el modo de lista de sugerencias se denomina modo de finalización rápida.

Puertos de estado del validador de direcciones

La transformación de validador de direcciones escribe información de estado sobre los elementos de dirección que lee y escribe en puertos de entrada y de salida. Utilice los puertos de información de estado para ver la información de estado.

Puede seleccionar los siguientes puertos de estado:

Código de resolución de direcciones

Describe los elementos de dirección no válidos en la dirección. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Tipo de dirección

Indica el tipo de dirección en casos en que el proveedor de servicios de correo reconoce más de una forma de una dirección. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Estado de entrada de elementos

Describe los niveles de similitud entre los elementos dirección de entrada y los datos de referencia. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Relevancia de elementos

Identifica los elementos de dirección que el proveedor de servicios de correo requiere para identificar un buzón para la dirección. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Estado de resultado de elementos

Describe cualquier actualización que la validación de direcciones realiza en la dirección de entrada. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Estado de resultado de elementos extendido

Indica la presencia de datos adicionales de una dirección en los datos de referencia. El puerto puede contener información detallada acerca de las actualizaciones que la validación de direcciones realiza en una dirección. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Estado de geocodificación

Describe el tipo de datos de geocodificación que la validación de direcciones devuelve para una dirección. Seleccione el puerto en el grupo de puertos Geocodificación del modelo básico.

Puntuación de viabilidad de envío de correo

Indica la probabilidad de que el proveedor de servicios de correo entregue un elemento de correo a la dirección. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Código de coincidencia

Describe los resultados de las operaciones de corrección de direcciones y validación de direcciones en la dirección de entrada. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Porcentaje de resultado

Representa el grado de similitud entre una dirección de entrada y la dirección de salida correspondiente como un valor de porcentaje. Seleccione el puerto en el grupo de puertos Información de estado del modelo básico.

Definiciones de código de estado de elemento

Los puertos de estado de entrada de elementos, relevancia de elementos, estado de resultado de elementos y estado de resultados de elementos extendidos proporcionan información de estado sobre la validez de los elementos de datos de entrada y salida. Seleccione los puertos de elementos para revisar el resultado de una operación de validación de direcciones.

Los códigos contienen la siguiente información:

- Los códigos de estado de entrada de elementos representan la calidad de la coincidencia encontrada entre los datos de dirección de entrada y los datos de referencia.
- Los códigos de relevancia de elementos identifican los elementos de dirección necesarios para la entrega en las direcciones en el país de destino.
- Los códigos de estado de resultado de elementos describen los cambios realizados en los datos de entrada durante el procesamiento.
- Los códigos de estado de resultado de elementos extendido indican que los datos de referencia de direcciones contienen información adicional acerca del elemento de dirección.

Cada puerto devuelve un código de 20 caracteres en el que cada carácter hace referencia a un elemento de datos de dirección distinto. Durante la lectura de los códigos de salida de los puertos de elementos, debe saber a qué elemento hace referencia cada carácter. Los 20 caracteres constan de 10 pares. Los dos

códigos de cada par representan un tipo de información de dirección. Por ejemplo, la primera posición del código de retorno representa información básica del código postal.

Nota: El puerto de código de resolución de dirección devuelve una cadena de 20 caracteres basada en los mismos elementos de dirección que los puertos de estado de elemento.

En la siguiente tabla, se describen los elementos de dirección que los valores en cada posición identifican:

Posición	Elemento de dirección	Descripción	Ejemplo de elemento de dirección
1	Nivel de código postal 0	Información básica del código postal, como un código postal de cinco dígitos.	Código postal de cinco dígitos 10118
2	Nivel de código postal 1	Información adicional del código postal, como los últimos cuatro dígitos de un código postal ZIP+4.	0110 en el código ZIP+4 10118-0110
3	Nivel de localidad 0	Ubicación principal, como una ciudad o un municipio.	Londres, en Inglaterra
4	Nivel de localidad 1	Localidad dependiente, como un barrio residencial o un pueblo.	Islington, en Londres
5	Provincia nivel 0	Región principal de un país, como un nombre de estado de Estados Unidos, un nombre de provincia de Canadá o un cantón de Suiza.	Estado de Nueva York
6	Nivel de provincia 1	Nombre de condado de Estados Unidos.	Condado de Queens, en el Estado de Nueva York
7	Nivel de calle 0	Información de calle principal.	South Great George's Street
8	Nivel de calle 1	Información de calle dependiente.	George's Arcade, en South Great George's Street
9	Nivel de número 0	Número de edificio o casa relacionado con la calle principal.	460, en South Great George's Street
10	Nivel de número 1	Número de edificio o casa relacionado con la calle dependiente.	81, en George's Arcade
11	Nivel de servicio de entrega 0	Descriptor y número de apartado de correos.	Apartado de correos 111
12	Nivel de servicio de entrega 1	Código de la oficina de correos responsable de la entrega.	MAIN STN
13	Nivel de edificio 0	Nombre o número de edificio. No identifica un número de domicilio.	Alice Tully Hall

Posición	Elemento de dirección	Descripción	Ejemplo de elemento de dirección
14	Nivel de edificio 1	Nombre o número de edificio adicional.	Starr Theater, en Alice Tully Hall
15	Nivel de edificio secundario 0	Nombre o número de apartamento, suite o planta.	80, en 350 5th Avenue, planta 80
16	Nivel de edificio secundario 1	Información del apartamento, suite o planta relacionada con la información de nivel de edificio secundario 0.	80-18 (18 es el número de estudio y 80 es el número de planta)
17	Nivel de organización 0	Nombre de compañía.	AddressDoctor® GmbH
18	Nivel de organización 1	Información corporativa adicional, como la empresa matriz.	Informatica Corporation
19	Nivel de país 0	Nombre de país.	Estados Unidos de América
20	Nivel de país 1	Territorio.	Islas Vírgenes de los Estados Unidos

Si un nombre de puerto incluye un sufijo de número, el nivel 0 hace referencia a los datos del número de puerto 1 y el nivel 1 hace referencia a los datos de los números de puerto de 2 a 6.

La información de nivel 0 se puede incluir antes o después de la información de nivel 1 en una dirección impresa. Por ejemplo, el nivel de código postal 1 se incluye después del nivel de código postal 0 y el nivel de localidad 1 se incluye antes del nivel de localidad 0.

Valores del puerto de salida del código de resolución de direcciones

El código de resolución de direcciones es una cadena de 20 caracteres en la que cada carácter de la cadena representa un elemento diferente de una dirección de entrada. El valor de un carácter describe cualquier elemento de dirección no válido en la posición correspondiente en la dirección.

La siguiente tabla describe los valores del puerto del código de resolución de direcciones:

Código	Descripción
2	El elemento de dirección es obligatorio para la entrega, pero no está presente en la dirección de entrada. Los datos de referencia de direcciones contienen el elemento de dirección que falta. Una salida de 2 indica que la dirección no es válida para la entrega sin el elemento de dirección.
3	El elemento de dirección es un número de domicilio o de calle que está fuera del intervalo válido de la dirección. Por ejemplo, el elemento de dirección contiene un número de domicilio que no existe en la calle especificada. El modo de lista de sugerencias devuelve direcciones alternativas.

Código	Descripción
4	La validación de direcciones no puede verificar ni corregir el elemento de dirección porque la dirección de entrada contiene más de una instancia del elemento.
5	El elemento de dirección es ambiguo en la dirección actual y los datos de referencia de direcciones contienen alternativas. La validación de direcciones copia el elemento de entrada en la dirección de salida. Por ejemplo, el elemento de dirección es un código postal válido que no coincide con una localidad válida en la dirección.
6	El elemento de dirección contradice otro elemento en la dirección. La validación de direcciones no puede determinar el elemento correcto de la dirección. La dirección de salida copia la dirección de entrada.
7	El elemento de dirección no se puede corregir sin realizar varios cambios en la dirección. La validación de direcciones puede corregir la dirección, pero el número de cambios indica que la dirección no es fiable.
8	Los datos no se ajustan a las reglas de validación del proveedor de servicios de correo.

Valores del puerto de salida del estado de entrada de elementos

El estado de entrada de elementos es una cadena de 20 caracteres en la que cada carácter representa un elemento diferente de una dirección de entrada. El valor de cada carácter representa el tipo de procesamiento realizado en el elemento de dirección.

Busque el puerto en el grupo de puertos de Información de estado.

En la siguiente tabla se describen los códigos que el estado de entrada de elementos puede devolver en cada posición de la cadena de salida en los modos por lotes, certificado o de lista de sugerencias:

Código	Descripción
0	La dirección de entrada no contiene datos en la posición actual.
1	Los datos de referencia no contienen datos en la posición actual.
2	No se pueden comprobar los datos porque faltan los datos de referencia.
3	Los datos en la posición actual son incorrectos. La base de datos de referencia sugiere que el número o el valor del servicio de entrega no está comprendido en el rango que esperan los datos de referencia. En el modo por lotes y el modo de certificación, la transformación transfiere los datos de entrada en la posición actual sin corregir como salida.
4	Los datos en la posición actual coinciden con los datos de referencia, pero contienen errores.
5	Los datos en la posición actual coinciden con los datos de referencia, pero la transformación ha corregido o estandarizado los datos.
6	Los datos en la posición actual coinciden con los datos de referencia y no contienen errores.

En la siguiente tabla se describen los códigos que el estado de entrada de elementos puede devolver en cada posición de la cadena de salida en el modo de análisis:

Código	Descripción
0	La dirección de entrada no contiene datos en la posición actual.
1	La transformación ha movido el elemento en la posición actual a otra posición de la dirección de salida.
2	El elemento en la posición actual coincide con el valor de datos de referencia, pero la transformación ha normalizado el elemento en la dirección de salida.
3	Los datos en la posición actual son correctos.

Valores del puerto de salida de relevancia de elementos

Los valores de relevancia de elementos indican si se necesita un elemento de dirección para la entrega postal. Busque el puerto en el grupo de puertos de Información de estado.

El valor del puerto de relevancia de elementos es una cadena de 20 caracteres en la que cada carácter puede representar un tipo distinto de datos de direcciones. Después de ejecutar la asignación de validación de direcciones, revise los datos de salida del puerto para identificar los elementos de dirección necesarios para cada dirección. Use los resultados para comprobar si ha seleccionado los puertos de salida correctos para los datos de direcciones. Si no selecciona ningún puerto de salida para un elemento de datos de direcciones relevante, la salida para esta dirección no será válida.

En la siguiente tabla se indican los códigos que el puerto de relevancia de elementos puede devolver en cada posición de la cadena de salida:

Código	Descripción
0	No es relevante para la entrega en la dirección.
1	Es relevante para la entrega en la dirección. El proveedor de servicios de correo nacional no puede realizar la entrega en la dirección sin los datos de esta posición en la cadena de salida.

Nota: Los valores de relevancia de elementos están disponibles para las direcciones con un valor de código de coincidencia de Cx o Vx en modo por lotes, o Cx, Vx, I3 o I4 en modo interactivo. Otros códigos de evaluación, como el Estado de entrada de elementos, el Estado de resultado de elementos, el Estado de resultado de elementos extendido y el Código de resolución de direcciones, devuelven valores independientemente del valor de código de coincidencia.

Valores del puerto de salida del estado de resultado de elementos

El estado de resultado de elementos es una cadena de 20 caracteres en la que cada carácter representa un elemento diferente de una dirección de entrada. El valor de cada carácter describe cualquier actualización que el proceso de validación hace en el elemento de dirección.

Busque el puerto en el grupo de puertos de Información de estado.

La siguiente tabla describe los valores del puerto de estado de resultado de elementos:

Código	Descripción
0	La dirección de salida no contiene datos en la posición actual.
1	La transformación no puede encontrar los datos de la posición actual en los datos de referencia. La transformación copia los datos de entrada en los datos de salida.
2	Los datos en la posición actual están desprotegidos, pero se han estandarizado.
3	Los datos en la posición actual están protegidos, pero no coinciden con los datos de referencia. Los datos de referencia sugieren que los datos de número no están comprendidos en el intervalo válido. La transformación copia los datos de entrada en el puerto de salida. Se aplica en el modo por lotes.
4	La transformación copia los datos de entrada en los datos de salida porque faltan los datos de referencia.
5	Los datos en la posición actual se han validado, pero no se han modificado debido a que existen varias coincidencias en los datos de referencia. Se aplica en el modo por lotes.
6	La validación de datos ha eliminado el valor de entrada en la posición actual.
7	Los datos en la posición actual se han validado, pero los datos de entrada tienen un error ortográfico. La validación ha corregido el error a partir de un valor de los datos de referencia.
8	Los datos en la posición actual se han validado y se han actualizado a partir de un valor de los datos de referencia. Un valor de 8 también puede indicar que la base de datos de referencia contiene datos adicionales del elemento de entrada. Por ejemplo, la validación puede añadir un número de edificio o de edificio secundario si encuentra una coincidencia perfecta para el nombre de calle o de edificio.
9	Los datos en la posición actual se han validado, pero no se han modificado y el estado de entrega es dudoso. Por ejemplo, el valor de validación de punto de entrega es incorrecto.
C	Los datos en la posición actual se han validado y verificado, pero los datos de nombres están desactualizados. La validación ha cambiado los datos de nombre.
D	Los datos en la posición actual se han validado y verificado, pero han cambiado de un exónimo a un nombre oficial.
E	Los datos en la posición actual se han validado y verificado. Sin embargo, la validación de direcciones ha estandarizado el idioma o la distinción de mayúsculas y minúsculas de caracteres. La validación de direcciones puede cambiar el idioma si el valor coincide exactamente con otro idioma. Por ejemplo, la validación de direcciones puede cambiar "Bruselas" por "Bruxelles" en una dirección belga.
F	Los datos en la posición actual se han validado y verificado, pero no se han modificado debido a que existe una coincidencia perfecta con los datos de referencia.

Las posiciones 19 y 20 de la cadena de salida están relacionadas con los datos de país.

En la siguiente tabla se describen los valores que la validación puede devolver para las posiciones 19 y 20:

Código	Descripción
0	La dirección de salida no contiene datos en la posición actual.
1	La validación de direcciones no reconoce los datos de país.
4	La validación de direcciones identifica el país a partir del valor de país predeterminado de la transformación del validador de direcciones.
5	La validación de direcciones no puede determinar el país debido a que los datos de referencia contienen varias coincidencias.
6	La validación de direcciones identifica el país a partir de un script.
7	La validación de direcciones identifica el país a partir del formato de dirección.
8	La validación de direcciones identifica el país a partir de los datos de municipio principal.
9	La validación de direcciones identifica el país a partir de los datos de provincia.
C	La validación de direcciones identifica el país a partir de los datos de territorio.
D	La validación de direcciones identifica el país a partir del nombre de país, pero el nombre contiene errores.
E	La validación de direcciones identifica el país a partir de los datos de direcciones, como un código ISO o un nombre de país.
F	La validación de direcciones identifica el país a partir del conjunto de valores de Forzar país de la transformación del validador de direcciones.

Valores de puerto de salida del estado de resultado de elementos extendido

El estado de resultado de elementos extendido es una cadena de 20 caracteres en la que cada carácter representa un elemento de dirección de entrada diferente. Los códigos de salida de puertos complementan los datos de estado en el puerto de estado de entrada de elementos y en el puerto de estado de resultado de elementos. El código de salida de puertos también puede indicar la presencia de información adicional acerca de un elemento de dirección en los datos de referencia.

Busque el puerto en el grupo de puertos de Información de estado.

La siguiente tabla describe los valores de puerto de estado de resultado de elementos extendido:

Código	Descripción
1	Los datos de referencia de direcciones contienen información adicional acerca del elemento de dirección. La validación de direcciones no requiere la información adicional.
2	La validación de direcciones ha actualizado el elemento de dirección para resolver un error de datos o de formato. La validación de direcciones no ha comprobado el elemento de dirección.

Código	Descripción
3	La validación de direcciones ha actualizado el elemento de dirección para resolver un error de datos o de formato. La validación de direcciones ha verificado los datos de número en el elemento de dirección.
4	La validación de direcciones ha movido el elemento de dirección a otro campo para resolver un error de formato.
5	Los datos de referencia de direcciones contienen una versión alternativa del elemento de dirección, como un nombre de localidad preferido.
6	La validación de direcciones no ha comprobado todos los componentes del elemento de dirección. El elemento incluye datos que la validación de direcciones no puede validar.
7	La validación de direcciones ha encontrado un elemento de dirección en una posición incorrecta de una dirección. La validación de direcciones ha movido el elemento de dirección a la posición correcta.
8	La validación de direcciones ha encontrado un elemento de dirección válido en el campo de datos incorrecto. La validación de direcciones ha movido el elemento de dirección al campo correcto.
9	La validación de direcciones ha generado el elemento de salida según las reglas de validación del proveedor de servicios de correo.
A	La validación de direcciones ha encontrado elementos de dirección de diferentes tipos de dirección que se pueden elegir para la posición actual. La validación de direcciones ha seleccionado el elemento de dirección de salida que cumple con las reglas del proveedor de servicios de correo en el país de destino.
B	La validación de direcciones no puede determinar la relevancia de elementos. La validación de direcciones devuelve el valor predeterminado del país que especifique la dirección.
C	Modo de lista de sugerencias. La validación de direcciones puede devolver sugerencias de dirección adicionales para el elemento de dirección. Para devolver las sugerencias adicionales, actualice la propiedad Recuento máximo de resultados de la transformación del validador de direcciones.
D	La validación de direcciones ha interpolado los datos numéricos en el elemento de dirección.
E	La validación de direcciones no puede devolver el elemento de dirección en el idioma preferido. La validación de direcciones devuelve el elemento en el idioma predeterminado.
F	Modo de búsqueda de código de dirección. La dirección de entrada ha caducado.

Valores del puerto de salida de la puntuación de viabilidad de envío de correo

El valor de puntuación de viabilidad de envío de correo representa una estimación de la viabilidad de entrega de la dirección de salida. Use la puntuación de viabilidad de envío de correo como indicador general de la viabilidad de entrega de la dirección. Busque el puerto en el grupo de puertos de Información de estado.

La transformación del validador de direcciones considera varios factores cuando calcula la puntuación de viabilidad de envío de correo. La transformación se basa en cálculos principalmente con el valor de código de coincidencia y el valor de estado del resultado del elemento para la dirección. Otros factores que influyen en la puntuación de viabilidad de envío de correo incluyen la relevancia postal de los valores de dirección y la granularidad de los datos de referencia para el país.

El valor del puerto de puntuación de viabilidad de envío de correo brinda una estimación de la viabilidad de entrega de la dirección. La puntuación no es un indicador preciso ni definitivo de la viabilidad de entrega de la dirección.

En la siguiente tabla se indican los códigos de salida de la puntuación de viabilidad de envío de correo:

Valor	Resumen	Descripción
5	Fiabilidad total	Indica que la validación de direcciones ha comprobado y verificado todos los elementos relevantes de la dirección de entrada.
4	Casi seguro	Indica cualquiera de los siguientes escenarios: <ul style="list-style-type: none">- Uno o varios elementos de dirección relevantes no se pueden comprobar debido a que faltan datos de referencia. Otros elementos de dirección se han verificado.- La validación de direcciones ha corregido uno o varios elementos relevantes con un nivel de confianza muy alto. Esto se produce cuando la validación de direcciones encuentra una coincidencia única entre la dirección de entrada y los datos de referencia, y el grado de variación es muy bajo.
3	No debería haber problemas	Indica que la validación de direcciones ha corregido uno o varios elementos relevantes de la dirección de entrada. La validación de direcciones ha encontrado una coincidencia única entre la dirección de entrada y los datos de referencia, y el grado de variación es aceptable.
2	Posibilidad razonable	Indica que la validación de direcciones no puede corregir ni verificar la dirección por uno de los siguientes motivos: <ul style="list-style-type: none">- La validación de direcciones no puede identificar una coincidencia candidata en los datos de referencia con la confianza suficiente.- La validación de direcciones ha encontrado varias coincidencias candidatas con niveles de confianza similares. El proveedor de servicios de correo podría realizar la entrega en la dirección.
1	Arriesgado	Indica que la validación de direcciones solo puede encontrar coincidencias de datos de referencia parciales relativas a la dirección de entrada.
0	No entregable	Indica que la validación de direcciones no puede encontrar una coincidencia relativa a la dirección en los datos de referencia. Faltan muchos elementos en la dirección de entrada, o la validación de direcciones no puede verificar la mayoría de los elementos de la dirección.

Valores del puerto de salida del código de coincidencia

El valor de código de coincidencia resume los resultados de la comparación de la dirección de entrada con los datos de referencia. El código también resume cualquier corrección que la transformación haya hecho a la dirección. Busque el puerto en el grupo de puertos de Información de estado.

En la siguiente tabla se describen los valores del puerto de salida del código de coincidencia.

Código	Descripción
A1	La búsqueda de códigos de dirección ha encontrado una dirección parcial o completa para el código de entrada.
A0	La búsqueda de códigos de dirección no ha encontrado ninguna dirección para el código de entrada.
C4	Corregido. Todos los elementos relevantes para el envío postal se han comprobado.
C3	Corregido. Algunos elementos no se pueden comprobar.
C2	Corregidos, aunque no se ha podido determinar el estado de entrega debido a la ausencia de datos de referencia.
C1	Corregidos, aunque no se ha podido determinar el estado de entrega debido a que la estandarización del usuario ha introducido errores.
I4	Los datos no se pueden corregir completamente, pero hay una única coincidencia con una dirección en los datos de referencia.
I3	Los datos no se pueden corregir completamente, pero hay varias coincidencias con direcciones en los datos de referencia.
I2	Los datos no se pueden corregir. El modo por lotes devuelve direcciones sugeridas parciales.
I1	Los datos no se pueden corregir. El modo por lotes no se puede sugerir una dirección.
N7	Error de validación. La validación no se produjo porque la validación de una sola línea no está desbloqueada.
N6	Error de validación. La validación no se produjo porque la validación de una sola línea no es compatible con el país de destino.
N5	Error de validación. La validación no se produjo porque la base de datos de referencia ha caducado.
N4	Error de validación. La validación no se produjo porque los datos de referencia están dañados o tienen un formato incorrecto.
N3	Error de validación. La validación no se produjo porque los datos de país no se pueden desbloquear.
N2	Error de validación. La validación no se produjo porque la base de datos de referencia requerida no está disponible.
N1	Error de validación. La validación no se produjo porque el país no se reconoce o no es compatible.

Código	Descripción
Q3	Modo de lista de sugerencias. La validación de direcciones puede recuperar una o más direcciones completas a partir de los datos de referencia de direcciones que se corresponden con la dirección de entrada.
Q2	Modo de lista de sugerencias. La validación de direcciones puede combinar los elementos de la dirección de entrada y los elementos de los datos de referencia de direcciones para crear una dirección completa.
Q1	Modo de lista de sugerencias. La validación de direcciones no puede sugerir una dirección completa. Para generar una sugerencia de dirección completa, añada datos a la dirección de entrada.
Q0	Modo de lista de sugerencias. No hay suficientes datos de entrada para generar una sugerencia.
RB	País reconocido por la abreviatura. Reconoce códigos de país ISO de dos y de tres caracteres. También puede reconocer abreviaturas habituales como "GER" para Alemania.
RA	País reconocido por la propiedad Forzar país.
R9	País reconocido por la propiedad País predeterminado.
R8	País reconocido por el nombre de país.
R7	País reconocido por el nombre de país, pero el proceso de validación identificó errores en los datos del país.
R6	País reconocido por los datos de territorio.
R5	País reconocido por los datos de provincia.
R4	País reconocido por los datos de municipio principal.
R3	País reconocido por el formato de dirección.
R2	País reconocido por un script.
R1	País no reconocido porque hay varias coincidencias disponibles.
R0	País no reconocido.
S4	Modo de análisis. La dirección se ha analizado correctamente.
S3	Modo de análisis. La dirección se ha analizado y se han obtenido varios resultados.
S1	Modo de análisis. Se ha producido un error de análisis debido a que un formato de entrada no coincide.
V4	Verificado. Los datos de entrada son correctos. La validación de direcciones ha comprobado todos los elementos relevantes para el envío postal y las entradas coinciden de manera perfecta.

Código	Descripción
V3	Verificado. Los datos de entrada son correctos, pero algunos o todos los elementos se han estandarizado, o bien la entrada contiene nombres que han caducado o que son exónimos.
V2	Verificado. Los datos de entrada son correctos, pero algunos elementos no se han podido verificar debido a que los datos de referencia están incompletos.
V1	Verificado. Los datos de entrada son correctos, pero la estandarización del usuario ha afectado la factibilidad de entrega de forma negativa. Por ejemplo, el código postal tiene una longitud demasiado corta.

Valores del puerto de salida del estado de geocodificación

En la siguiente tabla se describen los valores del puerto de salida del estado de geocodificación. Busque este puerto en el grupo de puertos de geocodificación.

Seleccione este puerto si ha instalado datos de referencia de geocodificación para un país de dirección de entrada.

Valor	Descripción
EGC0	No se pueden añadir geocódigos a la dirección de entrada debido a que los geocódigos no están disponibles para la dirección.
EGC1-3	Reservado para uso futuro.
EGC4	Los geocódigos tienen una precisión aproximada a nivel de código postal.
EGC5	Los geocódigos son precisos a nivel de código postal.
EGC6	Los geocódigos son precisos a nivel de localidad.
EGC7	Los geocódigos son precisos a nivel de calle.
EGC8	Los geocódigos son precisos a nivel de número de casa. Los geocódigos calculan la ubicación del número de casa e incluyen un desfase a un lado de la calle que contiene el buzón.
EGC9	Los geocódigos son precisos en relación con el punto de llegada o la dirección exacta.
EGCA	Los geocódigos son precisos en relación con el centro de la parcela de terreno.
EGCC	La base de datos de geocódigos está dañada.
EGCN	No se puede encontrar la base de datos de geocódigos.
EGCU	La base de datos de geocódigos está bloqueada.

Nota: Informatica ya no emite datos de referencia relativos a la geocodificación de direcciones exactas y centroides de parcelas.

Configuración general de la transformación del validador de direcciones

Ajuste la configuración general para configurar los parámetros necesarios para la validación de direcciones.

Puede configurar las siguientes propiedades en la vista **Configuración general**:

País predeterminado

Especifica el conjunto de datos de referencia de direcciones que utiliza la transformación si no puede identificar un destino de país en la dirección de entrada. Seleccione Ninguno si sus datos ya incluyen el país.

También puede establecer el país predeterminado como una propiedad avanzada en la transformación.

Forzar país

Propiedad opcional. Reemplaza el nombre o abreviatura del país en la dirección de entrada con el nombre o abreviatura del país predeterminado. Si la dirección de entrada no identifica a un país, la transformación anexa los datos del país predeterminado a la dirección.

Separador de línea

Especifica el símbolo delimitador que separa campos de datos en una dirección formada por una línea.

También puede especificar un separador de línea como una propiedad avanzada en la transformación.

Uso de mayúsculas y minúsculas

Define el estilo de formato de mayúsculas y minúsculas de los caracteres para los datos de salida. Seleccione la opción Mixto para seguir el estándar de datos de referencia de dirección de iniciales en mayúsculas. Seleccione la opción Preservado para escribir la dirección en el estilo de mayúsculas y minúsculas que utilizan los datos de referencia de direcciones.

También puede establecer el estilo de mayúsculas y minúsculas como una propiedad avanzada en la transformación.

Modo

Determina el tipo de validación que realiza la transformación.

Seleccione una de las siguientes opciones:

Tipo de modo	Descripción
Búsqueda del código de dirección	Devuelve una dirección parcial o una dirección completa de los datos de referencia cuando se proporciona un código de dirección como entrada. Son varios los países que admiten códigos de direcciones que representan la localidad, la calle, el edificio o un buzón único para una dirección.
Lote	Realiza la validación de direcciones en los registros de un conjunto de datos. La validación por lotes se centra en la integridad y la factibilidad de entrega de las direcciones. El modo por lotes no devuelve sugerencias para direcciones con una calidad deficiente. Este es el modo predeterminado.
Certificado	Realiza la validación de direcciones de los registros en un conjunto de datos con los estándares de certificación del país especificado. Los estándares de certificación requieren que cada dirección identifique un buzón único. Puede realizar la validación de direcciones certificadas en direcciones de Australia, Francia, Nueva Zelanda, Reino Unido y Estados Unidos.

Tipo de modo	Descripción
Reconocimiento del país	Determina un país de destino para la dirección postal. La transformación no realiza la validación de direcciones en el modo de reconocimiento del país.
Interactivo	Completa una dirección válida que está incompleta. Cuando una dirección de entrada incompleta coincide con más de una dirección en los datos de referencia, la transformación devuelve todas las direcciones válidas hasta el límite especificado en el recuento máximo de resultados.
Analizar	Analiza los datos en campos de dirección. La transformación no realiza la validación de direcciones en el modo de análisis.
Lista de sugerencias	Devuelve una lista de direcciones válidas de los datos de referencia cuando una dirección de entrada contiene información fragmentada. Si un fragmento de dirección coincide con más de una dirección en los datos de referencia, la transformación devuelve todas las direcciones válidas hasta el límite especificado en el recuento máximo de resultados.

También puede establecer el modo como una propiedad avanzada en la transformación.

Propiedades de validación de direcciones en la ventana Preferencias

Las propiedades del motor de validación de direcciones y los archivos de datos de referencia de direcciones que lee el motor se pueden ver en la herramienta Developer tool. La herramienta Developer tool expone las propiedades del motor que utiliza el servicio de integración de datos para ejecutar asignaciones de validación de direcciones. La herramienta Developer tool enumera las propiedades del servicio de administración de contenido que rigen las operaciones de validación de direcciones.

Utilice la ventana **Preferencias** en la herramienta Developer tool para revisar las propiedades. Seleccione la opción **Estado de contenido** en la ventana **Preferencias** para identificar el servicio de administración de contenido que utiliza el servicio de integración de datos actual. Para ver las propiedades, seleccione el servicio de administración de contenido local.

Puede ver las siguientes propiedades:

Datos de validación de direcciones

Las propiedades de datos de validación de direcciones enumeran los tipos de datos de referencia que el servicio de administración de contenido actual puede proporcionar al servicio de integración de datos. Las propiedades también indican los países en los que se aplican los datos de referencia.

Motor de validación de direcciones

Las propiedades del motor de validación de direcciones incluyen la versión actual del motor, el motor en el que se actualizaron por última vez los componentes de certificación y el método de precarga de datos.

Licencia de validación de direcciones

En las propiedades de licencia de validación de direcciones se incluye la información relativa a la licencia de los datos de referencia que el servicio de administración de contenido actual puede proporcionar al servicio de integración de datos.

Propiedades de datos de validación de direcciones

Las propiedades de datos de validación de direcciones enumeran los tipos de datos de referencia que el servicio de administración de contenido actual puede proporcionar al servicio de integración de datos. Las propiedades también incluyen los países en los que se aplican los datos de referencia.

En la siguiente tabla se describen las propiedades de datos que se muestran al seleccionar el servicio de administración de contenido en la vista **Estado de contenido**:

Propiedad	Descripción
Código ISO del país	El país al que se aplica el archivo de datos de referencia de direcciones. La propiedad muestra el código ISO de tres caracteres para el país.
Fecha de vencimiento	La fecha de vencimiento de la licencia actual. Informatica publica un archivo nuevo en la fecha de caducidad. Se puede utilizar el archivo de datos de referencia de direcciones actual una vez haya vencido, pero es posible que los datos que contenga ya no sean precisos.
Tipo de país	El tipo de validación de direcciones que puede llevar a cabo con los datos. Seleccione el tipo de procesamiento en la opción Modo de la ficha Configuración general . Si el modo que ha seleccionado no se corresponde con un archivo de datos de direcciones del dominio, se producirá un error al realizar la asignación de validación de direcciones.
Fecha de vencimiento de desbloqueo	La fecha de vencimiento de la licencia. Después de esta fecha no es posible utilizar ninguna versión del archivo. La propiedad Desbloquear fecha de vencimiento y la propiedad Fecha de vencimiento en la vista Propiedades de la licencia de validación de direcciones representan la misma información.
Fecha de inicio de desbloqueo	La fecha de entrada en vigor de la licencia para el modo que identifica la propiedad Tipo de país y el país que identifica la propiedad ISO del país. Antes de esta fecha no es posible utilizar ninguna versión del archivo.

Propiedades de licencia de validación de direcciones

En las propiedades de licencia de validación de direcciones se incluye la información relativa a la licencia de los datos de referencia que el servicio de administración de contenido actual puede proporcionar al servicio de integración de datos.

En la siguiente tabla se describen las propiedades de licencia que se muestran al seleccionar el servicio de administración de contenido en la vista **Estado de contenido**:

Propiedad	Descripción
Código de desbloqueo	El código de licencia que desbloquea los datos de referencia para el modo que identifica la propiedad Tipo de código. La herramienta Developer tool muestra los primeros cuatro caracteres del código y enmascara los otros.
Tipo de código	El modo de validación de direcciones que se puede realizar con los datos que especifica la licencia. Informatica emite un único código de licencia para cada modo. El código de licencia puede aplicarse a uno o varios países. Seleccione el tipo de procesamiento en la opción Modo de la ficha Configuración general . Si el modo que ha seleccionado no se corresponde con un archivo de datos de direcciones del dominio, se producirá un error al realizar la asignación de validación de direcciones.

Propiedad	Descripción
Lista de país	Los países en los que el código de desbloqueo actúa sobre los datos de referencia. La propiedad Lista de países contiene uno o varios códigos ISO de tres caracteres para cada país.
Estado	Estado del código de licencia. La propiedad devuelve OK cuando el archivo de licencia es válido.
Fecha de vencimiento	La fecha de vencimiento de la licencia. La propiedad Fecha de vencimiento y la propiedad Desbloquear fecha de vencimiento en la vista Propiedades de datos de validación de direcciones representan la misma información.

Propiedades del motor de validación de direcciones

Las propiedades del motor de validación de direcciones incluyen la versión actual del motor, el motor en el que se actualizaron por última vez los componentes de certificación y el método de precarga de datos.

En la siguiente tabla se describen las propiedades del motor que se muestran al seleccionar el servicio de administración de contenido en la vista **Estado de contenido**:

Propiedad	Valor
Versión de motor	Versión del motor de validación de direcciones que ejecuta el servicio de integración de datos.
Versión de CASS	Versión del motor de validación de direcciones en la que Informática ha actualizado por última vez los componentes de certificación CASS. Utilice la propiedad para identificar la versión del motor en un informe de certificación CASS. La propiedad también incluye el ciclo de certificación CASS que el motor admite. Por ejemplo, el motor podría admitir el ciclo de certificación N.
Versión de AMAS	Versión del motor de validación de direcciones en la que Informática ha actualizado por última vez los componentes de certificación AMAS. Utilice la propiedad para identificar la versión del motor en un informe de certificación AMAS.
Versión de SendRight	Versión del motor de validación de direcciones en la que Informática ha actualizado por última vez los componentes de certificación SendRight. Utilice la propiedad para identificar la versión del motor en un informe de certificación SendRight.
Versión de SERP	Versión del motor de validación de direcciones en la que Informática ha actualizado por última vez los componentes de certificación SERP. Utilice la propiedad para identificar la versión del motor en un informe de certificación SERP.
Versión de SNA	Versión del motor de validación de direcciones en la que Informática ha actualizado por última vez los componentes de certificación SNA. Utilice la propiedad para identificar la versión del motor en un informe de certificación SNA.
Método de precarga	Método que utiliza el servicio de integración de datos para precargar la base de datos de referencia en la memoria. Las propiedades del servicio de administración de contenido especifican los países para los que el servicio de integración de datos precarga los datos de referencia. Los valores posibles son MAP y LOAD. El valor predeterminado es MAP. Tanto el método MAP como el método LOAD asignan un bloque de memoria y, a continuación, leen los datos de referencia en el bloque. Sin embargo, el método MAP puede compartir los datos de referencia entre varios procesos.

Propiedad	Valor
Tamaño de la memoria caché	Tamaño de la memoria caché de datos que utiliza el servicio de integración de datos para los datos de referencia que el servicio no precarga. Los valores posibles son NONE, SMALL y LARGE. El valor predeterminado es LARGE.
Utilización máxima de memoria	Número de megabytes de memoria que el motor de validación de direcciones puede asignar. El valor predeterminado es 4096.
Número máximo de objetos de dirección	Número máximo de instancias de validación de direcciones que el servicio de integración de datos puede ejecutar al mismo tiempo. El valor predeterminado es 3.
Número máximo de subprocesos	Número máximo de subprocesos que puede utilizar la validación de direcciones. El valor predeterminado es 2.
Recuento máximo de resultados	Número máximo de direcciones que la validación de direcciones puede devolver cuando se ejecuta una asignación en el modo de lista de sugerencias. El valor predeterminado es 20. El límite superior de la propiedad es 100.
Fecha actual	La fecha actual. La herramienta Developer tool devuelve los valores de propiedad que se aplican en la fecha actual.
Escribir XML BOM	Indica si el servicio de integración de datos escribe una marca de orden de byte en el archivo GetConfig.xml. Los valores posibles son ALWAYS, IF_NECESSARY y NEVER. El valor predeterminado es IF_NECESSARY.
Codificación XML	Identifica la codificación XML que utiliza el motor de validación de direcciones para leer y escribir datos.

Propiedades avanzadas de validación de direcciones

Configure las propiedades avanzadas para ayudarle a determinar cómo el servicio de integración de datos procesa los datos para la transformación del validador de direcciones.

Alias de localidad

Determina si la validación de direcciones reemplaza un alias de localidad válido con el nombre oficial de la localidad.

Un alias de localidad es un nombre de localidad alternativo que USPS reconoce como un elemento en una dirección de entrega. Puede utilizar la propiedad cuando configure la transformación del validador de direcciones para validar registros de direcciones de Estados Unidos en el modo certificado.

En la siguiente tabla se describen las opciones de Alias de localidad:

Opción	Descripción
Desactivada	Deshabilita la propiedad Alias de localidad.
Oficial	Reemplaza cualquier nombre de localidad alternativo o alias de localidad con el nombre oficial de la localidad. Opción predeterminada.
Conservar	Conserva un nombre de localidad alternativo o un alias de localidad válidos. Si el nombre de localidad introducido no es válido, la validación de direcciones reemplaza el nombre con el nombre oficial.

Alias de calle

Determina si la validación de direcciones reemplaza un alias de calle con el nombre de calle oficial.

Un alias de calle es un nombre de calle alternativo que USPS reconoce como un elemento en una dirección de entrega. Puede utilizar la propiedad cuando configure la transformación del validador de direcciones para validar registros de direcciones de Estados Unidos en el modo certificado.

En la siguiente tabla se describen las opciones de Alias de calle:

Opción	Descripción
Desactivada	No se aplica la propiedad.
Oficial	Reemplaza cualquier nombre de calle alternativo o alias de calle con el nombre oficial de la calle. Opción predeterminada.
Conservar	Conserva un nombre de calle alternativo o un alias de calle válidos. Si el nombre de calle introducido no es válido, la validación de direcciones reemplaza el nombre con el nombre oficial.

Uso de mayúsculas y minúsculas

Especifica si la transformación aplica un estilo de mayúsculas o minúsculas a los datos de la dirección de salida.

En la siguiente tabla se describen las opciones del uso de mayúsculas y minúsculas:

Opción	Descripción
Asignar parámetro	Utiliza un parámetro definido para establecer el uso de mayúsculas y minúsculas.
Minúsculas	Escribe la dirección de salida en minúsculas.
Mixto	Ejemplifica el uso de mayúsculas y minúsculas en vigor en el país de destino siempre que sea posible.
Reservado	Aplica el uso de mayúsculas y minúsculas que utilizan los datos de referencia de direcciones. Opción predeterminada.

Opción	Descripción
No hay cambios.	No aplica ningún estilo de mayúsculas o minúsculas a la dirección. Nota: La opción No hay cambios no garantiza que el uso de mayúsculas o minúsculas de la dirección de salida coincida con el de la dirección de entrada. Si la validación de direcciones sustituye un elemento de la dirección con un elemento de los datos de referencia, el elemento sigue el uso de mayúsculas y minúsculas que utilizan los datos de referencia.
Mayúsculas	Escribe la dirección de salida en mayúsculas.

También puede configurar el uso de mayúsculas o minúsculas en la ficha **Configuración general**.

Uso de parámetros

Puede utilizar uno de los siguientes parámetros para especificar el uso de mayúsculas y minúsculas:

- LOWER. Escribe la dirección de salida en minúsculas.
- MIXED. Emplea el uso de mayúsculas y minúsculas en vigor en el país de destino siempre que sea posible.
- NATIVE. Aplica el uso de mayúsculas y minúsculas que utilizan los datos de referencia de direcciones. Opción predeterminada. Coincide con la opción **Reservado**.
- NOCHANGE. No aplica ningún estilo de mayúsculas o minúsculas a la dirección.
- UPPER. Escribe la dirección de salida en mayúsculas.

Introduzca el valor del parámetro en mayúsculas.

País de origen

Identifica el país en el que se envían los registros de direcciones.

Seleccione un país de la lista. De forma predeterminada, la propiedad está vacía.

Tipo de país

Determina el formato del nombre o la abreviatura del país en los datos de salida del puerto de línea de dirección formateada o dirección completa. La transformación escribe el nombre o abreviatura del país en el formato estándar del país que seleccione.

En la siguiente tabla se describen las opciones del tipo de país:

Opción	País
ISO 2	Código ISO del país de dos caracteres
ISO 3	Código ISO del país de tres caracteres
ISO #	Código ISO del país de tres dígitos
Abreviatura	(Reservado para su uso en el futuro)
CN	Canadá

Opción	País
DA	(Reservado para su uso en el futuro)
DE	Alemania
EN	Gran Bretaña (predeterminado)
ES	España
FI	Finlandia
FR	Francia
GR	Grecia
IT	Italia
JP	Japón
HU	Hungría
KR	Corea, República de
NL	Países Bajos
PL	Polonia
PT	Portugal
RU	Rusia
SA	Arabia Saudí
SE	Suecia

País predeterminado

Especifica el conjunto de datos de referencia de direcciones que utiliza la transformación cuando un registro de direcciones no identifica un país de destino.

Seleccione un país de la lista. Utilice la opción predeterminada si los registros de direcciones incluyen información del país. El valor predeterminado es Ninguno.

También puede configurar el país predeterminado en la ficha **Configuración general**.

Uso de parámetro

Puede usar un parámetro para especificar el país predeterminado. Cuando cree el parámetro, escriba el código ISO 3166-1 alfa-3 para el país como valor del parámetro. Cuando introduzca un valor de parámetro, utilice caracteres en mayúscula. Por ejemplo, si todos los registros de direcciones incluyen información del país, introduzca NONE.

Prioridad de dirección dual

Determina el tipo de dirección que validar. Establezca la propiedad cuando los registros de direcciones de entrada contengan más de un tipo de datos de direcciones válidos.

Por ejemplo, use la propiedad cuando un registro de direcciones contenga elementos de apartado de correos y elementos de calle. La validación de direcciones lee los elementos de datos que contienen el tipo de datos de dirección que especifique. La validación de direcciones omite los datos incompatibles en la dirección.

La siguiente tabla describe las opciones de la propiedad Prioridad de dirección dual:

Opción	Descripción
Servicio de entrega	Valida los elementos de datos del servicio de entrega de una dirección, como los elementos de apartado de correos.
Administración postal	Valida los elementos de dirección requeridos por el proveedor de servicios de correo local. Opción predeterminada.
Calle	Valida los elementos de datos de calle de una dirección, como los elementos de número de edificio y los elementos de nombre de calle.

Abreviatura de elemento

Determina si la transformación devuelve la forma abreviada de un elemento de dirección. Puede establecer la transformación para que devuelva la forma abreviada si los datos de referencia de direcciones contienen abreviaturas.

Por ejemplo, el servicio postal de Estados Unidos (USPS) tiene formas breves y largas de muchos nombres de calles y localidades. La forma corta de HUNTSVILLE BROWNSFERRY RD es HSV BROWNS FRY RD. Puede seleccionar la propiedad Abreviatura de elemento cuando los valores de calle o localidad excedan la longitud máxima de campo que especifica USPS.

La opción está desactivada de manera predeterminada. Active la propiedad para que devuelva los valores de dirección abreviados. La propiedad devuelve el nombre de localidad y el código de localidad abreviados cuando se utiliza la transformación en modo por lotes. La propiedad devuelve el nombre de calle, el nombre de localidad y el código de localidad abreviados cuando se utiliza la transformación en el modo de certificación.

Instancias de ejecución

Especifica el número de subprocesos que el servicio de integración de datos intenta crear para la transformación actual en tiempo de ejecución. El servicio de integración de datos tiene en cuenta el valor de Instancias de ejecución si se anula la propiedad en tiempo de ejecución Número máximo de paralelismos en la asignación que contiene la transformación. El valor predeterminado de Instancias de ejecución es 1.

El servicio de integración de datos tiene en cuenta varios factores a la hora de determinar el número de subprocesos que debe asignar a la transformación. Los principales factores son el valor de Instancias de ejecución y los valores en la asignación y en los servicios de aplicación asociados del dominio.

El servicio de integración de datos lee los siguientes valores al calcular el número de subprocesos que debe usar para la transformación:

- El valor de *Número máximo de paralelismos* del servicio de integración de datos. El valor predeterminado es 1.

- Cualquier valor de *Número máximo de paralelismos* que se haya establecido en el nivel de asignación. El valor predeterminado es Automático.
- El valor de *Instancias de ejecución* en la transformación. El valor predeterminado es 1.

Si reemplaza el valor de *Número máximo de paralelismos* en el nivel de asignación, el servicio de integración de datos intentará usar el valor más bajo de las propiedades para determinar el número de subprocesos.

Si utiliza el valor predeterminado de *Número máximo de paralelismos* en el nivel de asignación, el servicio de integración de datos omitirá el valor de *Instancias de ejecución*.

El servicio de integración de datos también tiene en cuenta la propiedad *Número máximo de objetos de dirección* en el servicio de administración de contenido al calcular el número de subprocesos que debe crear. La propiedad *Número máximo de objetos de dirección* determina el número máximo de instancias de validación de direcciones que pueden ejecutarse simultáneamente en una asignación. La propiedad *Número máximo de objetos de dirección* debe ser mayor o igual al valor de *Número máximo de paralelismos* en el servicio de integración de datos.

Reglas y directrices de la propiedad *Instancias de ejecución*

Tenga en cuenta las siguientes reglas y directrices cuando establezca el número de instancias de ejecución:

- Puede que varios usuarios ejecuten asignaciones simultáneas en un servicio de integración de datos. Para calcular el número correcto de subprocesos, divida el número de unidades de procesamiento central a las que puede acceder el servicio entre el número de asignaciones simultáneas.
- En PowerCenter, el archivo de configuración *AD50.cfg* especifica el número máximo de instancias de validación de direcciones que pueden ejecutarse simultáneamente en una asignación.
- Cuando se utiliza el valor predeterminado de *Instancias de ejecución* y los valores predeterminados de *Número máximo de paralelismos*, las operaciones de transformación no admitirán particiones.
- Cuando defina un valor de *Instancias de ejecución* superior a 1, cambie la transformación del validador de direcciones de una transformación pasiva a una transformación activa.

Expansión de rango flexible

Impone un límite práctico sobre el número de direcciones que la transformación del validador de direcciones devuelve al establecer la propiedad *Rangos para expandir*. Puede establecer la propiedad *Rangos para expandir* y la propiedad *Expansión de rango flexible* al configurar la transformación para que se ejecute en el modo de lista de sugerencias.

La propiedad *Rangos para expandir* determina cómo la transformación devuelve sugerencias de dirección si una dirección de entrada no contiene datos de número de domicilio. Si la dirección de entrada no incluye datos contextuales, como un código postal completo, la propiedad *Rangos para expandir* puede generar un gran número de direcciones muy similares. La propiedad *Expansión de rango flexible* limita el número de direcciones que la propiedad *Rangos para expandir* genera para una dirección individual. Habilite la propiedad *Expansión de rango flexible* al establecer la propiedad *Rangos para expandir* en Todos.

La siguiente tabla describe las opciones de la propiedad *Expansión de rango flexible*:

Opción	Descripción
Activada	La validación de direcciones limita el número de direcciones que la propiedad <i>Rangos para expandir</i> añade a la lista de sugerencias. Opción predeterminada.
Desactivada	La validación de direcciones no limita el número de direcciones que la propiedad <i>Rangos para expandir</i> añade a la lista de sugerencias.

Nota: La transformación del validador de direcciones aplica la propiedad Expansión de rango flexible de forma diferente para cada dirección que devuelve a la lista de sugerencias. La transformación no impone un límite fijo sobre el número de direcciones expandidas de la lista. La transformación también considera el valor de la propiedad Recuento máximo de resultados al calcular el número de direcciones ampliadas que incluir en la lista.

Tipo de datos de geocodificación

Determina el modo en que la transformación del validador de direcciones calcula los datos de geocodificación para una dirección. Los datos de geocodificación son las coordenadas de latitud y longitud.

Los resultados de la geocodificación que la transformación devuelve dependen de los datos de referencia de geocodificación que se instalen. Para obtener información sobre los datos de referencia de geocodificación, póngase en contacto con Informática.

Seleccione una de las siguientes opciones de geocodificación:

Punto de llegada

Devuelve las coordenadas de latitud y longitud de la entrada de un edificio o de una parcela de terreno. Opción predeterminada.

Puede seleccionar la opción de punto de llegada para las direcciones de los países siguientes:

Australia, Austria, Canadá, Croacia, Dinamarca, Estonia, Finlandia, Francia, Alemania, Hungría, Italia, Letonia, Liechtenstein, Lituania, Luxemburgo, México, Mónaco, Países Bajos, Noruega, Polonia, Eslovaquia, Eslovenia, Suecia, Suiza y Estados Unidos.

Si especifica geocódigos de punto de llegada y la transformación del validador de direcciones no puede devolverlos para una dirección, la transformación devolverá geocódigos interpolados.

Estándar

Devuelve las coordenadas de latitud y longitud estimadas de la entrada al edificio o a la parcela de terreno. Un geocódigo estimado también se conoce como geocódigo interpolado.

La transformación del validador de direcciones usa los geocódigos más cercanos disponibles en los datos de referencia para calcular los geocódigos de la dirección.

Nota: Informática ya no emite datos de referencia relativos a la geocodificación de direcciones exactas o centroides de parcelas.

Uso de parámetros

Puede utilizar un parámetro para especificar el tipo de geocodificación. Introduzca ARRIVAL_POINT o NONE. Para obtener las geocodificaciones estándar, introduzca NONE.

Introduzca el valor del parámetro en mayúsculas.

Longitud de campo máxima global

Determina el número máximo de caracteres en cualquier línea de la dirección. Si la transformación del validador de direcciones escribe una línea de dirección de salida que contiene más caracteres que los que especifique, la transformación abrevia los elementos de dirección de la línea.

Use la propiedad para controlar la longitud de línea de la dirección. Por ejemplo, los estándares del SNA requieren que una dirección no contenga más de 38 caracteres en una línea. Si se generan direcciones de acuerdo con el estándar del SNA, defina la longitud de campo máxima global en 38.

El valor predeterminado es 1024.

Uso de parámetros

Puede utilizar un parámetro para especificar el número máximo de direcciones. Para establecer el valor del parámetro, introduzca un número entero entre 0 y 1024.

Descriptor global preferido

Determina el formato de los descriptores de edificio, edificio secundario y calle que la transformación del validador de direcciones escribe en los datos de salida. Seleccione un descriptor cuando los datos de referencia de direcciones del país de destino contengan un intervalo de descriptores para uno o más elementos de datos.

La siguiente tabla describe las opciones de la propiedad:

Opción	Descripción
Base de datos	Devuelve el descriptor que la base de datos de referencia especifica para el elemento de la dirección. Si la base de datos no especifica un descriptor para la dirección, la transformación copiará el valor de entrada en la dirección de salida. La base de datos es el valor predeterminado.
Largo	Devuelve la forma completa del descriptor, por ejemplo <i>Calle</i> .
Conservar entrada	Copia la forma del descriptor de la dirección de entrada en la dirección de salida. Si el descriptor de entrada no es una versión válida del descriptor, la transformación devolverá un descriptor válido equivalente desde la base de datos de referencia.
Corto	Devuelve una forma abreviada del descriptor, por ejemplo <i>C/</i> .

Tipo de formato de entrada

Describe el tipo de información más común que contienen los datos de entrada que no se utilizan. Utilice la propiedad Tipo de formato de entrada cuando conecte datos de entrada a los puertos Dirección completa o Línea de dirección formateada. Seleccione la opción que mejor describa la información de los datos de origen de la asignación.

Seleccione una de las siguientes opciones:

- Todos
- Dirección
- Organización
- Contacto
- Organización/Contacto
La dirección incluye información de la organización e información de contacto.
- Organización/Departamento
La dirección incluye información de la organización e información del departamento.

La opción predeterminada es Todos.

Formato de entrada con país

Determina si la entrada contiene datos del país. Seleccione esta propiedad si conecta los datos de entrada con los puertos de entrada Dirección completa o Línea de dirección formateada y si los datos contienen información del país.

La opción está desactivada de manera predeterminada.

Separador de línea

Especifica el símbolo delimitador que indica los saltos de línea en una dirección formateada.

Seleccione una de las siguientes opciones:

- Asignar un parámetro para identificar el separador de línea
- Retorno de carro (CR)
- Coma
- Avance de línea (LF)
- Ninguna
- Punto y coma
- Tabulador
- Nueva línea en Windows (CRLF)

El valor predeterminado es punto y coma.

También puede configurar el separador de línea en la ficha **Configuración general**.

Uso de parámetros

Puede usar un parámetro para especificar el separador de línea. El valor del parámetro distingue entre mayúsculas y minúsculas. Introduzca el valor del parámetro en mayúsculas.

Especifique uno de los siguientes valores:

- CR
- CRLF
- COMMA
- LF
- PIPE
- SEMICOLON
- SPACE
- TAB

Alternativas coincidentes

Determina si la validación de direcciones reconoce nombres de lugares alternativos, como sinónimos o nombres históricos, en una dirección de entrada. La propiedad se aplica a datos de calle, localidad y provincia.

Nota: La propiedad Alternativas coincidentes no conserva los nombres alternativos en una dirección validada.

En la siguiente tabla se describen las opciones de las alternativas coincidentes:

Opción	Descripción
Todos	Reconoce todos los nombres de calles y nombres de lugares alternativos conocidos. Opción predeterminada.
Solo archivos	Reconoce solo los nombres históricos. Por ejemplo, la validación de direcciones valida "Constantinopla" como una versión histórica de "Estambul".
Ninguno	No reconoce nombres de calles y nombres de lugares alternativos.
Solo sinónimos	Solo reconoce sinónimos y exónimos. Por ejemplo, la validación de direcciones valida "Seville" como un exónimo de "Sevilla".

Coincidencia de archivo extendido

Determina si la validación de direcciones devuelve un código de punto de entrega único para una dirección japonesa obsoleta.

Los archivos de datos de referencia de direcciones de Japón incluyen datos para direcciones obsoleta o retiradas junto con las direcciones actuales de los buzones correspondientes. Cuando seleccione la propiedad Coincidencia de archivo extendido, la validación de direcciones devuelve el código de punto de entrega para la versión actual de cada dirección. La validación de direcciones también escribe un valor en el puerto de estado de resultado de elementos extendido para indicar que la dirección de entrada ha caducado.

Para recuperar la dirección actual de los datos de referencia de direcciones, escriba el código de dirección como un elemento de entrada.

La siguiente tabla describe las opciones de Coincidencia de archivo extendido:

Opción	Descripción
Desactivada	No se aplica la propiedad.
Activada	Devuelve el código de dirección de la versión actual de una dirección japonesa obsoleta.

La propiedad Estado de elementos extendido utiliza datos complementarios y datos de búsqueda de códigos de direcciones de Japón. Para aplicar la propiedad en la validación de direcciones, configure la transformación para que se ejecute en modo de búsqueda de códigos de direcciones.

Ámbito coincidente

Determina la cantidad de datos que la transformación coteja con los datos de referencia de direcciones durante la validación de direcciones.

En la siguiente tabla se describen las opciones del ámbito coincidente:

Opción	Descripción
Todos	Valida todos los puertos seleccionados. Opción predeterminada.
Punto de entrega	Valida los datos de la dirección de edificios y subedificios, además de los datos que valida la opción Calle.
Localidad	Valida los datos de provincia, localidad y código postal.
Calle	Valida datos de direcciones de calle además de los datos que valida la opción Localidad.

Recuento máximo de resultados

Determina el número máximo de direcciones que la validación de direcciones puede devolver en el modo de lista de sugerencias.

Puede establecer un número máximo en el intervalo de 1 a 100. El valor predeterminado es 20.

Nota: El modo de lista de sugerencias comprueba las direcciones en los datos de referencia de direcciones y devuelve una lista de direcciones que son posibles coincidencias para la dirección de entrada. Cuando se comprueba una dirección en el modo de lista de sugerencias, la validación de direcciones devuelve primero la mejor coincidencia.

Uso de parámetros

Puede utilizar un parámetro para especificar el número máximo de direcciones. Para establecer el valor del parámetro, introduzca un número entero entre 0 y 100.

Modo

Determina el tipo de análisis de direcciones que realiza la transformación. También puede configurar el modo en la ficha **Configuración general** de la transformación.

TEMAS RELACIONADOS

- [“ Configuración general de la transformación del validador de direcciones ” en la página 106](#)

Nivel de optimización

Determina cómo la transformación coteja datos de direcciones de entrada y datos de referencia de direcciones. La propiedad define el tipo de coincidencia que la transformación debe encontrar entre los datos de entrada y los datos de referencia antes de que puede actualizar el registro de direcciones.

En la siguiente tabla se describen las opciones del nivel de optimización:

Opción	Descripción
Estrecha	La transformación analiza los números de edificios o de domicilios con la información de la calle antes de realizar la validación. De lo contrario, valida los elementos de la dirección de entrada estrictamente según la estructura del puerto de entrada. La opción estrecha es la que realiza la validación de la dirección de forma más rápida, pero es posible que se obtengan resultados menos precisos que en otras opciones.
Estándar	La transformación analiza varios tipos de información de la dirección de los datos de entrada antes de llevar a cabo la validación. Al seleccionar la opción estándar, la transformación actualiza una dirección en caso de que pueda cotejar varios valores de entrada con los datos de referencia. El valor predeterminado es Estándar.
Ancho	La transformación utiliza la configuración del análisis estándar y realiza más operaciones de análisis en los datos de entrada. Al seleccionar esta opción, la transformación actualiza una dirección en caso de que pueda hacer coincidir al menos un valor de entrada con los datos de referencia. Esta opción aumenta los tiempos de ejecución de la asignación.

Uso de parámetros

Puede utilizar un parámetro para especificar el nivel de optimización. Introduzca NARROW, STANDARD o WIDE. Introduzca el valor del parámetro en mayúsculas.

Tipo de formato de salida

Describe el tipo de información más común que la transformación escribe en el puerto de salida de línea de dirección formateada o dirección completa. Seleccione la opción que mejor describa los datos que espera en el puerto de salida.

Seleccione una de las siguientes opciones:

- Todos
- Dirección
- Organización
- Contacto
- Organización/Contacto
La dirección incluye información de la organización e información de contacto.
- Organización/Departamento
La dirección incluye información de la organización e información del departamento.

La opción predeterminada es Todos.

Formato de salida con país

Determina si la transformación escribe datos de identificación de país en los puertos de salida de línea de dirección formateada o dirección completa.

La opción está desactivada de manera predeterminada.

Idioma preferido

Determina los idiomas en los que la transformación del validador de direcciones devuelve elementos de dirección cuando los conjuntos de datos de referencia contienen datos en varios idiomas. Se puede establecer un idioma preferido para las direcciones de Bélgica, Canadá, China, Finlandia, Hong Kong, Irlanda, Israel, Macao, Suiza y Taiwán.

La transformación del validador de direcciones puede devolver datos de dirección en los siguientes idiomas:

- El idioma predeterminado para la dirección en los datos de referencia de direcciones. El idioma por defecto es la principal idioma hablado en la región a la que pertenece cada dirección.
- Cualquier otro idioma que los datos de referencia de direcciones admite para una dirección. Por ejemplo, los datos de referencia de Bélgica contienen elementos de dirección en flamenco, francés y alemán.

Los datos de referencia de direcciones pueden contener datos de un elemento de dirección única o una dirección completa en varios idiomas. Por ejemplo, la validación de dirección puede devolver todos los elementos de direcciones para Irlanda en inglés y puede devolver información de la calle, la localidad y la provincia en irlandés. Además, los datos de referencia podrían especificar diferentes idiomas predeterminados para direcciones en diferentes partes de un país. Por ejemplo, en los datos de referencia de Suiza, el idioma predeterminado varía de región a región entre francés, alemán e italiano.

En la tabla siguiente se resumen las opciones que puede seleccionar en la propiedad Idioma preferido:

Opción	Descripción
Base de datos	Devuelve cada dirección en el idioma especificado por los datos de referencia de direcciones. Los datos de referencia de direcciones podrían especificar idiomas diferentes para direcciones de distintas regiones de un país. Base de datos es la opción predeterminada.
Alternativa 1, Alternativa 2, Alternativa 3	Devuelve los elementos de dirección en un idioma alternativo de los datos de referencia. Los idiomas alternativos dependen del país al que pertenece la dirección.
Inglés	Devuelve elementos de dirección en inglés cuando los datos de referencia contienen los datos en inglés. Devuelve los demás elementos de la dirección en el idioma predeterminado de la región a la que pertenece la dirección.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si la validación de direcciones no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Nota: Un conjunto de datos de referencia de direcciones podría contener algunos elementos de dirección en un idioma no predeterminado, pero no otros. Si la transformación no puede encontrar un elemento en el idioma especificado por la propiedad, devolverá el elemento en el idioma predeterminado.

Cuando establece una opción de idioma preferido, compruebe que el juego de caracteres que especifica la propiedad Codificación de caracteres preferida sea compatible con los datos de la dirección de salida que usted espera.

Soporte multilingüe para direcciones de Bélgica

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Bélgica:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve las direcciones en el idioma principal de la región a la que pertenece la dirección. El idioma puede ser flamenco, francés o alemán.
Inglés	Devuelve la información de la provincia, la localidad y la calle en inglés si los datos de referencia de la dirección contienen la información en inglés. Devuelve los demás elementos de la dirección en el idioma principal de la región a la que pertenece la dirección.
Alternativa 1	Devuelve la información de la provincia, la localidad y la calle en flamenco.
Alternativa 2	Devuelve la información de la provincia, la localidad y la calle en francés.
Alternativa 3	Devuelve la información de la provincia, la localidad y la calle en alemán.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Soporte multilingüe para direcciones de Canadá

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Canadá:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve las direcciones en inglés para todas las provincias excepto Quebec. Devuelve las direcciones de Quebec en francés.
Inglés	Devuelve todas las direcciones en inglés.
Alternativa 1	Devuelve todas las direcciones en inglés.

Opción	Descripción
Alternativa 2	Devuelve las direcciones de Quebec en francés. En el resto de provincias aparte de Quebec, la transformación devuelve los descriptores de la calle, las indicaciones para llegar y los nombres de las provincias en francés, y devuelve otros elementos de dirección en inglés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Soporte multilingüe para direcciones de China

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de China:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en chino.
Inglés	Devuelve las versiones en inglés del descriptor de calle y los valores direccionales. Devuelve el resto de la información en chino. Los elementos de las direcciones en inglés omiten elementos de transliteración como "shi".
Alternativa 1	Devuelve toda la información de la dirección en el idioma de la base de datos.
Alternativa 2	Devuelve toda la información de la dirección en el idioma de la base de datos.
Alternativa 3	Devuelve toda la información de la dirección en el idioma de la base de datos.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Tenga en cuenta las siguientes reglas y directrices al seleccionar el idioma preferido:

- Para devolver la dirección en chino, seleccione Base de datos, Alternativa 1, Alternativa 2 o Alternativa 3.
Para devolver la dirección en un conjunto de caracteres chinos, establezca la propiedad Codificación de caracteres preferida en Base de datos.
- Para devolver el descriptor de calle y las indicaciones direccionales en inglés, seleccione Inglés.
Para devolver la dirección en un conjunto de caracteres latinos o ASCII, establezca la propiedad Codificación de caracteres preferida en un valor LATIN o ASCII.
- Si selecciona un valor LATIN o ASCII como codificación de caracteres preferida y Base de datos como el idioma preferido, la validación de direcciones devuelve los datos de direcciones en Pinyin.

Soporte multilingüe para direcciones de Finlandia

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Finlandia:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en finés.
Alternativa 1	Devuelve toda la información de la dirección en el idioma de la base de datos.
Alternativa 2	Devuelve la información de la calle, la localidad y la provincia en sueco. Devuelve el resto de la información en finés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Soporte multilingüe para direcciones de Hong Kong

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Hong Kong:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en chino.
Inglés	Devuelve toda la información de direcciones en inglés.
Alternativa 1	Devuelve toda la información de la dirección en el idioma de la base de datos.
Alternativa 2	Devuelve toda la información de direcciones en inglés.
Alternativa 3	Devuelve toda la información de la dirección en el idioma de la base de datos.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Tenga en cuenta las siguientes reglas y directrices al seleccionar el idioma preferido para Hong Kong:

- Para devolver la dirección en un conjunto de caracteres chinos, establezca la propiedad Codificación de caracteres preferida en Base de datos.
- Para devolver la dirección en un conjunto de caracteres latinos o ASCII, establezca la propiedad Codificación de caracteres preferida en un valor LATIN o ASCII.
- El idioma de los datos de entrada puede afectar al funcionamiento de la opción Conservar entrada en una dirección de Hong Kong. La validación de direcciones identifica el idioma de entrada como inglés cuando los datos de entrada utilizan caracteres ASCII de 7 bits e incluyen un descriptor en inglés.

Soporte multilingüe para direcciones de Irlanda

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Irlanda:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en inglés.
Inglés	Devuelve toda la información de direcciones en inglés.
Alternativa 1	Devuelve toda la información de direcciones en inglés.
Alternativa 2	Devuelve la información de la calle, la localidad y el condado en irlandés. Devuelve el resto de información de las direcciones en inglés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Soporte multilingüe para direcciones de Israel

En la siguiente tabla se describen los idiomas que puede especificar para las direcciones de Israel:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en hebreo.
Inglés	Devuelve toda la información de direcciones en inglés.
Alternativa 1	Devuelve toda la información de direcciones en hebreo.
Alternativa 2	Devuelve toda la información de direcciones en inglés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Tenga en cuenta las siguientes reglas y directrices al seleccionar el idioma preferido:

- Para devolver las direcciones en un conjunto de caracteres hebreos, establezca la propiedad Codificación de caracteres preferida en Base de datos.
- Para devolver las direcciones en un conjunto de caracteres latinos o ASCII, establezca la propiedad Codificación de caracteres preferida en un valor LATIN o ASCII.
- Si selecciona un conjunto de caracteres latinos como codificación de caracteres preferida y selecciona hebreo como el idioma preferido, la validación de direcciones realiza la transliteración de las direcciones de hebreo a caracteres latinos. Para obtener los mejores resultados en un conjunto de caracteres latinos, seleccione inglés como idioma preferido.

Soporte multilingüe para direcciones de Macao

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Macao:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en chino.
Alternativa 1	Devuelve toda la información de la dirección en el idioma de la base de datos.
Alternativa 2	Devuelve toda la información de direcciones en inglés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

- Para devolver la dirección en un conjunto de caracteres chinos, establezca la propiedad Codificación de caracteres preferida en Base de datos.
- Para devolver la dirección en un conjunto de caracteres latinos o ASCII, establezca la propiedad Codificación de caracteres preferida en un valor LATIN o ASCII.
- El idioma de los datos de entrada puede afectar al funcionamiento de la opción Conservar entrada en una dirección de Macao. La validación de direcciones identifica el idioma de entrada como portugués cuando los datos de entrada utilizan caracteres ASCII de 7 bits e incluyen un descriptor en portugués.

Soporte multilingüe para Suiza

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Suiza:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve las direcciones en el idioma principal de la región a la que pertenece la dirección. Por ejemplo, la validación de direcciones devuelve una dirección de Zúrich en alemán y una dirección de Ginebra en francés.
Inglés	Devuelve la información de la localidad y la provincia en inglés cuando la base de datos de direcciones de referencia contiene la información en inglés. Devuelve los demás elementos de la dirección en el idioma principal de la región a la que pertenece la dirección. La validación de direcciones devuelve la información de algunas localidades en inglés, como Geneva y Zurich.
Alternativa 1	Devuelve la información de la provincia y la localidad en alemán.
Alternativa 2	Devuelve la información de la provincia y la localidad en francés.

Opción	Descripción
Alternativa 3	Devuelve la información de la provincia y la localidad en italiano.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Nota: La validación de direcciones también devuelve la información de la calle en las direcciones de Biel/Bienne en el idioma alternativo que configure.

Soporte multilingüe para Taiwán

En la tabla siguiente se describen los idiomas que puede especificar para las direcciones de Taiwán:

Opción	Descripción
Base de datos	Valor predeterminado. Devuelve toda la información de direcciones en chino.
Inglés	Devuelve toda la información de direcciones en inglés.
Conservar entrada	Devuelve la información de la dirección en el idioma de entrada. La validación de direcciones conserva el idioma cuando los datos de referencia contienen la información de la dirección en el idioma de entrada. Si la validación de direcciones detecta más de un idioma compatible en la dirección de entrada, devuelve la dirección en el idioma de la base de datos. Si Address Verification no puede devolver un elemento en el idioma de entrada, lo hace en el idioma de la base de datos.

Tenga en cuenta las siguientes reglas y directrices al seleccionar el idioma preferido:

- Para devolver la dirección en un conjunto de caracteres chinos, establezca el parámetro Codificación de caracteres preferida en Base de datos.
- Para devolver la dirección en un conjunto de caracteres latinos o ASCII, establezca la propiedad Codificación de caracteres preferida en un valor LATIN o ASCII.
- El idioma de los datos de entrada puede afectar al funcionamiento de la opción Conservar entrada en una dirección de Taiwán. La validación de direcciones identifica el idioma de entrada como inglés cuando los datos de entrada utilizan caracteres ASCII de 7 bits e incluyen un descriptor en inglés.

Codificación de caracteres preferida

Determina el conjunto de caracteres que la transformación del validador de direcciones usa para los datos de salida.

La siguiente tabla describe las opciones de la propiedad:

Opción	Descripción
ASCII (Simplificado)	Devuelve una dirección en caracteres ASCII.
ASCII (Extendido)	Devuelve una dirección en caracteres ASCII y amplía los caracteres especiales de una dirección. Por ejemplo, Õ se translitera a OE.
Base de datos	Devuelve una dirección en el conjunto de caracteres que utilizan los datos de referencia de las direcciones para el idioma predeterminado. Base de datos es el valor predeterminado.
Latín	Devuelve una dirección en el conjunto de caracteres latinos.
Latín (alt.)	Devuelve una dirección en un conjunto de caracteres latinos alternativo. Por ejemplo, especifique Latín para devolver una dirección de Corea del Sur en la transliteración de romanización revisada. Especifique Latín (alt.) para devolver una dirección de Corea del Sur en la antigua transliteración de ISO/TR 11941.
Administración postal	Devuelve una dirección en la codificación de caracteres que prefiere el servicio postal local para la dirección.
Administración postal (Alt.)	Devuelve una dirección en una codificación de caracteres que aprueba el servicio postal local para la dirección como codificación de caracteres alternativa.
Conservar entrada	Devuelve datos de dirección en el conjunto de caracteres que utiliza la dirección de entrada.

La transformación puede procesar un origen de datos que contiene datos en varios idiomas y conjuntos de caracteres. La transformación convierte todos los datos de entrada en el conjunto de caracteres Unicode UCS-2 y procesa los datos en ese formato. Después de que la transformación procese los datos, convierte los datos de cada registro de direcciones en el conjunto de caracteres que especifique en la propiedad. El proceso se llama transliteración.

La transliteración puede utilizar las representaciones numéricas de cada carácter de un conjunto de caracteres cuando convierte los caracteres para su procesamiento. La transliteración también puede convertir los caracteres fonéticamente cuando no hay representación numérica equivalente de un carácter. Si la transformación del validador de direcciones no puede asignar un carácter a UCS-2, lo convierte en un espacio.

Nota: Si actualiza el idioma preferido o la codificación de caracteres preferida en la transformación, compruebe que el idioma y el código de caracteres que seleccione sean compatibles.

Rangos para expandir

Determina cómo la transformación del validador de direcciones devuelve direcciones sugeridas para una dirección de calle que no especifica un número de domicilio. Use la propiedad cuando la transformación se ejecute en el modo de lista de sugerencias.

La transformación del validador de direcciones lee una dirección de calle parcial o incompleta en el modo de lista de sugerencias. La transformación compara la dirección con los datos de referencia de direcciones y

devuelve todas las direcciones similares para el usuario final. Si la dirección de entrada no contiene un número de domicilio, la transformación puede devolver una o más sugerencias de número de domicilio para la calle. La propiedad Rangos para expandir determina cómo devuelve la transformación las direcciones.

La transformación puede devolver el intervalo de números de domicilio válidos en una sola dirección, o puede devolver una dirección independiente para cada número de domicilio válido. La transformación también puede devolver una dirección para cada número del intervalo del número de domicilio más bajo al más alto de la calle.

La siguiente tabla describe las opciones de la propiedad:

Opción	Descripción
Todos	La validación de direcciones devuelve una dirección sugerida para cada número de domicilio en el rango de posibles de números de domicilio de la calle.
Ninguno	La validación de direcciones devuelve una dirección única que identifica el número de domicilio más bajo y el más alto del intervalo válido para la calle.
Solo con elementos válidos	La validación de direcciones devuelve una dirección sugerida para cada número de domicilio que los datos de referencia de direcciones reconocen como una dirección de entrega.

Nota: El modo de lista de sugerencias puede usar otros elementos de la dirección para especificar el intervalo válido de números de calle. Por ejemplo, un código postal puede identificar la manzana que contiene el buzón de la dirección. La transformación del validador de direcciones puede utilizar el código postal para identificar el número de domicilio más bajo y el más alto de la manzana.

Si la transformación no puede determinar un intervalo de números de domicilio dentro de los límites prácticos, el número de direcciones sugeridas puede alcanzar un tamaño inutilizable. Para restringir el número de direcciones que la propiedad Rangos para expandir genera, habilite la propiedad Expansión de rangos flexible.

Estandarizar direcciones no válidas

Determina si el proceso de validación de direcciones estandariza los valores de datos de una dirección no válida para la entrega. La propiedad se aplica a registros de direcciones que devuelven un estado de código de coincidencia en el intervalo de I1 a I4.

Cuando se estandarizan los datos, aumenta la probabilidad de que un proceso de datos de nivel inferior devuelva resultados precisos. Por ejemplo, una asignación de análisis duplicado puede devolver una mejor puntuación de coincidencia para dos registros de dirección que tienen elementos de dirección comunes con el mismo formato.

La validación de direcciones puede estandarizar los siguientes elementos de dirección:

- Elementos de sufijo de calle, como avenida y paseo.
- Elementos anteriores y posteriores de la dirección, como norte, sur, este y oeste.
- Elementos de servicios de entrega, como apartado de correos.
- Elementos secundarios del edificio, como apartamento, planta y suite.
- Nombres de estado o provincia. La estandarización devuelve la forma abreviada de los nombres.

En la siguiente tabla se describen las opciones de la propiedad:

Opción	Descripción
Desactivada	La validación de direcciones no corrige errores de datos. Opción predeterminada.
Activada	La validación de direcciones corrige errores de datos.

Uso de parámetros

Puede asignar un parámetro para especificar la directiva de estandarización respecto a los errores de datos. Introduzca OFF u ON como valor de parámetro. Introduzca el valor en mayúsculas.

Nivel de seguimiento

Define la cantidad de detalles que se incluyen en el registro.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Informes de certificación

Puede generar un informe que describa el estado de las direcciones que envía para validación en modo certificado. El informe verifica que el conjunto de direcciones cumple con los estándares de certificación que haya definido el proveedor de servicios de correo.

La transformación del validador de direcciones genera informes para los siguientes estándares:

Address Machine Approval System (AMAS)

Australia Post define el estándar de certificación AMAS.

Coding Accuracy Support System (CASS)

USPS define el estándar de certificación CASS.

SendRight

New Zealand Post define el estándar de certificación SendRight.

Software Evaluation and Recognition Program (SERP)

Canada Post define el estándar de certificación SERP.

Nota: Puede utilizar la transformación del validador de direcciones para certificar registros de direcciones francesas de acuerdo con el estándar de certificación del Servicio de administración de direcciones nacional (SNA). No obstante, no genera un informe para la certificación de direcciones de SNA en la transformación del validador de direcciones.

Al suministrar los elementos de correo que cumplen con los estándares de certificación al proveedor de correo, envía el informe de certificación con el conjunto de direcciones. El informe incluye datos sobre la organización. Los datos pueden introducirse cuando se configura la transformación del validador de direcciones. La transformación agrega los datos de la organización al archivo de informe.

Campos de informe AMAS

Cuando se configura un informe AMAS, se proporciona información acerca de la organización que envía el conjunto de registros de direcciones certificadas a Australia Post. Guarde o imprima el informe, e incluya el informe con los registros de direcciones que envíe a Australia Post.

Utilice la vista **Informes** para introducir la información.

La siguiente tabla describe la información que se especifica:

Campo	Descripción
Nombre del archivo de informe	Nombre y ubicación del informe creado por la validación de direcciones. De forma predeterminada, el validador de direcciones crea el informe en el directorio <code>bin</code> del equipo del servicio de integración de datos. Para grabar el archivo de informe en otra ubicación del equipo del servicio de integración de datos, especifique la ruta de acceso al archivo y el nombre del archivo. Puede especificar una ruta de acceso completa o una ruta de acceso relativa. La ruta de acceso relativa utiliza el directorio <code>bin</code> como directorio raíz. El directorio especificado debe existir antes de ejecutar la asignación de validación de direcciones.
Nombre de lista de direcciones	Nombre del conjunto de registros de direcciones que envíe a Australia Post.
Nombre de procesador de lista	Nombre de la organización que envía el conjunto de registros de direcciones.
Nombre del administrador/ propietario de la lista	Nombre del administrador o propietario de los datos de direcciones de la organización.
Número de teléfono	Número de teléfono de contacto de la organización que envía el conjunto de registros de direcciones.
Dirección	Dirección de la organización que envía el conjunto de registros de direcciones.

TEMAS RELACIONADOS

- [“Cómo definir un informe de certificación” en la página 136](#)

Campos de informe CASS

Cuando se configura un informe CASS, se proporciona información acerca de la organización que envía el conjunto de registros de direcciones certificadas a USPS. Guarde o imprima el informe, e incluya el informe con los registros de direcciones que envíe a USPS.

Utilice la vista **Informes** para introducir la información.

La siguiente tabla describe la información que se especifica:

Campo	Descripción
Nombre del archivo de informe	El nombre y la ubicación del informe creado por la validación de direcciones. De forma predeterminada, el validador de direcciones crea el informe en el directorio <code>bin</code> del equipo del servicio de integración de datos. Para escribir el archivo de informe en otra ubicación del equipo del servicio de integración de datos, especifique la ruta de acceso al archivo y el nombre de archivo. Puede especificar una ruta de acceso completa o una ruta de acceso relativa. La ruta de acceso relativa utiliza el directorio <code>bin</code> como directorio raíz. El directorio especificado debe existir antes de ejecutar la asignación de validación de direcciones.
Nombre/ID de lista	Nombre o número de identificación de la lista de direcciones que se envía al proveedor de servicios de correo.
Nombre de procesador de lista	Nombre de la organización que realiza la validación de direcciones.
Nombre/Dirección	Nombre y dirección postal de la organización que realiza la validación de direcciones.

Nota: La transformación del validador de direcciones lee los metadatos sobre los archivos de datos de referencia de CASS desde el archivo de datos de referencia interactivo y por lotes para Estados Unidos. Para generar un informe CASS, la transformación debe leer los archivos de datos de referencia CASS actuales y el archivo de datos de referencia interactivo y por lotes actual.

TEMAS RELACIONADOS

- [“Cómo definir un informe de certificación” en la página 136](#)

Informe SendRight

Cuando se configura un informe SendRight, se proporciona información acerca de la organización que envía el conjunto de registros de direcciones certificadas a New Zealand Post. Guarde o imprima el informe, e incluya el informe con los registros de direcciones que envíe a New Zealand Post.

Utilice la vista **Informes** para introducir la información.

La siguiente tabla describe la información que se especifica:

Campo	Descripción
Nombre del cliente	Nombre de la organización que envía el conjunto de registros de direcciones.
Número de NZP del cliente	Número de la cuenta de New Zealand Post de la organización que envía el conjunto de registros de direcciones. Si un servicio de correo envía los registros en nombre de la organización, especifique el número de identificación de transporte (TPID) del servicio de correo.
Base de datos del cliente	Nombre del archivo que contiene el conjunto de registros de direcciones. La transformación del validador de direcciones crea el informe en la ubicación que se especifique en el servicio de administración de contenido. Use Administrator Tool para definir la ubicación.
Dirección de cliente	Dirección de la organización que envía el conjunto de registros de direcciones.

TEMAS RELACIONADOS

- [“Cómo definir un informe de certificación” en la página 136](#)

Campos de informe SERP

Cuando se configura un informe SERP, se proporciona información acerca de la organización que envía el conjunto de registros de direcciones certificadas a Canada Post. Guarde o imprima el informe, e incluya el informe con los registros de direcciones que envíe al servicio de correos de Canadá.

Utilice la vista **Informes** para introducir la información.

La siguiente tabla describe la información que se especifica:

Campo	Descripción
Nombre del archivo de informe	<p>El nombre y la ubicación del informe creado por la validación de direcciones. De forma predeterminada, el validador de direcciones crea el informe en el directorio <code>bin</code> del equipo del servicio de integración de datos.</p> <p>Para escribir el archivo de informe en otra ubicación del equipo del servicio de integración de datos, especifique la ruta de acceso al archivo y el nombre de archivo. Puede especificar una ruta de acceso completa o una ruta de acceso relativa. La ruta de acceso relativa utiliza el directorio <code>bin</code> como directorio raíz.</p> <p>El directorio especificado debe existir antes de ejecutar la asignación de validación de direcciones.</p>
Número de cliente de CPC	El número de cliente que Canada Post Corporation emite para la organización que realiza la validación de direcciones.
Nombre/dirección del cliente	El nombre y la dirección de la organización que realiza la validación de direcciones.

TEMAS RELACIONADOS

- [“Cómo definir un informe de certificación” en la página 136](#)

Cómo configurar una transformación del validador de direcciones.

Utilice una transformación del validador de direcciones para validar y mejorar la calidad de los datos de direcciones postales.

La transformación del validador de direcciones lee datos de referencia de direcciones. Compruebe que Developer Tool puede acceder a los archivos de datos de referencia necesarios.

1. Abra la transformación.
2. Haga clic en la vista **Configuración general** y configure las propiedades generales.
3. Haga clic en la vista **Plantillas** para añadir puertos de entrada y de salida.
4. Haga clic en la vista **Informes** para generar informes para una certificación de dirección de servicio postal.

5. Haga clic en la vista **Avanzadas** para configurar las propiedades avanzadas de validación de direcciones.
6. Conecte los puertos de entrada y de salida.
Nota: Conecte los puertos de entrada que no desee que valide la transformación del validador de direcciones en el grupo de puertos de entrada **Transferencia**.

Cómo añadir puertos a la transformación del validador de direcciones

Utilice la vista **Plantillas** para añadir puertos a la transformación del validador de direcciones.

1. Haga clic en la vista **Plantillas**.
2. Expanda una plantilla.
 - Elija la plantilla **Modelo básico** para añadir campos de dirección comunes.
 - Elija la plantilla **Modelo avanzado** para añadir campos de dirección especializados.
3. Expanda el grupo de puertos de entrada que se corresponde con el formato de los datos de entrada. Los grupos de puertos de entrada son **Discreto**, **Varias líneas** e **Híbrido**.
4. Seleccione puertos de entrada.
Sugerencia: Haga clic en la tecla **CTRL** para seleccionar varios puertos.
5. Haga clic con el botón derecho en los puertos y seleccione **Añadir puerto a transformación**.
6. Expanda el grupo de puertos de salida que contiene los campos que necesita.
7. Haga clic con el botón derecho en los puertos y seleccione **Añadir puerto a transformación**.
8. Para añadir puertos de transferencia a columnas que no se desean validar, haga clic en la ficha **Puertos**, seleccione el grupo de puertos de entrada **Transferencia** y haga clic en **Nuevo**.

Cómo crear plantillas definidas por el usuario

Cree plantillas para agrupar los puertos de dirección que prevé reutilizar.

Puede crear plantillas personalizadas seleccionando puertos desde las plantillas Básica y Avanzada. Puede seleccionar las plantillas personalizadas cuando se crean transformaciones del validador de direcciones subsiguientes.

Nota: Las plantillas no son objetos del repositorio. Las plantillas residen en el equipo que se utiliza para crearlas.

1. Seleccione la vista **Plantillas**.
2. Haga clic en **Nueva**.
3. Especifique un nombre para la plantilla..
4. Expanda la plantilla **Modelo básico** o **Modelo avanzado** y seleccione los puertos que necesite.
5. Haga clic en **Aceptar**.

Cómo definir modelos del validador de direcciones

Los modelos del validador de direcciones definen los puertos de entrada y de salida para las transformaciones del validador de direcciones.

Las transformaciones del validador de direcciones no contienen puertos de entrada y de salida predeterminados. Sin embargo, puede definir un modelo para especificar los puertos de entrada y de salida que utilizan las transformaciones del validador de direcciones.

Nota: Los modelos no son objetos del repositorio. Los modelos residen en el equipo que se utiliza para crearlos.

Para definir un modelo del validador de direcciones, realice los siguientes pasos:

1. Seleccione la vista **Plantillas**.
2. Expanda la plantilla **Modelo básico** o **Modelo avanzado** y seleccione los puertos que necesite.
3. Seleccione **Crear modelo de validador de direcciones predeterminado mediante los puertos seleccionados**.
4. Para restablecer el modelo y quitar todos los puertos, seleccione **Borrar modelo de validador de direcciones predeterminado**.

Cómo definir un informe de certificación

Cuando se define un informe de certificación en la transformación del validador de direcciones, se configuran opciones en las vistas **Configuración general** e **Informes**.

1. En la vista **Configuración general**, establezca la opción **Modo** en **Certificado**.
2. En la vista **Informes**, seleccione el tipo de informes que generar. Se pueden seleccionar los siguientes tipos de informe:

Opción	Descripción
Informe AMAS	Contiene información que Australia Post requiere para procesar el conjunto de registros.
Informe CASS	Contiene información que USPS requiere para procesar el conjunto de registros.
Informe SendRight	Contiene información que New Zealand Post requiere para procesar el conjunto de registros.
Informe SERP	Contiene información que Canada Post requiere para procesar el conjunto de registros.

3. Especifique los detalles del informe para cada tipo de informe seleccionado.

Envíe el archivo de informe al proveedor de servicios de correo con la lista de registros de dirección que haya validado con la transformación del validador de direcciones.

TEMAS RELACIONADOS

- [“Campos de informe AMAS” en la página 132](#)
- [“Campos de informe CASS” en la página 132](#)
- [“Informe SendRight” en la página 133](#)
- [“Campos de informe SERP” en la página 134](#)

Transformación del validador de direcciones en un entorno no nativo

El procesamiento de transformaciones del validador de direcciones en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación del validador de direcciones en el motor de Blaze

La transformación del validador de direcciones es compatible con las siguientes restricciones:

- La transformación del validador de direcciones no puede generar un informe de certificación.
- La transformación del validador de direcciones no se puede validar si está configurada para ejecutarse en el modo interactivo o en el modo de lista de sugerencias.

Transformación del validador de direcciones en el motor de Spark

La transformación del validador de direcciones es compatible con las siguientes restricciones:

- La transformación del validador de direcciones no puede generar un informe de certificación.
- La transformación del validador de direcciones no se puede validar si está configurada para ejecutarse en el modo interactivo o en el modo de lista de sugerencias.
- Una asignación que se ejecuta con una transformación del validador de direcciones omite el valor de la propiedad avanzada **Instancias de ejecución** en la transformación.

Transformación del validador de direcciones en una asignación de transmisión

Las asignaciones de transmisión tienen las mismas reglas de procesamiento que las asignaciones por lotes en el motor de Spark.

Transformación del validador de direcciones en el motor de Databricks Spark

La transformación del validador de direcciones es compatible con las siguientes restricciones:

- La transformación del validador de direcciones no puede generar un informe de certificación.
- La transformación del validador de direcciones no se puede validar si está configurada para ejecutarse en el modo interactivo o en el modo de lista de sugerencias.
- Una asignación que se ejecuta con una transformación del validador de direcciones omite el valor de la propiedad avanzada **Instancias de ejecución** en la transformación.

CAPÍTULO 5

Transformación de agregación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de agregación, 138](#)
- [Transformaciones de agregación en asignaciones dinámicas, 139](#)
- [Cómo desarrollar una transformación de agregación, 139](#)
- [Puertos de transformación de agregación, 139](#)
- [Expresiones de agregado, 140](#)
- [Agrupar por puertos, 142](#)
- [Memorias caché de agregación, 145](#)
- [Entrada ordenada para una transformación de agregación, 145](#)
- [Propiedades avanzadas de la transformación de agregación, 147](#)
- [Cómo crear una transformación de agregación no reutilizable, 148](#)
- [Cómo crear una transformación de agregación no reutilizable, 148](#)
- [Consejos para transformaciones de agregación, 149](#)
- [Solución de problemas de transformaciones de agregación, 149](#)
- [Transformación de agregación en un entorno no nativo, 150](#)

Resumen de la transformación de agregación

Configure una transformación de agregación para realizar cálculos de agregado (como promedios y sumas) en grupos de datos. Puede utilizar una transformación de agregación para eliminar filas duplicadas. La transformación de agregación es una transformación activa.

La transformación de agregación difiere de la transformación de expresión porque la primera se puede configurar para realizar cálculos en grupos de datos. Una transformación de expresión devuelve un resultado por cada una de las filas.

Por ejemplo, puede calcular el promedio de salarios de los empleados de cada departamento de una organización. Configure un grupo por número de departamento. Configure una expresión para calcular el promedio de salarios y para devolver el resultado correspondiente a cada número de departamento único.

Utilice el lenguaje de transformación para crear expresiones de agregado.

El Servicio de integración de datos realiza cálculos de agregado mientras lee y almacena los datos en una memoria caché de agregado. Puede ordenar los datos de entrada para aumentar el rendimiento. El Servicio de integración de datos no crea la memoria caché si ordena los datos de entrada.

Transformaciones de agregación en asignaciones dinámicas

Puede utilizar una transformación de agregación en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados.

Puede realizar las siguientes tareas para configurar una transformación de agregación en una asignación dinámica:

Hacer referencia a un puerto dinámico o a un puerto generado como una columna Agrupar por.

Puede incluir puertos dinámicos y puertos generados como columnas Agrupar por. Si especifica un puerto dinámico como una columna de agrupación, está especificando que todos los puertos generados en dicho puerto dinámico son columnas de agrupación. Por este motivo, no puede especificar un puerto generado como columna de agrupación si se ha especificado el puerto dinámico primario como columna de agrupación. Si hace referencia a un puerto generado y dicho puerto no existe en tiempo de ejecución, la asignación fallará. Puede parametrizar las columnas Agrupar por para indicar qué columnas se van a agrupar en tiempo de ejecución.

Hacer referencia a un puerto generado en la expresión de agregado.

Puede hacer referencia a un puerto generado en la expresión de agregado. Si el puerto no existe en tiempo de ejecución, la asignación fallará. No se puede hacer referencia a un puerto dinámico en la expresión de agregado.

Crear la expresión de agregado en un puerto de salida.

No se puede crear una expresión de agregado en un puerto generado o un puerto dinámico. La expresión de agregado no puede ser una expresión dinámica.

Parametrizar los valores en la expresión de agregado.

Puede incluir parámetros en la expresión de agregado. Sin embargo, no puede utilizar un parámetro de expresión o un parámetro de puerto en dicha expresión.

Cómo desarrollar una transformación de agregación

Al crear una transformación de agregación, defina la expresión que se va a ejecutar en cada fila. Defina una lista de agrupación por puertos en función de la cual se deben devolver los resultados.

Haga lo siguiente para crear una transformación de agregación:

1. Defina la transformación y cree los puertos.
2. Configure una expresión de agregado en un puerto de variable o de salida.
3. Defina un grupo de puertos en función del cual se deben devolver los resultados.

Puertos de transformación de agregación

Una transformación de agregación tiene varios tipos de puerto que puede utilizar para configurar grupos y agregar expresiones.

La vista Puertos de una transformación de agregación tiene los siguientes campos:

Nombre

Nombre del puerto.

Tipo

Tipo de datos del puerto.

Precisión

Longitud del campo.

Escala

Número de posiciones a la derecha del punto decimal de los datos numéricos.

Entrada

Indica que los datos proceden de una transformación de nivel superior.

Salida

Indica que el puerto devuelve el valor de una expresión. La expresión puede incluir expresiones de no agregado y cláusulas condicionales. Puede crear varios puertos de salida agregados.

Transferencia

Indica que el puerto es un puerto de entrada-salida que devuelve los datos inalterados.

Variable

Indica que el puerto puede almacenar valores o cálculos para utilizarlos en una expresión. Los puertos de variable no pueden ser ni puertos de entrada ni de salida. Solamente pasan datos dentro de la transformación.

Expresión

Expresión para agregar filas o grupos de filas.

Valor predeterminado

Valor predeterminado de un puerto con valores nulos o no válidos.

Reglas de entrada

Conjunto de reglas que filtra los puertos que se van a incluir en la transformación o excluir de ella en función de los nombres de puerto o el tipo de datos. Configure reglas de entrada al definir puertos dinámicos.

Expresiones de agregado

Configure expresiones de agregado en los puertos de variable o en los puertos de salida de una transformación de agregación. No puede especificar una expresión si el puerto es un puerto de entrada, un puerto de transferencia o un puerto dinámico.

Una expresión de agregado puede incluir funciones y cláusulas condicionales que no sean funciones de agregado. La función de agregado puede incluir también una función de agregado anidada en otra función de agregado, como la siguiente:

```
MAX( COUNT( ITEM ) )
```

El resultado de una expresión de agregado varía según la agrupación por puertos de la transformación. Por ejemplo, la siguiente expresión de agregado halla la cantidad total de artículos vendidos:

```
SUM( QUANTITY )
```

Sin embargo, si se utiliza la misma expresión y se agrupa por el puerto ITEM, el Servicio de integración de datos devolverá la cantidad total por artículo.

Se puede crear una expresión de agregado en cualquier puerto de salida y usar varios puertos de agregado en una transformación.

Funciones de agregado

Configure funciones de agregación dentro de una transformación de agregación. Puede anidar una función de agregado dentro de otra función de agregado.

El lenguaje de la transformación incluye las siguientes funciones de agregado:

- ANY
- AVG
- COLLECT_LIST
- COLLECT_MAP
- COUNT
- FIRST
- LAST
- MAX (Fecha)
- MAX (Número)
- MAX (Cadena)
- MEDIAN
- MIN (Fecha)
- MIN (Número)
- MIN (Cadena)
- PERCENTILE
- STDDEV
- SUM
- VARIANCE

Si usa un puerto en una expresión de la transformación de agregación, pero no usa ese puerto en una función de agregado, el servicio de integración de datos usa la última fila del puerto para procesar la expresión.

Por ejemplo, puede crear una transformación de agregación que contenga los puertos COMMISSIONS y SALARY. El puerto SALARY es un puerto group-by.

En un puerto de salida se podría usar la siguiente expresión:

```
SUM(COMMISSIONS)
```

El servicio de integración de datos procesa la función de agregado y devuelve la suma de los valores en el puerto COMMISSIONS en el puerto de salida.

La expresión se podría modificar a la siguiente expresión:

```
SUM(COMMISSIONS) + COMMISSIONS
```

Para procesar la expresión, el servicio de integración de datos devuelve la suma de los valores en el puerto COMMISSIONS y suma el valor de la última fila en el puerto COMMISSIONS al valor de devolución en el puerto de salida.

En un puerto de salida diferente, se podría usar la siguiente expresión:

```
SUM(COMMISSIONS) + SALARY
```

Para procesar la expresión, el servicio de integración de datos devuelve la suma de los valores en el puerto `COMMISSIONS` y suma el valor de la última fila del puerto `SALARY` al valor de devolución en el puerto de salida. Fíjese en que los valores de cada fila del puerto `SALARY` son iguales porque el puerto `SALARY` es un puerto `group-by`.

Funciones de agregado anidadas

Se pueden incluir varias funciones de un nivel o varias funciones anidadas en distintos puertos de la transformación de agregación

No se pueden incluir tanto funciones de un nivel y funciones anidadas en una función de agregación. Por lo tanto, si una transformación de agregación contiene una función de un nivel en cualquier puerto de salida, no se puede utilizar una función anidada en ningún otro puerto de esa transformación. Cuando se incluyen funciones de un nivel y funciones anidadas en la misma transformación de agregación, Developer Tool marca la asignación o el mapplet como no válidos. Si necesita crear funciones de un nivel y funciones anidadas al mismo tiempo, cree transformaciones de agregación individuales.

Cláusulas condicionales en expresiones de agregado

Utilice cláusulas condicionales en la expresión de agregado para reducir el número de filas utilizadas en la agregación. La cláusula condicional puede ser cualquier cláusula cuya evaluación da como resultado `TRUE` o `FALSE`.

Por ejemplo, utilice la siguiente expresión para calcular el total de las comisiones de los empleados que hayan sobrepasado su cuota trimestral:

```
SUM( COMMISSION, COMMISSION > QUOTA )
```

Agrupar por puertos

Puede definir grupos de filas para agregarlos en lugar de ejecutar una agregación en todos los datos de entrada. Por ejemplo, puede calcular el total de ventas de la empresa o buscar el total de ventas agrupado por región.

Para definir un grupo para la expresión de agregado, seleccione los puertos de entrada, entrada/salida, salida y variables apropiados en la transformación de agregación. Puede seleccionar varias agrupaciones por puerto para crear un grupo nuevo para cada combinación única. Luego, el servicio de integración de datos realiza la agregación definida para cada grupo.

Cuando se agrupan valores, el servicio de integración de datos produce una fila para cada grupo. Si no se agrupan valores, el servicio de integración de datos devuelve una fila para todas las filas de entrada. El servicio de integración de datos devuelve la última fila de cada grupo con el resultado de la agregación. Puede especificar que se devuelva una fila específica. Por ejemplo, si utiliza la función de agregación `FIRST`, el servicio de integración de datos devolverá la primera fila.

Cuando se seleccionan varias agrupaciones por puerto en la transformación de agregación, el servicio de integración de datos utiliza el orden de los puertos para averiguar el orden en que se realizará la agrupación. El orden de grupos puede influir en los resultados. Ordene los grupos por puertos para procurar que la agrupación sea adecuada. Puede cambiar el orden de puertos después de seleccionarlos puertos en el grupo.

Por ejemplo, podría crear un puerto de salida denominado Price_Out. La expresión de Price_Out es SUM (Qty * Price). Puede definir los elementos Store_ID e Item como los grupos por puertos. La transformación devolverá el precio de total de cada artículo por almacén.

Las filas de entrada pueden contener los siguientes datos:

Store_ID	Item	Qty	Price
101	battery	3	2.99
101	battery	1	3.19
101	battery	2	2.59
101	AAA	2	2.45
201	battery	1	1.99
201	battery	4	1.59
301	battery	1	2.45

El servicio de integración de datos lleva a cabo el cálculo del agregado en los siguientes grupos únicos:

Store_Id	Item
101	battery
101	AAA
201	battery
301	battery

El servicio de integración de datos devuelve los elementos Store_ID, Item, Qty y Price de la última fila con la suma de (Price * Qty) de cada artículo por almacén:

Store_ID	Item	Qty	Price	Price_Out
101	battery	2	2.59	17.34
101	AAA	2	2.45	4.90
201	battery	4	1.59	8.35
301	battery	1	2.45	2.45

Configurar la agrupación por puertos

Defina la agrupación por puertos en la ficha **Agrupar por** de vista **Propiedades** de la transformación.

La siguiente imagen muestra la ficha Agrupar por:

La ficha Agrupar por contiene las siguientes opciones:

Especificar por

Seleccione **Valor** o **Parámetro**. Seleccione **Valor** para usar los nombres de puerto. Elija **Parámetro** para utilizar un parámetro de lista de puertos.

Añadir

Se acepta el nombre del puerto que escriba manualmente. Debe escribir un nombre de puerto válido antes de hacer clic en **Añadir**.

Elegir

Haga clic en **Elegir** para seleccionar los puertos que quiera añadir al grupo. Developer tool proporciona una lista de los puertos de la transformación para su selección.

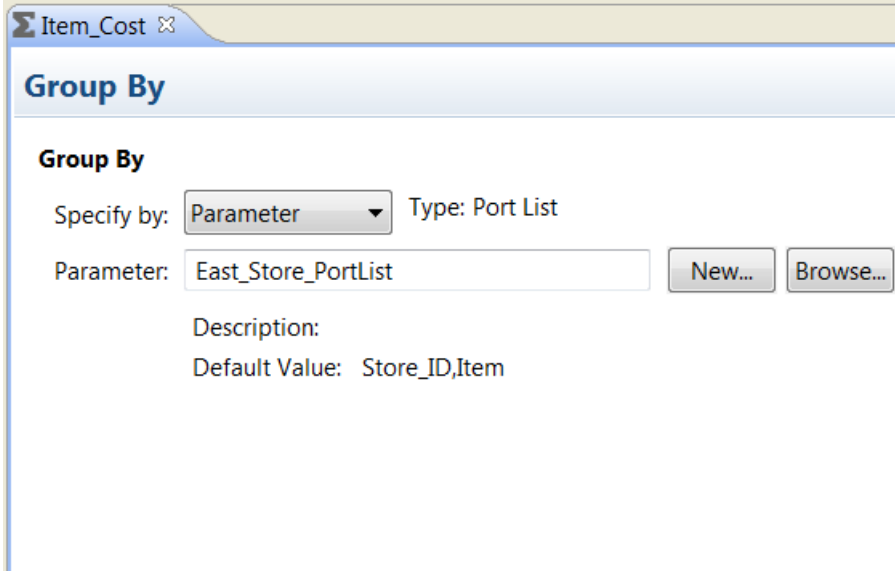
Subir y Bajar

Puede cambiar el orden de puertos del grupo. Seleccione el nombre de puerto y haga clic en uno de los botones para subirlo o bajarlo en la ordenación.

Parámetros de la agrupación por puertos

Puede configurar un parámetro de lista de puertos que contenga uno o más puertos que quiera incluir en un grupo. Para crear un parámetro de lista de puertos, seleccione puertos de una lista de puertos en la transformación.

La siguiente imagen muestra la ficha **Agrupar por** cuando utiliza un parámetro para identificar los puertos en el grupo:



Puede buscar un parámetro de lista de puertos o hacer clic en **Nuevo** para crear un parámetro de lista de puertos. Si decide crear un parámetro de lista de puerto, puede seleccionar los puertos de una lista de puertos en la transformación.

Valores predeterminados de Agrupar por puertos

El Servicio de integración de datos no crea un grupo cuando una agrupación por puertos contiene valores nulos. Puede definir un valor predeterminado por cada puerto en el grupo que reemplace los valores de entrada nulos. De este modo, el Servicio de integración de datos podrá incluir las filas en los totales de la agregación.

Expresiones de no agregado

Utilice expresiones de no agregado en agrupar por puertos para modificar o reemplazar grupos.

Por ejemplo, si desea reemplazar 'pila AAA' antes de la agrupación, puede crear un puerto de salida agrupar por llamado CORRECTED_ITEM utilizando la siguiente expresión:

```
IIF( ITEM = 'AAA battery', battery, ITEM )
```

Memorias caché de agregación

Cuando se ejecuta una asignación que utiliza una transformación de agregación, el Servicio de integración de datos crea una memoria caché de índice y una memoria caché de datos en la memoria para ejecutar la transformación. Si el Servicio de integración de datos necesita más espacio del que hay disponible en la memoria caché, almacena los datos desbordados en archivos de memoria caché.

El Servicio de integración de datos crea las siguientes memorias caché para la transformación de agregación:

- Memoria caché de índice que almacena valores de grupo tal y como están configurados en los puertos de agrupación.
- Memoria caché de datos que almacena los cálculos basados en los puertos de agrupación.

El Servicio de integración de datos no utiliza la memoria caché para ejecutar una transformación de agregación con los puertos ordenados. No necesita configurar la memoria caché para las transformaciones de agregación que utilizan puertos ordenados.

Entrada ordenada para una transformación de agregación

Puede mejorar el rendimiento de una transformación de agregación mediante la opción de entrada ordenada.

Cuando se utiliza la entrada ordenada, el servicio de integración de datos da por sentado que todos los datos están ordenados por grupo y realiza cálculos de agregado a medida que lee las filas de un grupo. En caso necesario, el servicio de integración de datos almacena la información sobre el grupo en la memoria. Para utilizar la opción Entrada ordenada, debe pasar datos ordenados a la transformación de agregación. Si utiliza la entrada ordenada, la transformación de agregación proporciona una salida ordenada.

Si no se utiliza la entrada ordenada, el servicio de integración de datos realiza los cálculos de agregados a medida que lee las filas. Puesto que los datos no están ordenados, el servicio de integración de datos almacena datos para cada grupo hasta que lee el origen por completo para asegurarse de que todos los cálculos de agregado se realizan de forma precisa.

Por ejemplo, una transformación de agregación tiene los puertos agrupar por STORE_ID e ITEM con la opción de entrada ordenada seleccionada. Cuando se pasan los siguientes datos a través de la agregación, el servicio de integración de datos realiza una agregación por cada tres filas en el grupo 101/batería cuando encuentra el grupo 201/batería:

STORE_ID	ITEM	QTY	PRICE
101	'battery'	3	2.99

STORE_ID	ITEM	QTY	PRICE
101	'battery'	1	3.19
101	'battery'	2	2.59
201	'battery'	4	1.59
201	'battery'	1	1.99

Si utiliza la entrada ordenada y no ordena los datos previamente de manera correcta, el servicio de integración de datos genera un error en la ejecución de la asignación.

Condiciones de entrada ordenada

Hay ciertas condiciones que impiden que se utilice la entrada ordenada.

No se puede usar la entrada ordenada si se cumple una de las siguientes condiciones:

- La expresión de agregado contiene funciones de agregado anidadas.
- Los datos de origen dependen de los datos.

Cuando se cumple alguna de estas condiciones, el servicio de integración de datos procesa la transformación como si no se utilizara la entrada ordenada.

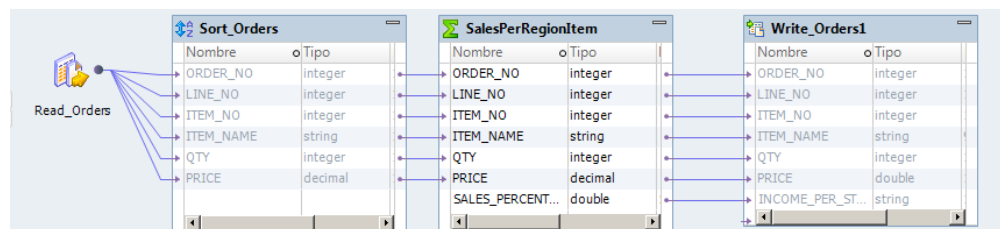
Cómo ordenar datos en una transformación de agregación

Para utilizar una entrada ordenada, pase datos ordenados por una transformación de agregación.

Debe ordenar los datos por los puertos agrupar por del agregador en el orden que aparecen en la transformación de agregación.

En el caso de entradas relacionales y de archivo sin formato, utilice la transformación de ordenación para ordenar los datos de la asignación antes de pasarlos a la transformación de agregación. Puede poner la transformación de ordenación en cualquier lugar de la asignación antes del agregador en el caso de que ninguna transformación cambie el orden de los datos ordenados. La agrupación por columnas en la transformación de agregación debe estar en el mismo orden que aparece en la transformación de ordenación.

La siguiente asignación muestra una transformación de ordenación configurada para ordenar los datos de origen en orden ascendente según ITEM_NO:



La transformación de ordenación ordena los datos del siguiente modo:

ITEM_NO	ITEM_NAME	QTY	PRICE
345	Soup	4	2.95
345	Soup	1	2.95

ITEM_NO	ITEM_NAME	QTY	PRICE
345	Soup	2	3.25
546	Cereal	1	4.49
546	Cereal	2	5.25

Con la entrada ordenada, la transformación de agregación devuelve los siguientes resultados:

ITEM_NAME	QTY	PRICE	INCOME_PER_ITEM
Cereal	2	5.25	14.99
Soup	2	3.25	21.25

Propiedades avanzadas de la transformación de agregación

Configure propiedades para ayudar a determinar cómo el Servicio de integración de datos procesa los datos para la transformación de agregación.

Configure las siguientes propiedades avanzadas de una transformación de agregación:

Directorio de la memoria caché

Directorio en el que el Servicio de integración de datos crea los archivos de memoria caché de índice y de datos. Compruebe que el directorio existe y que tiene suficiente espacio en disco para los archivos de memoria caché.

Introduzca varios directorios separados por punto y coma para aumentar el rendimiento durante la creación de particiones en la memoria caché. La creación de particiones en la memoria caché crea una memoria caché independiente para cada partición que procesa la transformación.

El valor predeterminado es el parámetro del sistema CacheDir. Para esta propiedad, puede configurar otro parámetro del sistema o definido por el usuario.

Tamaño de la memoria caché de datos

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de datos para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Tamaño de la memoria caché de índice

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de índice para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Entrada ordenada

Indica que los datos de entrada se ordenan previamente por grupos. Seleccione esta opción solamente si la asignación pasa datos ordenados a la transformación de agregación.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

TEMAS RELACIONADOS

- [“Tamaño de la memoria caché” en la página 74](#)

Cómo crear una transformación de agregación no reutilizable

Cree una transformación de unión no reutilizable para utilizarla en varias asignaciones o mapplets.

1. Seleccione un proyecto o una carpeta en la vista **Object Explorer**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de agregación.
4. Haga clic en **Siguiente**.
5. Especifique un nombre para la transformación.
6. Haga clic en **Finalizar**.
La transformación aparece en el editor.
7. Haga clic en el botón **Nueva** para añadir un puerto a la transformación.
8. Edite el puerto para definir el nombre, el tipo de datos y la precisión.
9. Determine el tipo de cada puerto: entrada, salida, transferencia o variable.
10. Haga clic en el campo Expresión para configurar expresiones de agregado para puertos de salida. Puede seleccionar puertos y parámetros para definir la expresión de agregados.
11. Haga clic en la vista **Avanzadas** y edite las propiedades de la transformación.

Cómo crear una transformación de agregación no reutilizable

Cree una transformación de agregación no reutilizable como parte de una asignación o de un mapplet.

1. En una asignación o en un mapplet, arrastre una transformación de agregación desde la paleta Transformación hasta el editor.
A continuación, la transformación aparece en el editor.
2. En la vista **Propiedades**, edite el nombre y la descripción de la transformación.

3. En la ficha **Puertos**, haga clic en el botón **Nuevo** para añadir puertos a la transformación.
4. Edite los puertos para definir el nombre, el tipo de datos y la precisión.
5. Determine el tipo de cada puerto: entrada, salida, transferencia o variable.
6. Configure expresiones de agregado para puertos de salida.
7. En la vista **Avanzadas**, edite las propiedades de la transformación.

Consejos para transformaciones de agregación

Puede seguir estos consejos para utilizar las transformaciones de agregación de forma más eficaz.

Utilizar la entrada ordenada para reducir el uso de memorias cachés de agregado.

La entrada ordenada reduce la cantidad de datos que se guardan en la memoria caché durante la ejecución de una asignación y aumenta el rendimiento. Utilice esta opción con la transformación de ordenación para pasar datos ordenados a la transformación de agregación.

Limitar el número de puertos de entrada/salida o de salida conectados.

Limite el número de puertos de entrada/salida o de salida conectados para reducir la cantidad de datos que la transformación de agregación almacena en la memoria caché de datos.

Filtrar los datos antes de agregarlos.

Si se utiliza una transformación de filtro en la asignación, colóquela antes de la transformación de agregación para reducir toda agregación innecesaria.

Solo una transformación de agregación ordenada proporciona una salida ordenada.

Si utiliza una entrada no ordenada y desea producir una salida ordenada, tiene que utilizar una transformación de ordenación después de la transformación de agregación.

Solución de problemas de transformaciones de agregación

Se pueden solucionar los problemas de transformaciones de agregación.

He seleccionado la entrada ordenada, pero la asignación tarda lo mismo que antes.

No se puede usar la entrada ordenada si se cumple una de las siguientes condiciones:

- La expresión de agregado contiene funciones de agregado anidadas.
- Los datos de origen dependen de los datos.

Cuando se cumple alguna de estas condiciones, el servicio de integración de datos procesa la transformación como si no se utilizara la entrada ordenada.

Una asignación con una transformación de agregación ralentiza el rendimiento.

Es posible que el servicio de integración de datos esté paginando en el disco. En las propiedades de la transformación, puede aumentar el rendimiento incrementando los tamaños de la memoria caché de índices y datos.

Transformación de agregación en un entorno no nativo

El procesamiento de transformaciones de agregación en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de agregación en el motor de Blaze

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Funciones de agregado

Si usa un puerto en una expresión en la transformación de agregación, pero no usa ese puerto en una función de agregado, el motor en tiempo de ejecución podría usar cualquier fila del puerto para procesar la expresión.

Puede que la fila que utiliza el motor en tiempo de ejecución no sea la última fila del puerto. El procesamiento es distribuido y, por tanto, es posible que el motor en tiempo de ejecución no pueda determinar la última fila del puerto.

Optimización de la memoria caché de datos

La memoria caché de datos de la transformación de agregación se ha optimizado para que utilice la longitud variable para almacenar los tipos de datos binarios y de cadena que pasen por la transformación de agregación. La optimización está habilitada en tamaños de registros de hasta 8 MB. Si el tamaño de los registros es mayor que 8 MB, la optimización de la longitud variable estará deshabilitada.

Si se usa la longitud variable para almacenar datos que pasan por la transformación de agregación en la memoria caché de datos, la transformación de agregación se optimiza para usar la entrada ordenada, y se inserta una transformación de ordenación de paso antes de la transformación de agregación en la asignación en tiempo de ejecución.

Para ver la transformación de ordenación, vea la asignación optimizada o vea el plan de ejecución en el entorno de validación de Blaze.

Durante la optimización de la memoria caché de datos, la memoria caché de datos y la memoria caché de índice de la transformación de agregación están establecidas en Automático. La memoria caché de ordenación de la transformación de ordenación está establecida en el mismo tamaño que la memoria caché de datos de la transformación de agregación. Para configurar la memoria caché de ordenación, debe configurar el tamaño de la memoria caché de datos de la transformación de agregación.

Transformación de agregación en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Funciones de agregado

Si usa un puerto en una expresión en la transformación de agregación, pero no usa ese puerto en una función de agregado, el motor en tiempo de ejecución podría usar cualquier fila del puerto para procesar la expresión.

Puede que la fila que utiliza el motor en tiempo de ejecución no sea la última fila del puerto. El procesamiento es distribuido y, por tanto, es posible que el motor en tiempo de ejecución no pueda determinar la última fila del puerto.

Optimización de la memoria caché de datos

La memoria caché de datos no se puede optimizar para que la transformación almacene datos usando la longitud variable.

Transformación de agregación en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- Un canal de transmisión contiene más de una transformación de agregación.
- Un canal de transmisión contiene una transformación de agregación y una transformación de rango.
- Una transformación de agregación está en un nivel superior a una transformación de búsqueda.
- Una transformación de agregación está en el mismo canal de transmisión que una transformación de búsqueda pasiva configurada con una condición de búsqueda de desigualdad.

Transformación de agregación en el motor de Databricks Spark

Algunas reglas de procesamiento del motor de Databricks Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Funciones de agregado

Si usa un puerto en una expresión en la transformación de agregación, pero no usa ese puerto en una función de agregado, el motor en tiempo de ejecución podría usar cualquier fila del puerto para procesar la expresión.

Puede que la fila que utiliza el motor en tiempo de ejecución no sea la última fila del puerto. El procesamiento es distribuido y, por tanto, es posible que el motor en tiempo de ejecución no pueda determinar la última fila del puerto.

Optimización de la memoria caché de datos

La memoria caché de datos no se puede optimizar para que la transformación almacene datos usando la longitud variable.

CAPÍTULO 6

Transformación de asociación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de asociación, 153](#)
- [Asignación de memoria, 154](#)
- [Propiedades avanzadas de transformación de asociación, 155](#)

Resumen de la transformación de asociación

La transformación de asociación procesa datos de salida de una transformación de coincidencia. Crea vínculos entre registros duplicados que se asignan a distintos clústeres de coincidencia, de modo que estos registros pueden asociarse juntos en operaciones de administración de datos principales y de consolidación de datos.

La transformación de asociación genera un valor **AssociationID** para cada fila en un grupo de registros asociados y escribe los valores de ID en un puerto de salida.

La transformación de consolidación lee la salida de la transformación de asociación. Utilice una transformación de consolidación para crear un registro principal basado en registros con valores de ID de asociación comunes.

La transformación de asociación acepta los valores de cadena y numéricos en los puertos de entrada. Si añade un puerto de entrada, la transformación convierte los valores de datos del puerto en cadenas.

El puerto de salida AssociationID escribe datos de enteros. La transformación puede escribir datos de cadenas en un puerto AssociationID si se ha configurado la transformación en una versión anterior de Informatica Data Quality.

Ejemplo: Cómo asociar salidas de transformación de coincidencia

La siguiente tabla contiene tres registros que pueden identificar a un mismo individuo:

ID	Nombre	Dirección	Ciudad	Estado	Código postal	SSN
1	David Jones	100 Admiral Ave.	New York	NY	10547	987-65-4321
2	Dennis Jones	1000 Alberta Ave.	New Jersey	NY	-	987-65-4321
3	D. Jones	Admiral Ave.	New York	NY	10547-1521	-

Una operación de análisis de duplicados definida en una transformación de coincidencia no identifica los tres registros como duplicados por las siguientes razones:

- Si define una búsqueda de duplicados en datos de nombres y de direcciones, los registros 1 y 3 se identifican como duplicados, pero se omite el registro 2.
- Si define una búsqueda de duplicados en datos de nombres y números de seguridad social, los registros 1 y 2 se identifican como duplicados, pero se omite el registro 3.
- Si se define una búsqueda de duplicados en los tres atributos (nombre, dirección y número de seguridad social), puede que la transformación de coincidencia no identifique ninguno de los registros como una coincidencia.

La transformación de asociación enlaza datos de distintos clústeres de coincidencia, de modo que los registros que comparten un ID de clúster reciben un valor AssociationID común. En este ejemplo, los tres registros reciben el mismo AssociationID, tal como se muestra en la siguiente tabla:

ID	Nombre	Dirección	Ciudad	Estado	Código postal	SSN	ID de clúster de nombre y dirección	ID de clúster de nombre y SSN	ID de asociación
1	David Jones	100 Admiral Ave.	New York	NY	10547	987-65-4320	1	1	1
2	Dennis Jones	1000 Alberta Ave.	New Jersey	NY	-	987-65-4320	2	1	1
3	D. Jones	Alberta Ave.	New York	NY	10547-1521	-	1	2	1

Puede consolidar los datos de registros duplicados en la transformación de consolidación.

Asignación de memoria

Puede definir la cantidad mínima de memoria caché que utiliza la transformación de asociación. El ajuste predeterminado es 400.000 bytes.

Defina el valor en la propiedad **Tamaño de la memoria caché** de la ficha **Avanzadas**.

El valor predeterminado representa la cantidad mínima de memoria que utiliza la transformación. La transformación de asociación intenta obtener un múltiplo del valor predeterminado en función del número de puertos que se asocian. La transformación utiliza esta fórmula para obtener la memoria caché:

$(\text{Número de puertos de asociación} + 1) \times \text{memoria caché predeterminada}$

Por ejemplo, si se configuran cuatro puertos de asociación, la transformación intenta asignar 3,2 millones de bytes, o 3,05 MB, a la memoria caché.

Si se cambia el ajuste predeterminado, la transformación no intenta obtener memoria adicional.

Nota: Si se especifica un valor de memoria caché menor que 65536, la transformación de asociación lee el valor en megabytes.

Propiedades avanzadas de transformación de asociación

La transformación de asociación contiene propiedades avanzadas que determinan el comportamiento de la memoria caché y el nivel de seguimiento.

Puede configurar las siguientes propiedades avanzadas:

Directorio de archivo de memoria caché

Especifica el directorio en el que el servicio de integración de datos escribe los datos temporales de la transformación actual. El servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos de entrada es mayor que la memoria disponible del sistema. El servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del servicio de integración de datos. El servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de archivo de memoria caché

Determina la cantidad de memoria del sistema que utiliza el servicio de integración de datos para ordenar los datos de entrada en la transformación.

Antes de ordenar los datos, el servicio de integración de datos asigna la cantidad de memoria que se haya especificado. Si la operación de ordenación genera una cantidad de datos mayor, el servicio de integración de datos escribirá los datos en exceso en el directorio de archivos de memoria caché. Si la operación de clasificación necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

La transformación lee el valor en bytes. El valor predeterminado es 400 000 bytes. El valor máximo es 2 147 483 647 bytes. Puede utilizar un parámetro para especificar el tamaño del archivo de memoria caché.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

CAPÍTULO 7

Transformación de excepción de registros incorrectos

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de excepción de registros incorrectos, 156](#)
- [Tipos de registros de salida de excepción de registros incorrectos, 157](#)
- [Flujo de proceso de administración de excepciones de registros incorrectos, 158](#)
- [Asignaciones de excepción de registros incorrectos, 159](#)
- [Puertos de excepción de registros incorrectos , 161](#)
- [Vista de configuración de excepción de registros incorrectos, 162](#)
- [Asignación problemática de excepción de registros incorrectos , 164](#)
- [Propiedades avanzadas de Transformación de excepción, 165](#)
- [Cómo configurar una transformación de excepción de registros incorrectos, 166](#)
- [Ejemplo de asignación de excepción de registros incorrectos, 167](#)

Resumen de la transformación de excepción de registros incorrectos

La transformación de excepción de registros incorrectos es una transformación activa que lee la salida de los procesos de calidad de datos e identifica registros que requieren la revisión manual. La transformación de excepción de registros incorrectos es una transformación de varios grupos.

Configure una transformación de excepción de registros incorrectos para analizar la salida de un proceso que identifica los problemas de calidad de datos en los registros. Una excepción es un registro que tiene un problema de calidad de datos que necesita más revisión.

La transformación de excepción de registros incorrectos recibe la entrada de otra transformación o de un objeto de datos en otra asignación. La entrada de la transformación de registros incorrectos debe contener uno o varios puertos de problemas de calidad que reciban las descripciones en texto de los problemas de calidad de datos. La entrada de la transformación de excepción de registros incorrectos también puede contener una puntuación de registro numérica que la transformación puede utilizar para determinar la calidad de los datos de cada registro. Establezca un umbral de puntuación superior e inferior en la transformación de excepción para clasificar los registros como calidad correcta e incorrecta según la puntuación del registro. La transformación de excepción de registros incorrectos escribe las excepciones y la descripción de los problemas de calidad asociados en una tabla de registros incorrectos.

Por ejemplo, una organización necesita validar las direcciones del cliente antes de enviarle el correo. Un desarrollador crea una asignación que valida la ciudad, el estado y el código postal del cliente frente a las tablas de referencia con una transformación de etiquetador. La transformación de etiquetador valida los campos y añade una puntuación de registro para cada fila según los resultados. La transformación de etiquetador también añade el texto que describe los problemas de calidad para cada registro que tenga un error. La transformación de etiquetador añade el texto del problema de calidad como `ciudad no válida` o `código postal en blanco` para cada excepción. La transformación de excepción de registros incorrectos escribe los registros de los clientes que deberían revisarse de forma manual en la tabla de registros incorrectos. Los analistas de datos revisan y corrigen los registros incorrectos en la herramienta del analista.

Tipos de registros de salida de excepción de registros incorrectos

La excepción de registros incorrectos examina las puntuaciones de registros de entrada para determinar la calidad de los registros. Devuelve los registros a grupos de salida diferentes

La transformación de excepción identifica los siguientes tipos de registros en función de cada puntuación de registros:

Registros correctos

Registros con puntuaciones mayores o iguales al umbral superior. Los registros correctos son válidos y no necesitan revisarse. Por ejemplo, si configura el umbral superior como 90, cualquier registro con una puntuación de 90 o más no necesitará revisarse.

Registros incorrectos

Los registros con puntuaciones inferiores al umbral superior y puntuaciones mayores o iguales que el umbral inferior. Los registros incorrectos son las excepciones que hay que revisar en Analyst Tool. Por ejemplo, cuando el umbral inferior es 40, cualquier registro con una puntuación de 40 a 90 necesita revisión manual.

Registros rechazados

Los registros con puntuaciones inferiores al umbral inferior. Los registros rechazados no son válidos. De forma predeterminada, la transformación de excepción suelta registros rechazados del flujo de datos. Para este ejemplo, cualquier registro con una puntuación de 40 o inferior es un registro rechazado.

Nota: Si los campos de problemas de calidad tienen valores nulos, el registro no es una excepción. Cuando un problema de calidad contiene texto o contiene una cadena vacía, el registro es una excepción. Compruebe que un puerto de problema de calidad contenga valores nulos cuando un campo no tenga ningún error. Si los puertos de problema de calidad contienen espacios en blanco en lugar de valores nulos, la transformación de excepción marca cada registro como una excepción. Cuando un usuario necesita corregir los problemas en Analyst Tool, el usuario no puede filtrar las excepciones por problema de calidad de datos.

Cuando un registro tiene una puntuación inferior a cero o mayor que 100, la fila no es válida. El servicio de integración de datos registra un mensaje de error de que la fila no es válida y omite el procesamiento del registro.

Si no se conecta una puntuación de registros como entrada de la transformación de excepción, la transformación escribe todos los registros que contienen los problemas de calidad a la tabla de registros incorrectos.

Cuando se incluye la transformación de excepción de los registros incorrectos a una tarea de asignación, puede configurar una tarea humana en el mismo flujo de trabajo para incluir una revisión manual de las excepciones. La tarea humana se inicia cuando una tarea de asignación del flujo de trabajo finaliza. La tarea

humana requiere que los usuarios accedan a Analyst Tool para solucionar los problemas de calidad. Un usuario puede actualizar los datos y cambiar el estado de la calidad de cada registro en la tabla de registros incorrectos.

Flujo de proceso de administración de excepciones de registros incorrectos

La transformación de excepción recibe puntuaciones de registros de las transformaciones de calidad de datos y crea las tablas que contienen los registros con distintos niveles de calidad de datos. Debe configurar las transformaciones de calidad de los datos para encontrar problemas de calidad y especificar una puntuación de registros para cada fila.

Puede configurar las transformaciones de calidad de datos en una asignación individual, o puede crear asignaciones en distintas etapas de los procesos de calidad de datos.

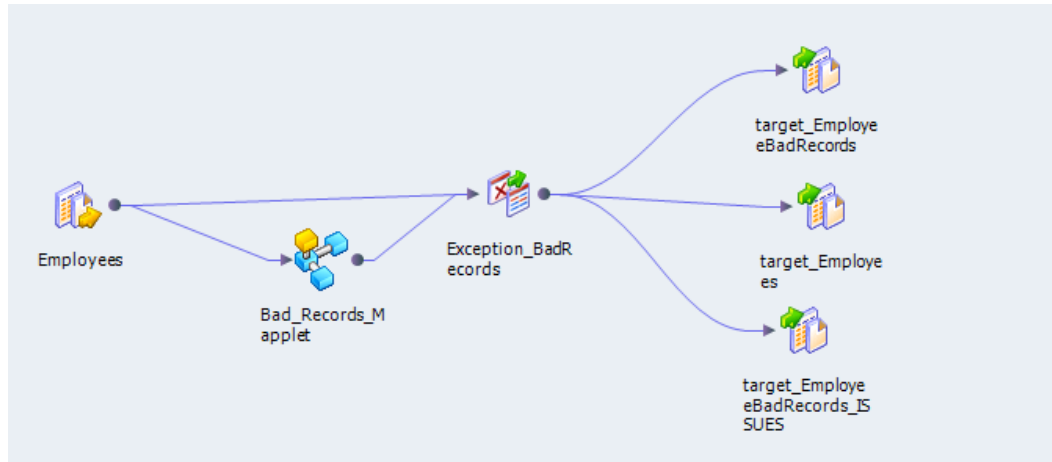
Realice las siguientes tareas de administración de excepciones de registros incorrectos:

1. En la herramienta Developer, defina las transformaciones que generan los valores de puntuación para datos de origen según los problemas de calidad de datos que se definan. Defina las transformaciones que devuelven el texto para describir la calidad de los datos de origen. Por ejemplo, puede configurar una transformación de etiquetador para comprobar los datos de origen en las tablas de referencia y luego escribir una etiqueta descriptiva para cada comparación. Puede definir una regla IF/THEN en una transformación de decisión para examinar un campo de datos. Puede definir varias transformaciones y mapplets que realicen diferentes operaciones de calidad de los datos.
2. Configure una transformación de excepción para analizar las puntuaciones del registro que recibe de las operaciones de calidad de datos. Configure la transformación para escribir los registros en las tablas de base de datos según los valores de puntuación de los registros. Puede crear tablas independientes para los registros correctos, registros incorrectos, problemas de calidad y registros rechazados.
3. Asigne un puerto de problema de calidad a cada puerto de entrada que pueda contener datos incorrectos.
4. Opcionalmente, configure los objetos de datos de destino para los registros correctos e incorrectos. Conecte los puertos de salida de la transformación de excepción a los objetos de datos de destino de la asignación.
5. Cree el objeto de datos de destino para los registros incorrectos. Elija generar una tabla de registros incorrectos y añádala a la asignación. Cuando se genere una tabla de registros incorrectos, Developer Tool también generará una tabla de problemas de calidad. Añada la tabla de problemas de calidad a la asignación.
6. Añada la asignación a un flujo de trabajo.
7. Configure una tarea humana para asignar la revisión manual de registros incorrectos a los usuarios. Los usuarios pueden revisar y actualizar los registros incorrectos en la herramienta del analista.

Asignaciones de excepción de registros incorrectos

Cuando cree una asignación que identifique excepciones de registros incorrectos, configure la asignación para escribir en uno o más destinos de la base de datos según la calidad de los datos en los registros.

La siguiente figura muestra un ejemplo de asignación de excepción de registros incorrectos:



La asignación contiene los siguientes objetos:

Origen de datos

Un origen de datos de empleados que contiene los registros para analizar la calidad de datos.

Mapplet

Bad_Records_Mapplet contiene transformaciones que buscan los problemas de calidad y las puntuaciones de registros, y los añaden a los registros de origen. Las reglas son transformaciones que analizan los datos y buscan los problemas de calidad. Por ejemplo, puede incluir una transformación de etiquetador para comparar datos de entrada con tablas de referencia. Según los resultados, puede configurar la transformación de etiquetador para devolver los problemas de calidad como columnas adicionales en las filas. Puede configurar una transformación de decisión que utiliza las instrucciones IF, THEN, ELSE para examinar los datos y aplicar problemas de calidad y puntuaciones de registros a los datos de entrada.

Transformación de excepción

La transformación de excepción determina qué registros se deben escribir en los destinos de datos, incluidas la tabla de registros incorrectos y la tabla de problemas.

Tabla de registros correctos

La transformación de excepción escribe todos los registros de buena calidad en la tabla target_Employees.

Tabla de registros incorrectos

La transformación de excepción escribe todos los registros de mala calidad en la tabla target_EmployeeBadRecords. Los registros incorrectos requieren revisión manual.

Tabla de problemas

La transformación de excepción escribe problemas de calidad en la tabla target_EmployeeBadRecords_ISSUES. Al visualizar los registros incorrectos en la herramienta del analista, la interfaz de usuario vincula los problemas de calidad a los registros incorrectos.

Opcionalmente, la transformación de excepción puede escribir los registros rechazados en una tabla de registros rechazados. Debe crear un grupo de salida independiente para los registros rechazados en la vista **Configuración** de la transformación.

Problemas de calidad de excepción de registros incorrectos

Los problemas de calidad son cadenas de texto que describen el tipo de problema de calidad de datos que ha generado una puntuación de registros baja. La transformación de excepción de registros incorrectos recibe problemas de calidad asociados con cada fila del origen que contiene una puntuación de registros baja. Puede configurar distintos tipos de transformaciones que determinen problemas de calidad y puntuaciones de registros.

Por ejemplo, puede crear una transformación de decisión que examine el número de teléfono. La transformación de decisión genera la puntuación de registros y los problemas de calidad para el número de teléfono.

La siguiente estrategia de decisión identifica los números de teléfono de longitud incorrecta en una transformación de decisión:

```
IF LENGTH(Phone_Number) > 10 THEN
  Score:=50
  Phone_Quality_Issue:='Phone num too long'
ELSEIF LENGTH(Phone_Number) < 10 THEN
  Score:=50
  Phone_Quality_Issue:=' Phone num too short'
ELSE
  Score:=90
ENDIF
```

Cuando configure la transformación de excepción, debe asociar Phone_Quality_Issue con el puerto Phone_Number. Los puertos son de diferentes grupos de entrada.

La transformación de excepción lee las puntuaciones generadas por la transformación de decisión y asigna registros con una puntuación de "50" al grupo de registros incorrectos de puertos de salida. Escribe Phone_Quality_Issue en el grupo de problemas de puertos de salida.

Tareas humanas

Cuando configura un flujo de trabajo que contiene una transformación de excepción, incluya la asignación en una tarea de asignación. Se añade una tarea humana al mismo flujo de trabajo. La tarea humana requiere que uno o más usuarios corrijan los registros de excepción de la herramienta del analista.

La tarea de asignación identifica los registros en los datos de origen que contienen problemas de calidad de datos sin resolver. Los analistas de datos utilizan la herramienta del analista para resolver estos problemas y para actualizar el estado de la calidad de datos de cada registro.

Cuando configure una tarea humana, puede crear una o más instancias de tarea y uno o más pasos de tarea. Una instancia de tarea representa el conjunto de datos en los que un usuario debe trabajar. Un paso de tarea representa el tipo de trabajo que un usuario realiza en los registros de su instancia de tarea. Puede crear varias instancias de tarea de modo que diferentes usuarios trabajen en distintas partes de los datos en la herramienta del analista.

Un usuario puede actualizar el estado de los registros incorrectos en la herramienta del analista de una de las siguientes maneras:

- Si un registro es válido, el usuario actualiza los metadatos de la tabla para confirmar el registro por el almacenamiento persistente en la base de datos.
- Si un registro no es válido, el usuario actualiza los metadatos de la tabla para eliminar dicho registro de la base de datos en una fase posterior del flujo de trabajo.

- Si el estado del registro no se confirma, el usuario actualiza los metadatos de la tabla de modo que el registro vuelve al flujo de trabajo para seguir procesándolo en una tarea de asignación.

Para obtener más información sobre las tareas humanas, consulte la *Guía de flujo de trabajo de Informática Developer*.

Puertos de excepción de registros incorrectos

Configure los puertos de entrada y salida en la ficha **Puertos** de la transformación de excepción de registros incorrectos.

La transformación de excepción de registros incorrectos contiene grupos de puertos de entrada y salida.

La siguiente figura muestra la ficha **Puertos**:

Puertos								
Nombre	Tipo	Precisión	Escala	Entrada	Salida	Predeterminado	Descripción	
[-] Entradas								
[-] Datos (3)								
1	EmployeeID	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
2	Employee_Name	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
3	Phone_Number	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
[-] Problemas de calidad (3)								
1	EmployeeID_Quality	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
2	Name_Quality_Issue	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
3	Phone_Quality_Issue	string	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
[-] Control (1)								
1	Puntuación	double	15	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
[-] Salidas								
[-] Salida estándar (4)								
1	Puntuación	double	15	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
2	EmployeeID	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
3	Employee_Name	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
4	Phone_Number	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
[-] Registros incorrectos (6)								
1	Workflow_ID	string	64	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
2	Row_Identifier	bigint	19	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
3	Record_Status	string	20	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
4	EmployeeID	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
5	Employee_Name	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
6	Phone_Number	string	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
[-] Problemas (2)								
1	Workflow_ID	string	64	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
2	Row_Identifier	bigint	19	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		

Puertos de entrada de transformación de excepción de registros incorrectos

Una transformación de excepción de registros incorrectos tiene distintos grupos de entrada para los datos, los problemas de calidad y la puntuación de registros.

La transformación de excepción de registros incorrectos incluye los siguientes grupos de entrada:

Datos

Los campos de datos de origen.

Problemas de calidad

Contiene puertos que describen los problemas de calidad de los registros. Los puertos de problemas de calidad contienen cadenas como "Excess_Characters" o "Bad_Data_Format". Puede tener varios problemas de calidad en cada registro. La transformación no asociará los puertos del grupo de

problemas de calidad a los campos de datos de origen hasta que se asignen los problemas a los puertos de datos en la vista **Asignación problemática**.

Control

La puntuación de registros. La transformación de excepción analiza la puntuación de registros para determinar si las filas de entrada son las excepciones. Si no se conecta el puerto de puntuación, la transformación de excepción identifica una fila como una excepción si un puerto de problemas de calidad contiene datos.

Salida de transformación de excepción de registros incorrectos

Una transformación de excepción de registros incorrectos tiene varios grupos de salida.

La transformación de excepción de registros incorrectos tiene los siguientes grupos de salida:

Salida estándar

Los registros de calidad satisfactoria que no es necesario examinar para detectar problemas de calidad de datos.

Cada registro del grupo de salida estándar contiene un puerto de puntuación que representa la calidad de datos del registro.

Registros incorrectos

Las excepciones que necesita examinar para detectar problemas de calidad de datos.

Cada registro del grupo de registros incorrectos contiene un ID de flujo de trabajo, un identificador de fila y un puerto de estado del registro.

Problemas

Los problemas de calidad para los registros del grupo de registros incorrectos. Los problemas de calidad son elementos de metadatos que la herramienta del analista muestra cuando se revisan los registros incorrectos.

Cada registro del grupo de problemas contiene un ID de flujo de trabajo y un puerto identificador de filas que identifica a qué fila de registros incorrectos pertenecen los problemas.

Registros rechazados

Grupo opcional que contiene los registros que pueden eliminarse de la base de datos. Cada registro del grupo de registros rechazados contiene una puntuación de registros baja en el puerto de puntuación.

Vista de configuración de excepción de registros incorrectos

La vista **Configuración** especifica los umbrales superior e inferior que utiliza la transformación para identificar los registros correctos y los registros incorrectos. La vista **Configuración** también identifica las tablas de destino de los registros con puntuaciones superiores o inferiores a los umbrales.

La siguiente figura muestra la vista **Configuración** de la transformación de excepción:

Umbral inferior :

Umbral superior :

Opciones de distribución de datos

Tipo	Salida estándar	Tabla de registros incorrectos
Registros correctos (sobre el umbral superior)	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Registros incorrectos (entre umbrales)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Registros rechazados (por debajo del umbral inferior)	<input type="checkbox"/>	<input type="checkbox"/>

Crear un grupo de salida aparte para los registros rechazados

Puede configurar las siguientes propiedades en la vista **Configuración**:

Umbral inferior

Límite inferior para el intervalo de puntuaciones de registros incorrectos. La transformación identifica los registros con puntuaciones inferiores al umbral inferior como registros rechazados.

Umbral superior

Límite superior para el intervalo de puntuaciones de registros incorrectos. La transformación identifica los registros con puntuaciones mayores o iguales al umbral superior como registros correctos.

Opciones de distribución de datos

Los tipos de registros de salida. En la configuración predeterminada, la transformación escribe los registros correctos en la salida estándar y los registros incorrectos en la tabla de registros incorrectos. De forma predeterminada, la transformación no escribe los registros rechazados en una tabla de base de datos.

Salida estándar

Los tipos de registro que la transformación escribe en los puertos de salida estándar. El valor predeterminado es Registros correctos.

Tabla de registros incorrectos

Los tipos de registro que la transformación escribe en los puertos de salida de registros incorrectos. El valor predeterminado es Registros incorrectos.

Crear un grupo de salida aparte para los registros rechazados

Crea un grupo de salida independiente para los registros rechazados. La opción está desactivada de forma predeterminada.

Generar tabla de registros incorrectos

Crea una tabla de base de datos que contiene datos de registros incorrectos. Cuando se selecciona esta opción, la transformación de excepción crea la tabla de la base de datos, añade un objeto de datos al repositorio de modelos y añade una instancia del objeto al lienzo de la asignación. Puede generar la tabla de registros incorrectos para una instancia de transformación de excepción en una asignación.

Cuando se genera la tabla de registros incorrectos, Developer tool también crea una tabla de problemas para almacenar metadatos descriptivos sobre los registros.

Nota: Developer tool añade un sufijo de 12 caracteres a cada nombre de columna de las tablas de registros incorrectos. Si utiliza una base de datos de Oracle, el nombre de columna de origen no puede contener más de 18 caracteres.

Cómo generar la tabla de registros incorrectos y la tabla de problemas

Cuando se agrega la transformación a una asignación, se pueden generar la tabla de registros incorrectos y la tabla de problemas. La herramienta del desarrollador añade las tablas al repositorio de modelos.

1. Haga clic en **Generar tabla de registros incorrectos** para generar la tabla.
A continuación, aparece el cuadro de diálogo **Crear objeto de datos relacionales**.
2. Examine las conexiones con la base de datos. Seleccione una conexión a la base de datos donde guardar la tabla.
3. Especifique un nombre para la tabla de registros incorrectos. La herramienta del desarrollador aplica el nombre que se ha especificado a la tabla de registros incorrectos y a la tabla de problemas.
La herramienta del desarrollador añade la siguiente cadena al nombre de la tabla de problemas:
_ISSUE
Si se conecta a una base de datos de Oracle, el nombre de tabla de registros incorrectos no debe tener más de 24 caracteres.
4. Especifique un nombre para el objeto de datos de registros incorrectos del repositorio de modelos.
5. Haga clic en **Finalizar**.
La herramienta del desarrollador añade las tablas al lienzo de la asignación y al repositorio de modelos.

Asignación problemática de excepción de registros incorrectos

Debe asignar puertos y prioridades a problemas de calidad de datos.

La siguiente figura muestra la vista **Asignación problemática**:

Asignación de problemas		
Asignar puertos y prioridades a problemas de calidad		
Problema de calidad	Entrada	
EmployeeID_Quality_Issue	EmployeeID	1
Name_Quality_Issue	EmployeeName	1
Phone_Quality_Issue		1

La vista **Asignación problemática** contiene los siguientes campos:

Problema de calidad

Cada puerto de problema de calidad que se define en el grupo de entrada de problemas de calidad aparece en la columna **Problema de calidad**.

Entrada

La columna **Entrada** contiene los puertos de datos que se asignan a los problemas de calidad en la vista **Asignación problemática**. Asocie un puerto de entrada con cada puerto de problema de calidad. Cada puerto de entrada que contiene los datos de calidad incorrectos debe tener al menos un puerto de problema de calidad que indica un tipo de problema. Puede seleccionar el puerto Phone_Number para Phone_Quality_Issue, por ejemplo. Puede asignar un puerto a más de un problema de calidad.

Prioridad de problema

Las prioridades de problema determinan qué problemas tienen la mayor importancia a la hora de asignar el mismo puerto de entrada a varios problemas de calidad. Si se produce más de un problema para un campo de entrada, el servicio de integración de datos aplica el problema de calidad que tiene la prioridad más alta. En el caso de que exista más de un problema de calidad para un puerto de entrada y los problemas tengan todos la misma prioridad, el servicio de integración de datos aplica el problema de calidad que aparece primero en la lista. Especifique una prioridad entre 1 y 99, donde 1 representa la prioridad más alta.

Defina las prioridades de problema para filtrar los registros en la herramienta del analista.

Asignar puertos a problemas de calidad

Asigne un puerto para asociarlo con cada problema de calidad. Developer Tool crea puertos en el grupo de salida de problemas para cada asociación que se añada a la vista **Asignación problemática**.

1. Para cada problema de calidad, haga clic en el campo **Entrada** para mostrar una lista de los puertos de entrada.
2. Seleccione un puerto de entrada para asociarlo con el problema de calidad.
Puede elegir el mismo puerto para más de un problema.
3. Haga clic en la columna **Problema** y seleccione una prioridad para el problema de calidad.

Propiedades avanzadas de Transformación de excepción

Configure las propiedades que determinan cómo el servicio de integración de datos procesa los datos para la transformación de excepción.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Cómo configurar una transformación de excepción de registros incorrectos

Cuando configure una transformación de excepción de registros incorrectos, configure los puertos de entrada y los problemas de calidad que pueden aparecer en cada puerto. Defina los umbrales superior e inferior para determinar la calidad de los datos. Configure dónde escribir las excepciones y los registros rechazados.

1. Cree una transformación de excepción de registros incorrectos reutilizable o no reutilizable.
 - Para crear una transformación reutilizable, haga clic en **Archivo > Nuevo > Transformación** y seleccione una transformación de excepción de registros incorrectos.
 - Para crear una transformación no reutilizable, abra una asignación y añada una transformación de excepción en el lienzo de la asignación. Seleccione una transformación de excepción de registros incorrectos en el asistente.
2. Haga clic en **Siguiente** o en **Finalizar**.

Si hace clic en **Siguiente**, puede actualizar los valores de umbral predeterminados y las opciones de distribución de datos antes de crear la transformación.
3. Configure los puertos de entrada.
 - Si crea una transformación reutilizable, seleccione la ficha **Puertos** y añada puertos para los datos que desea conectar a la transformación.
 - Si crea una transformación no reutilizable, añada otros objetos al lienzo de la asignación y arrastre los puertos de entrada para la transformación.
4. Seleccione la vista **Configuración**.
5. Configure los umbrales de puntuación superior e inferior.
6. En la sección **Opciones de distribución de datos**, configure las propiedades estándar de la salida y de la tabla de excepciones para establecer dónde la transformación escribirá cada tipo de registro.

Configure dónde escribir los registros correctos, registros incorrectos y registros rechazados. Puede escribirlos en la salida estándar o en la tabla de registros incorrectos.
7. Abra la vista **Asignación problemática**. Asigne los problemas de calidad de datos a los puertos de datos.

Asignar una prioridad a cada problema. Si un puerto contiene valores con varios problemas, la transformación muestra el problema con mayor prioridad.
8. Seleccione la opción para generar una tabla de registros incorrectos. Introduzca la información de conexión de la base de datos y el nombre de la tabla. La tabla debe proceder del esquema predeterminado.
 - Si genera una tabla de registros incorrectos, genere una tabla para los registros y una tabla adicional para los problemas de calidad de datos relacionados con los registros. La transformación crea un objeto de base de datos en el repositorio de modelos.
9. Conecte los puertos de salida de la transformación a uno o varios destinos de datos. Conecte los puertos de salida a los objetos de datos que correspondan a las opciones de salida que ha establecido en la vista **Configuración**.
 - Si crea una transformación reutilizable, añada la transformación a una asignación y conecte los puertos de salida.
 - Si crea una transformación no reutilizable, la transformación conecta los puertos a la tabla de registros incorrectos. Conecte los puertos de salida a cualquier otro destino de datos.

Ejemplo de asignación de excepción de registros incorrectos

Una organización dirige un proyecto de datos para revisar nuevos datos de clientes. La organización necesita comprobar que los datos de contacto son válidos. El siguiente ejemplo muestra cómo definir una transformación de excepción de registros incorrectos que recibe los registros de un mapplet que realiza un análisis de calidad de datos de registros de los clientes.

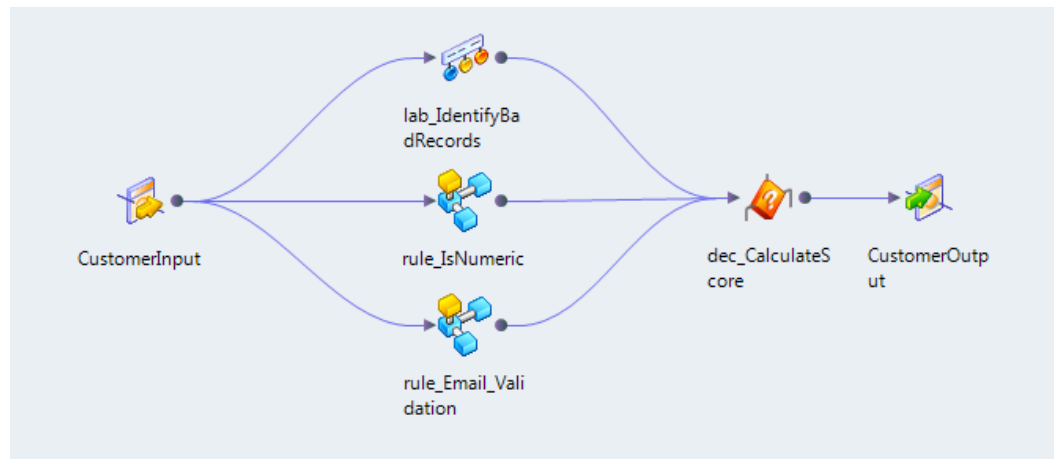
Cree un mapplet con transformaciones de calidad de datos que evalúen el formato y la exactitud de los datos de los clientes. El mapplet incluye transformaciones que generan una puntuación de registros según los resultados del análisis de calidad de datos. Las transformaciones también definen los problemas de calidad de los datos según los resultados del análisis.

Mapplet de excepción de registros incorrectos

Cree un mapplet que contenga las transformaciones de calidad de datos para comprobar los valores de determinados campos. Las transformaciones comprueban las tablas de referencia y los conjuntos de contenido para determinar si los campos de los registros son válidos. Las transformaciones aplican una puntuación de registros para cada registro según los resultados. La transformación de excepción recibe los registros del mapplet y distribuye cada registro a la salida correspondiente según la puntuación de registros.

El mapplet se compone de transformaciones de etiquetador, transformaciones de decisión y transformaciones de expresión.

La siguiente figura muestra los objetos en el mapplet:



El mapplet realiza las siguientes tareas:

- Una transformación de etiquetador comprueba la localidad, el estado, el código de país, el código postal, y el código postal de los datos que recibe en los puertos de entrada. La transformación contiene una estrategia para cada puerto. Las estrategias comparan los datos de origen con las tablas de referencia e identifican los valores que no son válidos.
- Un mapplet de transformación de expresión comprueba que el número de teléfono es numérico y que contiene 10 dígitos.
- Una transformación de etiquetador y un mapplet de transformación de expresión comprueban que la dirección de correo electrónico es válida. La transformación de expresión comprueba la estructura de la

cadena de correo electrónico. La transformación de etiquetador comprueba la dirección IP en una tabla de referencia de sufijos de dirección IP internacionales.

- Una transformación de decisión recibe la salida de la transformación y de los mapplets. Se calcula una puntuación de registros general para el registro de contactos del cliente.

Cree una asignación de excepción de registros incorrectos que incluya el mapplet. La asignación de excepción de registros incorrectos incluye una transformación de excepción que escribe las excepciones en una tabla de base de datos de registros incorrectos. Un analista de datos investiga y actualiza los registros de excepciones en la tabla de registros incorrectos con la herramienta del analista.

Grupos de entrada de ejemplo de excepción de registros incorrectos

La transformación de excepción tiene tres grupos de entrada. La transformación tiene un grupo de datos que recibe los datos de origen. Tiene el grupo de problemas de calidad, que recibe los problemas de calidad de los datos que las transformaciones de calidad de los datos encuentran. También tiene un grupo de control que contiene la puntuación de registros para la fila.

La siguiente figura muestra los grupos de entrada de la transformación de excepción:

	Nombre	Tipo	Precisión	Escala	Entrada	Salida
[-] Entradas						
[-] Datos (11)						
1	CUST_ID	decimal	20	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	COMPANY	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	CONTACT	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4	TITLE	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	ADDR1	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	ADDR2	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
7	ADDR3	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
8	ADDR4	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
9	COUNTRY	string	20	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
10	PHONE	string	50	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11	EMAIL	string	100	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
[-] Problemas de calidad (6)						
1	CompanyStat...	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	LocalityStatus	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	ProvinceStatus	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4	CountryStatus	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	ZipStatus	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	EmailStatus	string	30	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
[-] Control (1)						
[-] Salidas						

Configuración de ejemplo de excepción de registros incorrectos

Defina los umbrales superior e inferior en la vista **Configuración**. Identifique dónde escribe la transformación los registros correctos, los registros incorrectos y los registros rechazados.

Acepte la configuración predeterminada para la distribución de los registros correctos, registros incorrectos y problemas.

La siguiente figura muestra la vista **Configuración** de la transformación de excepción:

Umbral inferior :

Umbral superior :

Opciones de distribución de datos

Tipo	Salida estándar	Tabla de registros incorrectos
Registros correctos (sobre el umbral superior)	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Registros incorrectos (entre umbrales)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Registros rechazados (por debajo del umbral inferior)	<input type="checkbox"/>	<input type="checkbox"/>

Crear un grupo de salida aparte para los registros rechazados

La siguiente tabla describe las opciones de configuración:

Opción	Valor
Umbral inferior	10
Umbral superior	90
Registros correctos	Salida estándar
Registros incorrectos	Tabla de registros incorrectos
Registros rechazados	-

Haga clic en **Generar tablas de registros incorrectos** para crear las tablas de registros incorrectos y de problemas.

Salida de asignación de ejemplo de excepción de registros incorrectos

Añada una transformación de escritura a la asignación y conecte los puertos de salida estándar al objeto de datos. La asignación también contiene el objeto de la base de datos de registros incorrectos y los problemas de objeto de base de datos que ha creado en la vista **Configuración**.

Tabla de registros incorrectos

La tabla de registros incorrectos contiene las excepciones de registros con puntuaciones entre los umbrales inferior y superior.

La siguiente figura muestra los registros incorrectos que la transformación de excepción devuelve:

Salida								
Nombre: Excepción_Bad_Output_DI								
Workflow_ID	Row_Identifier	Record_Status	CUST_ID	COMPANY	CONTACT	TITLE	ADDR1	ADDR2
1	DummyWorkflowID	0	1001590	E-AGENCY	OAKLAND	FederalExpress...	291 3RD STREET	2500 WESTERF...
2	DummyWorkflowID	1	1001604	KPMG PEAT...	WOODCLIFF...	US Postal2-day	530 CHESNUT...	"50 BROAD ST...
3	DummyWorkflowID	2	1001622	INVESTEX	NEW YORK	UnitedParcel S...	"50 BROAD ST...	20TH FL"
4	DummyWorkflowID	3	7121564	"ARTHURVA...	INC."	WHITEHOUSE...	FederalExpress...	310 BRIER ROAD
5	DummyWorkflowID	4	7121569	RICE VOELKER	MANDEVILLE	FederalExpress...	3840HIGHWAY...	"2ND & D STRE...
6	DummyWorkflowID	5	1001658	HOUSEINFO...	WASHINGTON	FederalExpress...	"2ND & D STRE...	SW HOUSEA...
7	DummyWorkflowID	6	1001694	BANK OF MO...	BOSTON	US Postal2-day	"125 BROAD ST...	38TH FLOOR"
8	DummyWorkflowID	7	1001738	"INTERFACE...	INC."	ANN ARBOR	UnitedParcel S...	5855 INTERFAC...
9	DummyWorkflowID	8	1001798	NORTHWEST...	MILWAUKEE	US Postal2-day	EAST MASON S...	EAST MASON S...
10	DummyWorkflowID	9	1001735	TOPSPIN CR...	NEW YORK	FederalExpress...	"276 FIFTH AVE.	#902"

La tabla de registros incorrectos incluye todos los campos del registro de origen. Un registro incorrecto también incluye los siguientes campos:

Workflow_ID

El nombre del flujo de trabajo que incluyó la transformación de excepción. El flujo de trabajo contiene la tarea de asignación de transformación de excepción y la tarea humana para revisar los problemas. El ID de flujo de trabajo contiene `DummyWorkflowID` si la transformación de excepción no está en un flujo de trabajo.

Row_Identifier

Un número único que identifica cada fila.

Record_Status

Un estado del registro para la herramienta del analista. Cada registro de la tabla de registros incorrectos recibe un estado `No válido`. Puede conservar el estado del registro al actualizar registros en la herramienta del analista.

Tabla de problemas

La tabla de problemas contiene una fila para cada fila de la tabla de registros incorrectos. Cada fila contiene los problemas que el análisis de calidad de datos ha encontrado para el registro de origen.

La siguiente figura muestra las columnas de la tabla de problemas:

Salida				
Nombre: exc_BadRecord_Issues_DI				
Workflow_ID	Row_Identifier	CUST_ID	DQAPRIORITY_CUST_ID	
1	DummyWorkflowID	0	6803 S. TUCSON WAY	1
2	DummyWorkflowID	1	6803 S. TUCSON WAY	1
3	DummyWorkflowID	2	20TH FL"	1
4	DummyWorkflowID	3	310 BRIER ROAD	1
5	DummyWorkflowID	4	767 FIFTH AVENUE	1
6	DummyWorkflowID	5	6803 S. TUCSON WAY	1
7	DummyWorkflowID	6	6803 S. TUCSON WAY	1
8	DummyWorkflowID	7	6803 S. TUCSON WAY	1
9	DummyWorkflowID	8	6803 S. TUCSON WAY	1
10	DummyWorkflowID	9	SW HOUSE ANNEX 2"	1
11	DummyWorkflowID	10	6803 S. TUCSON WAY	1

La tabla de problemas contiene las siguientes columnas:

Workflow_ID

Identifica el flujo de trabajo que ha creado el registro. El flujo de trabajo contiene la tarea de asignación y la tarea humana de la transformación de excepción para revisar el problema.

Row_Identifier

Identifica la fila del registro en la tabla de base de datos. El identificador de filas identifica qué fila en la tabla de registros incorrectos se corresponde con la fila en la tabla de problemas.

Nombre de campo del problema

El nombre de campo es el nombre del campo que podría tener problemas de calidad. Cuando el campo contiene un error el valor de la columna es el texto del problema de calidad. En la figura anterior, el nombre de campo ADDR2 contiene el problema de calidad `invalid_locality`.

DQAPriority

La prioridad del problema. Cuando se producen varios problemas para el mismo campo, el problema con la prioridad más alta aparece en el problema el nombre de campo del problema.

Tabla de registros correctos

Cada registro de la tabla de registros correctos tiene una puntuación de registro mayor que el umbral superior. En este ejemplo, el umbral superior es 90.

La siguiente figura muestra los registros correctos que la transformación de excepción devuelve:

Salida							
Nombre: exc_BadRecord_Output_DI							
	Score	CUST_ID	COMPANY	CONTACT	TITLE	ADDR1	ADDR2
1	<null>	<null>	<null>	1001658	HOUSEINFOR...	WASHINGTON	Federal Express 2-day
2	<null>	<null>	<null>	15951451	D E FREY & CO...	DENVER	Federal Express Overnight
3	<null>	<null>	<null>	15951453	SONANGOL CO...	HOUSTON	United Parcel Service Red
4	<null>	<null>	<null>	15951457	CITADEL INVES...	CHICAGO	United Parcel Service Intern
5	<null>	<null>	<null>	15951458	IOMEGA	ROY	US Postal International
6	<null>	<null>	<null>	15951461	VICTORY SBSF...	NEW YORK	US Postal International
7	<null>	<null>	<null>	15951465	WILLIAMSENE...	TULSA	United Parcel Service Blue
8	<null>	<null>	<null>	15951469	PAINE WEBBER	DALLAS	US Postal 2-day
9	<null>	<null>	<null>	15951475	PINKSHEETS.C...	NEW YORK	United Parcel Service Red
10	<null>	<null>	<null>	15951482	JAMES A LUSTIG	DENVER	United Parcel Service Blue

Los registros de la tabla de registros correctos contienen los campos de la puntuación de registro y los datos de origen.

CAPÍTULO 8

Transformación de conversión de mayúsculas y minúsculas

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de conversión de mayúsculas y minúsculas, 172](#)
- [Propiedades de la estrategia de distinción de mayúsculas y minúsculas, 173](#)
- [Cómo configurar una estrategia de conversión de mayúsculas y minúsculas, 173](#)
- [Propiedades avanzadas de Transformación de conversión de mayúsculas y minúsculas, 174](#)
- [Transformación de conversión de mayúsculas y minúsculas en un entorno no nativo, 174](#)

Resumen de la transformación de conversión de mayúsculas y minúsculas

La transformación de conversión de mayúsculas y minúsculas es una transformación pasiva que normaliza el uso de mayúsculas o minúsculas de los caracteres del alfabeto en las cadenas de entrada.

Puede seleccionar un formato de conversión del formato de mayúsculas y minúsculas, como mayúsculas, minúsculas, formato de título y formato de frase. También puede invertir el uso de mayúsculas o minúsculas actual de cada carácter en los datos de entrada.

La transformación de conversión de mayúsculas y minúsculas puede utilizar los valores de la columna válida de una tabla de referencia para definir el uso de mayúsculas o minúsculas de los caracteres de entrada. Cuando la transformación encuentra una coincidencia entre un valor de entrada y un valor válido, la transformación aplica el formato de mayúscula o minúscula del valor válido en el valor de entrada. Puede usar tablas de referencia cuando el tipo de conversión de formato de mayúsculas y minúsculas sea **Tipo título** o **Tipo oración**.

Puede crear múltiples estrategias de conversión de formatos de mayúscula o minúscula en una transformación de conversión de mayúsculas y minúsculas. Cada estrategia utiliza un tipo de conversión individual.

Propiedades de la estrategia de distinción de mayúsculas y minúsculas

Se pueden configurar propiedades para estrategias de conversión de mayúsculas y minúsculas.

En la vista **Estrategias** puede configurar las siguientes propiedades para la conversión de mayúsculas y minúsculas:

Tipo de conversión

Define el método de conversión de mayúsculas y minúsculas que utiliza una estrategia. Se pueden aplicar los siguientes tipos de conversión de mayúsculas y minúsculas.

- **Mayúsculas.** Convierte todas las letras a mayúsculas.
- **Tipo de frase.** Capitaliza la primera letra de la cadena de datos del campo.
- **Alternar mayúsculas y minúsculas.** Convierte las letras minúsculas en mayúsculas y viceversa.
- **Tipo de título.** Capitaliza la primera letra de cada subcadena.
- **Minúsculas.** Convierte todas las letras a minúsculas.

El método de conversión de mayúsculas y minúsculas predeterminado es mayúsculas.

No modificar palabras en mayúscula

Reemplaza la capitalización elegida para las cadenas en mayúscula.

Delimitadores

Define cómo funciona la capitalización en la conversión de mayúsculas y minúsculas. Por ejemplo, elija un guión como delimitador para transformar "smith-jones" en "Smith-Jones". El delimitador predeterminado es el carácter de espacio.

Tabla de referencia

Aplica el formato de capitalización especificado por una tabla de referencia. Se aplica únicamente si la opción de conversión de mayúsculas y minúsculas es **Formato de mayúsculas y minúsculas de título** o **Formato de mayúsculas y minúsculas de frase**. Haga clic en **Nuevo** para añadir una tabla de referencia a la estrategia.

Nota: Si se produce una coincidencia de tabla de referencia al principio de un token, el siguiente carácter de ese token se cambia a mayúscula. Por ejemplo, si la cadena de entrada es mcdonald y la tabla de referencia tiene una entrada para Mc, la cadena de salida será McDonald.

Cómo configurar una estrategia de conversión de mayúsculas y minúsculas

Para cambiar el formato de mayúsculas y minúsculas de las cadenas de entrada, configure los ajustes en la vista **Estrategias** de una transformación de conversión de mayúsculas y minúsculas.

1. Seleccione la vista **Estrategias**.
2. Haga clic en **Nueva**.
Se abre el asistente **Nueva estrategia**.
3. Si lo desea, edite el nombre y la descripción de la estrategia.

4. Haga clic en los campos **Entradas** y **Salidas** para seleccionar los puertos de la estrategia.
5. Configure las propiedades de la estrategia. La estrategia de conversión predeterminada es **Mayúsculas**.
6. Haga clic en **Siguiente**.
7. Si lo desea, añada tablas de referencia para personalizar las opciones de mayúsculas y minúsculas de los datos de entrada que coinciden con las entradas de la tabla de referencia. La personalización del formato de mayúsculas y minúsculas de la tabla de referencia solamente se aplica a las estrategias de formato de mayúsculas y minúsculas de títulos y de frases.
8. Haga clic en **Finalizar**.

Propiedades avanzadas de Transformación de conversión de mayúsculas y minúsculas

Configure las propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de conversión de mayúsculas y minúsculas.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de conversión de mayúsculas y minúsculas en un entorno no nativo

El procesamiento de transformaciones de conversión de mayúsculas y minúsculas en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 9

Transformación de clasificador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de clasificador, 175](#)
- [Modelos de clasificador, 176](#)
- [Algoritmos clasificadores, 176](#)
- [Opciones de la transformación de clasificador, 176](#)
- [Estrategias de clasificador, 177](#)
- [Propiedades avanzadas de Transformación de clasificador, 177](#)
- [Cómo configurar una estrategia de clasificador, 178](#)
- [Ejemplo de análisis de clasificador, 178](#)
- [Transformación de clasificador en un entorno no nativo, 183](#)

Resumen de la transformación de clasificador

La transformación de clasificador es una transformación pasiva que analiza los campos de entrada e identifica el tipo de información de cada campo. Utilice una transformación de clasificador cuando los campos de entrada contengan varios valores de texto.

Cuando configure la transformación de clasificador, seleccione un modelo de clasificador y un algoritmo de clasificador. Un modelo de clasificador es un tipo de objeto de datos de referencia. Un algoritmo de clasificador es un conjunto de reglas que calcula el número de palabras similares en una cadena y la posición relativa de las palabras. La transformación compara el análisis del algoritmo con el contenido del modelo de clasificador. La transformación devuelve la clasificación del modelo que identifica el tipo de información dominante en la cadena.

La transformación de clasificador puede analizar cadenas de longitud considerable. Por ejemplo, puede utilizar la transformación para clasificar el contenido de mensajes de correo electrónico, mensajes de medios sociales y texto de documentos. Puede pasar el contenido de cada documento o mensaje a un campo en una columna del origen de datos, y conectar la columna a una transformación de clasificador. En cada caso, prepare el origen de datos de modo que cada campo incluya el contenido completo de un documento o de cadena que desee analizar.

Modelos de clasificador

La transformación de clasificador utiliza un objeto de datos de referencia llamado modelo de clasificador para analizar datos de entrada. Debe seleccionar el modelo de clasificador cuando configure una transformación de clasificador. La transformación compara una columna de datos de entrada con los datos del modelo de clasificador y devuelve una etiqueta que describe el tipo de información en cada campo de entrada.

Un modelo de clasificador contiene filas de datos de referencia y valores de etiqueta. Las filas representan los datos de entrada en el puerto que puede conectar a la transformación de clasificador. Los valores de etiqueta describen los tipos de información que contienen las filas de datos. Al configurar un modelo de clasificador, puede asignar una etiqueta a cada fila de datos de referencia del modelo.

Para vincular las filas de datos de referencia a las etiquetas de un modelo de clasificador, debe compilar el modelo. El proceso de compilación genera una serie de asociaciones lógicas entre las filas de datos y los valores de etiqueta. Cuando se ejecuta una asignación que lee el modelo, el Servicio de integración de datos aplica la lógica del modelo a los datos de entrada de la transformación de clasificador. El Servicio de integración de datos devuelve las etiquetas que describen mejor la información en cada campo de datos de entrada.

Puede crear un modelo de clasificador en Developer tool. El repositorio de modelos almacena el objeto del modelo de clasificador. Developer tool escribe las filas de datos, las etiquetas y los datos de compilación en un archivo de la estructura del directorio de Informatica.

Algoritmos clasificadores

Cuando se añade un modelo clasificador a una estrategia de transformación, también debe seleccionar un algoritmo clasificador. El algoritmo determina cómo la transformación compara los datos del modelo de clasificador con los datos de entrada.

Puede seleccionar el algoritmo **Bayesiano Ingenuo** o el algoritmo **Entropía máxima**.

Tenga en cuenta los siguientes factores cuando seleccione un algoritmo:

- El algoritmo Entropía máxima realiza un análisis más exhaustivo que el algoritmo Bayesiano Ingenuo.
- Una asignación que utiliza el algoritmo Bayesiano Ingenuo se ejecuta más rápidamente que una asignación que utiliza el algoritmo Entropía máxima en los mismos datos.
- Seleccione el algoritmo Entropía máxima con el modelo de clasificador que Informatica incluye en el acelerador del núcleo.

Opciones de la transformación de clasificador

La transformación de clasificador muestra opciones configurables en una serie de fichas o vistas en Developer Tool.

Cuando abre una transformación reutilizable, las opciones aparecen en una serie de fichas en el editor de transformaciones. Cuando abre una transformación no reutilizable en una asignación, las opciones aparecen en una serie de vistas en el editor de asignaciones. Seleccione la ficha de propiedades de asignación para ver las vistas en una transformación no reutilizable.

Puede seleccionar las siguientes vistas:

General

Ver y actualizar el nombre y la descripción de la transformación.

Puertos

Ver los puertos de entrada y salida en la transformación.

Nota: En una transformación de clasificador reutilizable, las vistas General y Puertos se combinan en la ficha **Resumen**.

Estrategias

Añadir, quitar o editar una estrategia.

Dependencias

Ver los puertos de entrada y salida en cada estrategia.

Avanzadas

Establecer el nivel de detalle que la transformación escribe en los archivos de registro.

Estrategias de clasificador

Una estrategia es un conjunto de operaciones de análisis de datos que una transformación realiza en los datos de entrada. Debe crear como mínimo una estrategia en la transformación de clasificador. Una estrategia de clasificador lee un solo puerto de entrada.

Puede definir una o más operaciones en una estrategia. Una operación de clasificador identifica un modelo de clasificador y un algoritmo de clasificador para aplicar a los datos de los puertos de entrada. Cada operación escribe para un puerto de salida diferente. Puede crear varias operaciones en una estrategia cuando desee analizar un puerto de entrada de diferentes maneras.

Nota: Cuando desee identificar los idiomas utilizados en los datos de origen, seleccione el algoritmo Entropía máxima en la operación de clasificador.

Propiedades avanzadas de Transformación de clasificador

Configure propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de clasificador.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Cómo configurar una estrategia de clasificador

Puede configurar una estrategia para identificar los tipos de información en los datos. Cada estrategia analiza un puerto de entrada.

En una transformación no reutilizable, conecte los puertos de entrada a la transformación antes de configurar la estrategia.

1. Abra la transformación y seleccione la vista **Estrategias**.
2. Haga clic en **Nueva estrategia**.
Se abrirá el asistente para la creación de estrategias.
3. Escriba un nombre y una descripción opcional para la estrategia.
4. Seleccione un puerto de entrada del campo Entradas.
5. Compruebe que el valor de precisión del puerto de entrada es lo suficientemente alto para leer todos los campos del puerto de entrada. El puerto trunca los datos de entrada cuando alcanzan el límite de precisión.
6. Seleccione o borre la opción para añadir datos de puntuación a la salida de estrategia.
7. Haga clic en **Siguiente**.
8. Confirme el tipo de operación de clasificador y haga clic en **Siguiente**.
9. Seleccione un algoritmo de clasificador. Puede seleccionar los siguientes algoritmos:
 - Bayesiano Ingenuo
 - Entropía máxima**Nota:** Para identificar el idioma utilizado en los datos de origen, seleccione el algoritmo de entropía máxima.
10. Compruebe el puerto de salida.
La transformación crea un solo puerto de salida para cada operación en una estrategia. Puede editar el nombre de puerto y la precisión.
11. Seleccione un modelo de clasificador.
El asistente enumera los objetos de modelo de clasificador en el repositorio de modelos.
12. Haga clic en **Siguiente** para añadir otra operación a la estrategia. De lo contrario, haga clic en **Finalizar**.

Ejemplo de análisis de clasificador

Trabaja como gestor de datos en una empresa de software que ha lanzado una nueva aplicación para smartphones. La empresa quiere entender la respuesta del público a la aplicación y la cobertura que recibe en los medios de comunicación. La compañía le pide a usted y a su equipo que analicen los comentarios de las redes sociales sobre la aplicación.

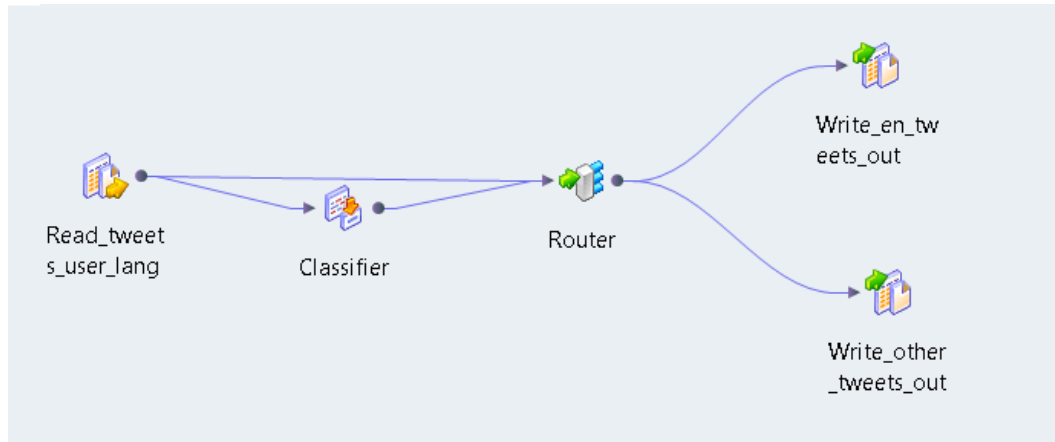
Decide capturar datos de fuentes de Twitter que analizan smartphones. Utiliza la interfaz de programación de aplicaciones de Twitter para filtrar el flujo de datos de Twitter. Crea un origen de datos que contiene los datos de Twitter que desea analizar.

Dado que las fuentes de Twitter contienen mensajes en varios idiomas, debe identificar el idioma utilizado en cada mensaje. Decide usar una transformación de clasificador para analizar los idiomas. Crea una asignación que identifica los idiomas en los datos de origen y escribe los mensajes de Twitter en destinos de datos en inglés e idiomas distintos al inglés.

Crear la asignación de clasificador

Puede crear una asignación que lea un origen de datos, clasifique los idiomas en los datos y escriba los datos en destinos según los idiomas que contengan.

La siguiente imagen muestra la asignación en Developer Tool:



La asignación que cree contiene los siguientes objetos:

Nombre de objeto	Descripción
Read_tweet_user_lang	Origen de datos. Contiene los mensajes de Twitter
Clasificador	Transformación de clasificador. Identifica los idiomas utilizados en los mensajes de Twitter.
Enrutador	Transformación de enrutador. Distribuye los mensajes de Twitter a los objetos de destino de datos según los idiomas contienen.
Write_en_tweets_out	Destino de datos. Contiene mensajes de Twitter en inglés.
Write_other_tweets_out	Destino de datos. Contiene mensajes de Twitter en idiomas distintos al inglés.

Muestra de datos de entrada

El siguiente fragmento de datos muestra un ejemplo de los datos de Twitter que analiza en la asignación:

Twitter Message

```
RT @GanaphoneS3: Faltan 10 minutos para la gran rifa de un iPhone 5...
RT @Clarified: How to Downgrade Your iPhone 4 From iOS 6.x to iOS 5.x (Mac)...
RT @jerseyjazz: The razor was the iPhone of the early 2000s
RT @KrissiDevine: Apple Pie that I made for Thanksgiving. http://t.com/s9ImzFxO
RT @sophieHz: Dan yang punya 2 kupon undian. Masuk dalam kotak undian yang berhadiah Samsung
RT @IsabelFreitas: o galaxy tem isso isso isso e a bateria ã melhor que do iPhone
RT @PremiusIpad: Faltan 15 minutos para la gran rifa de un iPhone 5...
RT @payyton3: I want apple cider
RT @wiesteronder: Retweet als je iets van Apple, Nike, Adidas of microsoft hebt!
```

Configuración de origen de datos

El origen de datos contiene un solo puerto. Cada fila en el puerto contiene un único mensaje de Twitter.

La siguiente tabla describe la configuración del origen de datos:

Nombre de puerto	Tipo de puerto	Precisión
text	n/a	200

Configuración de transformación de clasificador

La transformación de clasificador utiliza un solo puerto de entrada y salida. El puerto de entrada de la transformación lee el campo de texto del origen de datos. El puerto de salida contiene el idioma identificado para cada mensaje de Twitter en el campo de texto. La transformación de clasificador utiliza códigos de países ISO para identificar el idioma.

La siguiente tabla describe la configuración de la transformación de clasificador:

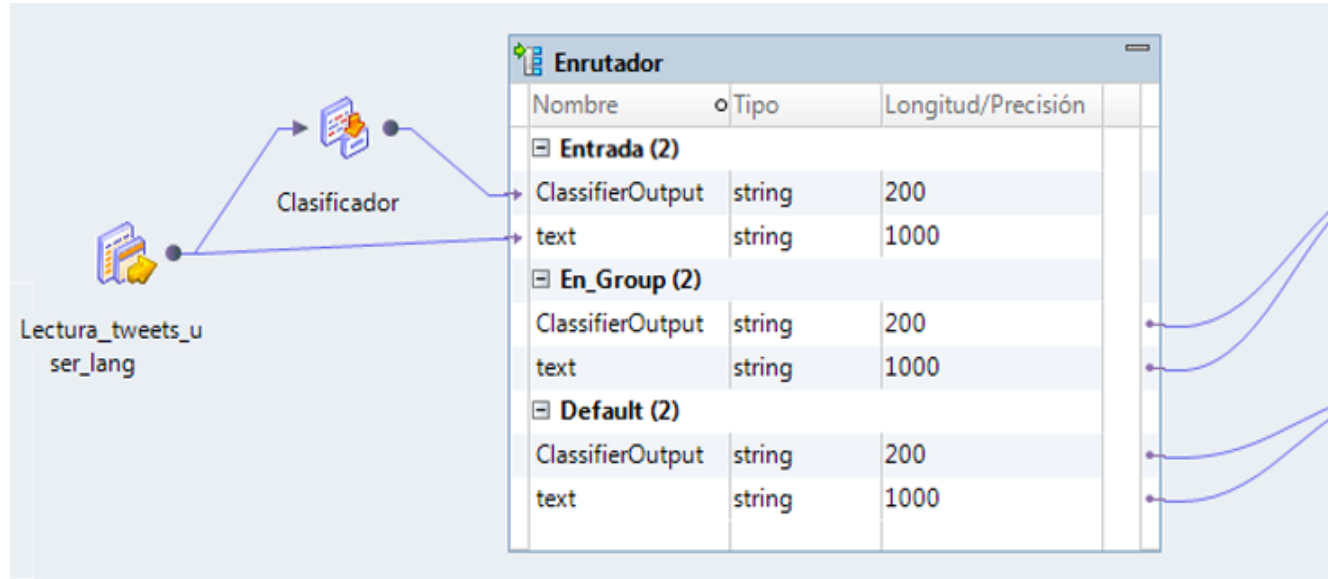
Nombre de puerto	Tipo de puerto	Precisión	Estrategia
text_input	Entrada	200	Classifier1
Classifier_Output	Salida	2	Classifier1

Configuración de transformación de enrutador

La transformación de enrutador usa dos puertos de entrada. Lee los mensajes de Twitter desde el origen de datos y los códigos ISO de países desde el clasificador de transformación. La transformación de enrutador

envía los datos en los puertos de entrada a distintos puertos de salida en función de una condición que especifique.

La siguiente imagen muestra los grupos de puertos y las conexiones de puertos de la transformación de enrutador:



La siguiente tabla describe la configuración de la transformación de enrutador:

Nombre de puerto	Tipo de puerto	Grupo de puertos	Precisión
Classifier_Output	Entrada	Entrada	2
text	Entrada	Entrada	200
Classifier_Output	Entrada	Predeterminado	2
text	Entrada	Predeterminado	200
Classifier_Output	Entrada	En_Group	2
text	Entrada	En_Group	200

Configure la transformación para crear flujos de datos para mensajes en inglés y para mensajes en otros idiomas. Para crear un flujo de datos, añada un grupo de puertos de salida a la transformación. Utilice las opciones de **Grupos** en la transformación para añadir el grupo de puertos.

Para determinar cómo la transformación enruta los datos a cada flujo de datos, debe definir una condición en un grupo de puertos. La condición identifica un puerto y especifica un valor posible en el puerto. Cuando la transformación encuentra un valor de puerto de entrada que coincide con la condición, enruta los datos de entrada al grupo de puertos que cumple la condición.

Defina la siguiente condición en En_Group:

```
ClassifierOutput='en'
```

Nota: La transformación de enrutador lee datos de dos objetos en la asignación. La transformación puede combinar los datos en cada grupo de salida porque no modifica la secuencia de filas definida en los objetos de datos.

Configuración de destino de datos

La asignación contiene un destino de datos para los mensajes de Twitter en inglés y un destino para los mensajes en otros idiomas. Debe conectar los puertos de un grupo de salida de transformación de enrutador a un destino de datos.

La siguiente tabla describe la configuración de los destinos de datos:

Nombre de puerto	Tipo de puerto	Precisión
text	n/a	200
Classifier_Output	n/a	2

Resultados de la asignación de clasificador

Cuando se ejecuta la asignación, la transformación de clasificador identifica el idioma de cada mensaje de Twitter. La transformación de enrutador escribe el texto del mensaje para los destinos de datos en función de las clasificaciones de idioma.

El siguiente fragmento de datos muestra un ejemplo de datos de destino del inglés:

ISO Country Code	Twitter Message
en	RT @Clarified: How to Downgrade Your iPhone 4 From iOS 6.x to iOS 5.x (Mac)...
en	RT @jerseyjazz: The razor was the iPhone of the early 2000s
en	RT @KrisiDevine: Apple Pie that I made for Thanksgiving. http://t.com/s9ImzFxO
en	RT @payyton3: I want apple cider

El siguiente fragmento de datos muestra un ejemplo de los datos de destino identificados para otros idiomas:

ISO Country Code	Twitter Message
es	RT @GanaphoneS3: Faltan 10 minutos para la gran rifa de un iPhone 5...
id	RT @sophieHz: Dan yang punya 2 kupon undian. Masuk dalam kotak undian yang berhadiah Samsung Champ.
pt	RT @IsabelFreitas: o galaxy tem isso isso isso e a bateria ã melhor que do iPhone

ISO Country Code	Twitter Message
es	RT @PremiusIpad: Faltan 15 minutos para la gran rifa de un iPhone 5...
nl	RT @wiesteronder: Retweet als je iets van Apple, Nike, Adidas of microsoft hebt! http://t.co/Je6Ts00H

Transformación de clasificador en un entorno no nativo

El procesamiento de transformaciones de clasificador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 10

Transformación de comparación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de comparación, 184](#)
- [Estrategias para la coincidencia de campos, 184](#)
- [Estrategias de coincidencia de identidades, 187](#)
- [Cómo configurar una estrategia de comparación, 188](#)
- [Propiedades avanzadas de Transformación de comparación, 188](#)
- [Transformación de comparación en un entorno no nativo, 189](#)

Resumen de la transformación de comparación

La transformación de comparación es una transformación pasiva que evalúa las similitudes entre pares de cadenas de entrada y calcula el grado de similitud para cada par en forma de puntuación numérica.

Cuando configure la transformación, seleccione un par de columnas y asígneles una estrategia de coincidencia.

La transformación de comparación produce puntuaciones de coincidencia en un intervalo de 0 a 1, donde 1 indica una coincidencia perfecta.

Nota: Las estrategias disponibles en la transformación de comparación también están disponibles en la transformación de coincidencia. Utilice la transformación de comparación para definir operaciones de comparación de coincidencia que se añadirán a un mapplet de coincidencia. Puede añadir varias estrategias de comparación al mapplet. Utilice la transformación de coincidencia para definir comparaciones de coincidencia en una transformación individual. Puede incrustar un mapplet de coincidencia en una transformación de coincidencia.

Estrategias para la coincidencia de campos

La transformación de comparación incluye estrategias para la coincidencia de campos predefinidas que comparan pares de campos de datos de entrada.

Bigram

Utilice el algoritmo Bigram para comparar cadenas de texto largas, como las direcciones postales introducidas en un único campo.

El algoritmo Bigram calcula una puntuación de coincidencia para dos cadenas de datos según la ocurrencia de caracteres consecutivos en ambas cadenas. El algoritmo busca pares de caracteres consecutivos que sean comunes en ambas cadenas. Divide el número de pares que coincide en ambas cadenas por el número total de pares de caracteres.

Ejemplo de Bigram

Considere las siguientes cadenas:

- larder
- lerder

Estas cadenas producen los siguientes grupos Bigram:

```
l a, a r, r d, d e, e r  
l e, e r, r d, d e, e r
```

Observe que la segunda ocurrencia de la cadena "e r" dentro de la cadena "lerder" no tiene coincidencia, puesto que no hay una segunda ocurrencia correspondiente de "e r" en la cadena "larder".

Para calcular la puntuación de coincidencia Bigram, la transformación divide el número de pares coincidentes (6) por el número total de pares en ambas cadenas (10). En este ejemplo, las cadenas son similares en un 60% y la puntuación de coincidencia es de 0,60.

Distancia de Hamming

Utilice el algoritmo Distancia de Hamming cuando la posición de los caracteres de datos sea un factor crítico; por ejemplo, en campos numéricos o de códigos como números de teléfono, códigos postales o códigos de producto.

El algoritmo de la distancia de Hamming calcula una puntuación de coincidencia para dos cadenas de datos calculando el número de posiciones en las que los caracteres difieren de una cadena de datos a otra. En el caso de que las cadenas tengan una longitud diferente, cada carácter adicional de la cadena más larga se cuenta como una diferencia de una cadena a otra.

Ejemplo de distancia de Hamming

Considere las siguientes cadenas:

- Morlow
- Marlowes

Los caracteres resaltados indican las posiciones que el algoritmo de Hamming identifica como diferentes.

Para calcular la puntuación de coincidencia Hamming, la transformación divide el número de caracteres coincidentes (5) por la longitud de la cadena más larga (8). En este ejemplo, las cadenas son similares en un 62,5% y la puntuación de coincidencia es de 0,625.

Distancia de edición

Utilice el algoritmo Distancia de edición para comparar palabras o cadenas de texto cortas, como un nombre.

El algoritmo Distancia de edición calcula el "coste" mínimo de transformar una cadena en otra por medio de la inserción, eliminación o reemplazo de caracteres.

Ejemplo de Distancia de edición

Considere las siguientes cadenas:

- Levenston
- Levenshtein

Los caracteres resaltados indican las operaciones necesarias para transformar una cadena en otra.

El algoritmo Distancia de edición divide el número de caracteres no modificados (8) entre la longitud de la cadena más larga (11). En este ejemplo, las cadenas son similares en un 72,7% y la puntuación de coincidencia es de 0,727.

Distancia de Jaro

Utilice el algoritmo Distancia de Jaro para comparar dos cadenas cuando la similitud entre los caracteres iniciales de las cadenas tenga prioridad.

La puntuación de coincidencia de la Distancia de Jaro refleja el grado de similitud entre los primeros cuatro caracteres de ambas cadenas y el número de transposiciones de carácter identificadas. La transformación pondera la importancia de la coincidencia entre los primeros cuatro caracteres utilizando el valor que se introduce en la propiedad Penalización.

Propiedades de la distancia de Jaro

Cuando se configura un algoritmo Distancia de Jaro, puede configurar las siguientes propiedades:

Penalización

Determina la penalización para la puntuación de coincidencia si los primeros cuatro caracteres en dos cadenas comparadas no son idénticos. La transformación resta el valor completo de la penalización cuando no hay ninguna coincidencia con el primer carácter. La transformación resta fracciones de la penalización según la posición que ocupan los otros caracteres sin coincidencia. El valor predeterminado de la penalización es 0,20.

Distinguir mayúsculas de minúsculas

Determina si el algoritmo Distancia de Jaro tiene en cuenta el formato de mayúsculas y minúsculas de los caracteres cuando los compara.

Ejemplo de la distancia de Jaro

Considere las siguientes cadenas:

- 391859
- 813995

Si utiliza el valor de Penalización predeterminado de 0,20 para analizar estas cadenas, el algoritmo Distancia de Jaro devuelve una puntuación de coincidencia de 0,513. Esta puntuación de coincidencia indica que las cadenas tienen una similitud del 51,3%.

Invertir la distancia de Hamming

Utilice el algoritmo Invertir la distancia de Hamming para calcular el porcentaje de posiciones de carácter que difiere entre dos cadenas, leídas de derecha a izquierda.

El algoritmo de la distancia de Hamming calcula una puntuación de coincidencia para dos cadenas de datos calculando el número de posiciones en las que los caracteres difieren de una cadena de datos a otra. En el caso de que las cadenas tengan una longitud diferente, el algoritmo cuenta cada carácter adicional de la cadena más larga se cuenta como una diferencia de una cadena a otra.

Ejemplo de Invertir la distancia de Hamming

Considere las siguientes cadenas, que utilizan la alineación de derecha a izquierda para imitar el algoritmo Invertir la distancia de Hamming:

- 1-999-9999
- **011-01**-999-9991

Los caracteres resaltados indican las posiciones que el algoritmo Invertir la distancia de Hamming identifica como diferentes.

Para calcular la puntuación de coincidencia de Invertir la distancia de Hamming, la transformación divide el número de caracteres coincidentes (9) entre la longitud de la cadena más larga (15). En este ejemplo, la puntuación de coincidencia es 0,6, lo que indica que las cadenas tienen una similitud del 60%.

Estrategias de coincidencia de identidades

La transformación de comparación incluye estrategias de coincidencia de identidades que puede usar para encontrar coincidencias en individuos, direcciones o entidades empresariales.

En la siguiente tabla se describe la operación de coincidencia que realiza cada estrategia de coincidencia de identidades:

Estrategias de coincidencia de identidades	Operación de coincidencia
Dirección	Identifica una coincidencia de dirección.
Contacto	Identifica un contacto de una organización en una ubicación individual.
Entidad corporativa	Identifica una organización por su nombre corporativo legal.
División	Identifica una organización en una dirección.
Familia	Identifica una familia por nombre de familia y dirección o número de teléfono.
Campos	Identifica los campos personalizados seleccionados.
Doméstico	Identifica los miembros de una familia que residen en la misma vivienda.
Individuo	Identifica un individuo por su nombre y por su ID o fecha de nacimiento.
Organización	Identifica una organización por su nombre.
Nombre de persona	Identifica una persona por su nombre.
Residente	Identifica una persona en una dirección.

Estrategias de coincidencia de identidades	Operación de coincidencia
Contacto en general	Identifica un contacto de una organización independientemente de cual sea su ubicación.
Doméstico en general	Identifica los miembros de una familia independientemente de su ubicación.

Nota: Las estrategias de coincidencia de identidades leen archivos de datos de referencia llamados **poblaciones**. Póngase en contacto con el usuario de Informática Administrator para obtener información sobre los archivos de poblaciones instalados en el sistema.

Cómo configurar una estrategia de comparación

Para configurar una estrategia de comparación, edite los ajustes en la vista **Estrategias** de una transformación de comparación.

1. Seleccione la vista **Estrategias**.
2. Seleccione una estrategia de comparación de la sección **Estrategias**.
3. En la sección **Campos**, haga doble clic en una celda de la columna **Campos disponibles** para seleccionar una entrada.

Nota: Debe seleccionar una entrada para cada fila que muestra el nombre de entrada en negrita en la columna **Campos de entrada**.

Propiedades avanzadas de Transformación de comparación

Configure propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de comparación.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de comparación en un entorno no nativo

El procesamiento de transformaciones de comparación en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 11

Transformación de consolidación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de consolidación, 190](#)
- [Asignaciones de consolidación, 191](#)
- [Puertos de transformación de consolidación, 191](#)
- [Vistas de transformación de consolidación, 192](#)
- [Estrategias simples, 194](#)
- [Estrategias basadas en filas, 195](#)
- [Estrategias avanzadas, 196](#)
- [Funciones de consolidación simple, 196](#)
- [Funciones de consolidación basada en filas, 201](#)
- [Ejemplo de asignación de consolidación, 204](#)
- [Cómo configurar una transformación de consolidación, 206](#)
- [Transformación de consolidación en un entorno no nativo, 206](#)

Resumen de la transformación de consolidación

La transformación de consolidación es una transformación activa que analiza los grupos de registros relacionados y crea un registro consolidado para cada grupo. Utilice la transformación de consolidación para consolidar los grupos de registros generados por transformaciones como las transformaciones de generador de claves, de coincidencia y de asociación.

La transformación de consolidación genera registros consolidados aplicando estrategias a grupos de registros relacionados. La transformación contiene un puerto de salida que indica qué registro es el consolidado. Puede optar por limitar la salida de transformación para incluir solamente los registros consolidados.

Por ejemplo, puede consolidar los grupos de registros duplicados de empleados que la transformación de coincidencia genera. La transformación de consolidación puede crear un registro consolidado a que contenga los datos fusionados de todos los registros del grupo.

Puede configurar la transformación de consolidación para utilizar diferentes tipos de estrategias según los requisitos de consolidación. Utilice estrategias simples para crear un registro consolidado a partir de varios registros. Cuando se utilizan estrategias simples, se debe especificar una estrategia para cada puerto. Utilice una estrategia basada en filas para analizar las filas en el grupo de registros y crear un registro consolidado con los valores de una de las filas. Utilice una estrategia avanzada para crear un registro consolidado aplicando una expresión que cree.

Asignaciones de consolidación

Para consolidar registros, cree una asignación que cree grupos de registros relacionados. Añada una transformación de consolidación a una asignación y configure la transformación para consolidar cada grupo de registros en un único registro principal.

Conecte una transformación de consolidación a otras transformaciones según los objetivos empresariales y los requisitos de datos. Para consolidar los registros coincidentes, se puede conectar la transformación de consolidación a una transformación de coincidencia. Para consolidar los registros como parte de la administración de registros de excepción, conecte la transformación de consolidación a una transformación de excepción. Si utiliza una transformación de generador de claves para agrupar registros, puede conectar una transformación de consolidación directamente a la transformación de generador de claves. La transformación de consolidación crea un registro consolidado para cada grupo que crea la transformación de generador de claves.

Salida de la asignación en entornos nativos y de Hadoop

Si ejecuta una asignación de consolidación en un entorno nativo y en un entorno de Hadoop, la transformación de consolidación puede generar resultados diferentes. Como la asignación se ejecuta en varios nodos en Hadoop, los registros de entrada pueden entrar en la transformación de consolidación en un orden diferente al del entorno nativo. Como resultado, la transformación puede generar diferentes conjuntos de registros supervivientes en cada entorno para el mismo conjunto de datos de entrada. Los cálculos de la transformación y los resultados consolidados son apropiados para el orden de filas de entrada en cada caso.

Para generar los mismos registros supervivientes en entornos de Hadoop y en entornos nativos, configure la transformación de consolidación para ordenar los registros en el orden siguiente:

- Primero, ordene los registros en el puerto de agrupación.
- A continuación, ordene los registros en el orden en que aparecen los puertos de entrada en la transformación.

Puertos de transformación de consolidación

Developer Tool crea un puerto de salida para cada puerto de entrada que se añade. No se pueden añadir puertos de salida a la transformación manualmente. La transformación de consolidación también incluye un puerto de salida **IsSurvivor** que indica los registros consolidados.

Uno de los puertos de entrada que añade a la transformación de consolidación debe contener claves de grupo. La transformación de consolidación requiere información de la clave de grupo porque las estrategias de consolidación procesan grupos de registros en lugar de conjuntos de datos completos.

Cuando se añade un puerto de entrada, Developer Tool crea un nombre de puerto de salida añadiendo el sufijo "1" al nombre del puerto de entrada. La transformación también incluye el puerto de salida **IsSurvivor**, que indica si un registro es el consolidado. En el caso de los registros consolidados, la transformación de consolidación escribe la cadena "Y" en el puerto **IsSurvivor**. En el caso de los registros de entrada, la transformación de consolidación escribe la cadena "N" en el puerto **IsSurvivor**.

Vistas de transformación de consolidación

La transformación de consolidación contiene vistas para los puertos, las estrategias y las propiedades avanzadas.

Vista Estrategias de la transformación de consolidación

La vista **Estrategias** contiene propiedades para las estrategias simples, basadas en filas y avanzadas.

La siguiente lista describe los tipos de estrategias de consolidación:

Estrategia simple

Una estrategia simple analiza todos los valores de un puerto de un grupo de registros y selecciona uno. Se debe especificar una estrategia simple para cada puerto. La transformación de consolidación utiliza los valores de puerto seleccionados por todas las estrategias simples para crear un registro consolidado. Algunos ejemplos de estrategias simples son el valor más frecuente de un puerto, el valor más largo de un puerto o el valor más frecuente que no está en blanco de un puerto.

Estrategia basada en filas

Una estrategia basada en filas analiza las filas del grupo de registros y selecciona una. La transformación de consolidación utiliza los valores de puerto de la fila para crear un registro consolidado. Algunos ejemplos de estrategias basadas en filas son el mayor número de caracteres, el menor número de campos en blanco o el mayor número de campos más frecuentes.

Estrategia avanzada

Una estrategia de avanzada analiza un grupo de registros mediante estrategias definidas por el usuario. Las estrategias avanzadas se construyen utilizando funciones de consolidación en una expresión. La transformación de consolidación crea un registro consolidado con base en la salida de la expresión. La expresión creada por el usuario también puede usar todas las funciones disponibles en la transformación de decisión.

Propiedades avanzadas de Transformación de consolidación

La transformación de consolidación contiene las propiedades avanzadas que determinan el comportamiento de ordenación, el modo de salida, el comportamiento de la memoria caché y el nivel de seguimiento.

Puede configurar las siguientes propiedades avanzadas:

Ordenar

Determina si la transformación ordena las filas de entrada en el puerto de datos **Agrupar por**. La propiedad está habilitada de forma predeterminada.

Seleccione la propiedad si las filas de entrada no se han ordenado previamente.

Ordenar con distinción de mayúsculas y minúsculas

Determina si la operación de ordenación distingue entre mayúsculas y minúsculas. La propiedad está habilitada de forma predeterminada.

Modo de salida

Determina si la transformación escribe todos los registros como salidas o si escribe los registros consolidados como salidas. El valor predeterminado es Todos.

Directorio de archivo de memoria caché

Especifica el directorio en el que el servicio de integración de datos escribe los datos temporales de la transformación actual. El servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos de entrada es mayor que la memoria disponible del sistema. El servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del servicio de integración de datos. El servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de archivo de memoria caché

Determina la cantidad de memoria del sistema que utiliza el servicio de integración de datos para ordenar los datos de entrada en la transformación.

Antes de ordenar los datos, el servicio de integración de datos asigna la cantidad de memoria que se haya especificado. Si la operación de ordenación genera una cantidad de datos mayor, el servicio de integración de datos escribirá los datos en exceso en el directorio de archivos de memoria caché. Si la operación de clasificación necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

La transformación lee el valor en bytes. El valor predeterminado es 400 000 bytes. El valor máximo es 2 147 483 647 bytes. Puede utilizar un parámetro para especificar el tamaño del archivo de memoria caché.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Tamaño de archivo de memoria caché

La propiedad de tamaño de archivo de memoria caché determina la cantidad de memoria del sistema que el servicio de integración de datos asigna a la transformación de consolidación para operaciones de ordenación. Configure la propiedad con un valor que sea menor o igual a la cantidad de RAM del equipo host del servicio de integración de datos.

Para un rendimiento óptimo, especifique un tamaño de archivo de memoria caché de 16 MB como mínimo. Establezca el valor del tamaño del archivo de memoria caché en bytes en la vista **Propiedades avanzadas**.

Antes de iniciar una operación de ordenación, el servicio de integración de datos asigna la memoria que especifica la propiedad de tamaño de archivo de memoria caché. El servicio de integración de datos transfiere todos los datos entrantes a la transformación de consolidación antes de realizar una operación de ordenación.

Si el volumen de datos de entrada es mayor que el tamaño del archivo de memoria caché, el servicio de integración de datos escribirá los datos en el directorio del archivo de memoria caché. Cuando escribe datos en el directorio del archivo de memoria caché, el servicio de integración de datos consume una cantidad de espacio en disco que representa al menos el doble del volumen de datos de entrada.

Utilice la siguiente fórmula para determinar el tamaño de los datos entrantes:

```
[number_of_input_rows * ( Sum (column_size) + 16]
```

La siguiente tabla enumera los posibles tipos de datos y valores de tamaño de columna que se aplicarán a los cálculos de datos de archivo de memoria caché:

Tipo de datos	Tamaño de columna
Binario	Precisión + 8. Redondee al múltiplo de 8 más cercano.
Fecha/Hora	29
Decimal, alta precisión desactivada (todas las precisiones)	16
Decimal, alta precisión activada (precisión <=18)	24
Decimal, alta precisión activada (precisión >18, <=28)	32
Decimal, alta precisión activada (precisión >28)	16
Decimal, alta precisión activada (escala negativa)	16
Doble	16
Real	16
Entero	16
Cadena, texto	Modo Unicode: 2*(precisión + 5) Modo ASCII: precisión + 9

Estrategias simples

Una estrategia simple analiza un puerto de un grupo de registros y devuelve un valor. Se debe especificar una estrategia simple para cada puerto. La transformación de consolidación utiliza los valores de puerto seleccionados por todas las estrategias simples para crear un registro consolidado.

Cuando configura una estrategia en la vista **Estrategias** de la transformación, la estrategia muestra el siguiente texto como el método de consolidación:

Utilizar predeterminado.

La estrategia predeterminada es "ID de fila máximo".

Puede elegir de entre las siguientes estrategias simples:

Media

Analiza un puerto del grupo de registros y devuelve el promedio de todos los valores.

En el caso de los tipos de datos String y Date/time, la estrategia devuelve el valor que se repite con más frecuencia.

El más largo

Analiza un puerto del grupo de registros y devuelve el valor con el mayor número de caracteres. Si hay dos o más valores con el mayor número de caracteres, la estrategia devuelve el primer valor apto.

Máximo

Analiza un puerto del grupo de registros y devuelve el valor más alto.

Para el tipo de datos String, la estrategia devuelve la cadena más larga. Para el tipo de datos Date/time, la estrategia devuelve la fecha más reciente.

Mínimo

Analiza un puerto del grupo de registros y devuelve el valor más bajo.

Para el tipo de datos String, la estrategia devuelve la cadena más corta. Para el tipo de datos Date/time, la estrategia devuelve la fecha más temprana.

Más frecuente

Analiza un puerto del grupo de registros y devuelve el valor más frecuente, incluidos los valores en blanco o nulos. Si dos o más valores son los más frecuentes, la estrategia devuelve el primer valor apto.

Más frecuente que no está en blanco

Analiza un puerto del grupo de registros y devuelve el valor más frecuente, excluyendo los valores en blanco o nulos. Si hay dos valores con el número más alto de elementos que no están en blanco, la estrategia devuelve el primer valor apto.

Más corto

Analiza un puerto del grupo de registros y devuelve el valor con el menor número de caracteres. Si hay dos o más valores que tienen el menor número de caracteres, la estrategia devuelve el primer valor apto.

Id de fila máximo

Analiza un puerto del grupo de registros y devuelve el valor con el ID de fila más alto.

Estrategias basadas en filas

Una estrategia basada en filas analiza las filas del grupo de registros y selecciona una. La transformación de consolidación utiliza los valores de puerto de la fila para crear un registro consolidado. La estrategia predeterminada es "mayor parte de los datos".

Elija una de las siguientes estrategias basadas en filas:

Mayor parte de los datos

Selecciona la fila con el mayor número de caracteres. Si hay dos o más filas con el mayor número de caracteres, la estrategia devuelve el último valor apto.

Más completos

Selecciona la fila con el número más alto de columnas que no están en blanco. Si hay dos o más filas con el mayor número de columnas que no están en blanco, la estrategia devuelve el último valor apto.

Exacto modal

Selecciona la fila con el mayor número de valores más frecuentes que no están en blanco. Por ejemplo, imaginemos una fila que tiene tres puertos que contienen a su vez los valores más frecuentes del grupo de registros. El número de valores más frecuentes de esa fila es "3".

Si hay dos o más filas con el mayor número de valores más frecuentes que no están en blanco, la estrategia devuelve el último valor apto.

Ejemplo de estrategia basada en filas

La siguiente tabla muestra un grupo de registros de ejemplo. La última columna describe los motivos por los que cada estrategia basada en filas selecciona filas diferentes en este grupo de registros.

ID de producto	Nombre	Apellido	Código postal	Selección de estrategia
2106	Bartholomew		28516	La estrategia Mayor parte de los datos selecciona esta fila porque contiene más caracteres que el resto de filas.
2236	Bart	Smith	28579	La estrategia Más completos selecciona esta fila porque tiene más columnas que no están en blanco que el resto de filas.
2236	<En blanco>	Smith	28516	La estrategia Exacto modal selecciona esta fila porque contiene el mayor número de valores más frecuentes.

Estrategias avanzadas

Puede utilizar las estrategias avanzadas para crear estrategias de consolidación a partir de funciones predefinidas. Puede utilizar funciones de consolidación y otras funciones de Informática.

Puede crear expresiones que contengan funciones de consolidación simple o basada en filas. Las funciones de consolidación simple se utilizan para construir un registro consolidado según los valores de puerto del grupo de registros. Las funciones de consolidación basada en filas se utilizan para seleccionar una fila del grupo de registros.

Las expresiones de consolidación deben llenar todos los puertos de salida en la transformación de consolidación. Si las expresiones de consolidación no utilizan todos los puertos de salida, la transformación hace que las asignaciones fallen.

Puede utilizar una estrategia simple o basada en filas como plantilla para una estrategia de avanzada. Configure una estrategia simple o basada en filas y, a continuación, seleccione Avanzada. La transformación de consolidación genera una expresión con las funciones que realizan la estrategia. Puede añadir más funciones para implementar requisitos adicionales.

Funciones de consolidación simple

Las funciones de consolidación simple seleccionan un valor de todos los valores de puerto de un grupo de registros. Cuando se utiliza una función de consolidación simple, se debe proporcionar la función con un puerto y el puerto agrupar por.

CONSOL_AVG

Analiza un puerto del grupo de registros y devuelve el promedio de todos los valores.

Sintaxis

```
CONSOL_AVG(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El promedio de todos los valores de un puerto.

Para los tipos de datos String y Date/time, la función devuelve el valor que más se repite.

Ejemplo

La siguiente expresión usa la función `CONSOL_AVG` para buscar el valor promedio del puerto de entrada `SalesTotal`:

```
SalesTotal1:= CONSOL_AVG(SalesTotal, GroupKey)
```

En esta expresión, la función `CONSOL_AVG` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro de ese grupo de registros, la función analiza el puerto `SalesTotal` y devuelve el valor promedio. La expresión escribe el valor promedio del puerto de salida `SalesTotal1`.

CONSOL_LONGEST

Analiza un puerto del grupo de registros y devuelve el valor con el mayor número de caracteres.

Sintaxis

```
CONSOL_LONGEST(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor de puerto con el mayor número de caracteres.

Si hay dos o más valores con el mayor número de caracteres, la estrategia devuelve el primer valor apto.

Ejemplo

La siguiente expresión usa la función `CONSOL_LONGEST` para analizar el puerto de entrada `FirstName` y buscar el valor con el mayor número de caracteres:

```
FirstName1:= CONSOL_LONGEST(FirstName, GroupKey)
```

En esta expresión, la función `CONSOL_LONGEST` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro de ese grupo de registros, la función analiza el puerto `FirstName` y devuelve el valor más largo. La expresión escribe este valor en el puerto de salida `FirstName1`.

CONSOL_MAX

Analiza un puerto del grupo de registros y devuelve el valor más alto.

Sintaxis

```
CONSOL_MAX(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor de puerto más alto.

Para el tipo de datos String, la función devuelve la cadena más larga. Para el tipo de datos Date/time, la función devuelve la fecha más reciente.

Ejemplo

La siguiente expresión usa la función `CONSOL_MAX` para analizar el puerto de entrada `SalesTotal` y buscar el valor máximo:

```
SalesTotal1:= CONSOL_MAX(SalesTotal, GroupKey)
```

En esta expresión, la función `CONSOL_MAX` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro del grupo de registros, la función analiza el puerto `SalesTotal` y devuelve el valor más alto. La expresión escribe este valor en el puerto de salida `SalesTotal1`.

CONSOL_MIN

Analiza un puerto del grupo de registros y devuelve el valor más bajo.

Sintaxis

```
CONSOL_MIN(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor de puerto más bajo.

Para el tipo de datos String, la función devuelve la cadena más corta. Para el tipo de datos Date/time, la función devuelve la fecha más temprana.

Ejemplo

La siguiente expresión usa la función `CONSOL_MIN` para analizar el puerto de entrada `SalesTotal` y buscar el valor mínimo:

```
SalesTotal1:= CONSOL_MIN(SalesTotal, GroupKey)
```

En esta expresión, la función `CONSOL_MIN` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro del grupo de registros, la función analiza el puerto `SalesTotal` y devuelve el valor más bajo. La expresión escribe este valor en el puerto de salida `SalesTotal1`.

CONSOL_MOSTFREQ

Analiza un puerto del grupo de registros y devuelve el valor más frecuente, incluidos los valores en blanco o nulos.

Sintaxis

```
CONSOL_MOSTFREQ(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor más frecuente, incluidos los valores en blanco o nulos.

Si dos o más valores son los más frecuentes, la estrategia devuelve el primer valor apto.

Ejemplo

La siguiente expresión usa la función `CONSOL_MOSTFREQ` para analizar el puerto de entrada `Company` y buscar el valor más frecuente:

```
Company1:= CONSOL_MOSTFREQ(Company, GroupKey)
```

En esta expresión, la función `CONSOL_MOSTFREQ` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro del grupo de registros, la función analiza el puerto `Company` y devuelve el valor más frecuente. La expresión escribe este valor en el puerto de salida `Company1`.

CONSOL_MOSTFREQ_NB

Analiza un puerto del grupo de registros y devuelve el valor más frecuente, excluyendo los valores en blanco o nulos.

Sintaxis

```
CONSOL_MOSTFREQ_NB(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor más frecuente, excluyendo los valores en blanco o nulos.

Si dos o más valores son los más frecuentes, la estrategia devuelve el primer valor apto.

Ejemplo

La siguiente expresión usa la función `CONSOL_MOSTFREQ_NB` para analizar el puerto de entrada `Company` y buscar el valor más frecuente:

```
Company1:= CONSOL_MOSTFREQ_NB(Company, GroupKey)
```

En esta expresión, la función `CONSOL_MOSTFREQ_NB` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro del grupo de registros, la función analiza el puerto `Company` y devuelve el valor más frecuente. La expresión escribe este valor en el puerto de salida `Company1`.

CONSOL_SHORTEST

Analiza un puerto del grupo de registros y devuelve el valor con el menor número de caracteres.

Sintaxis

```
CONSOL_SHORTEST(string, group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

El valor de puerto con el menor número de caracteres.

Si hay dos o más valores que tienen el menor número de caracteres, la estrategia devuelve el primer valor apto.

Ejemplo

La siguiente expresión usa la función `CONSOL_SHORTEST` para analizar el puerto de entrada `FirstName` y buscar el valor con el menor número de caracteres:

```
FirstName1:= CONSOL_SHORTEST(FirstName, GroupKey)
```

En esta expresión, la función `CONSOL_SHORTEST` utiliza el puerto `GroupKey` para identificar un grupo de registros. Dentro del grupo de registros, la función analiza el puerto `FirstName` y devuelve el valor más corto. La expresión escribe este valor en el puerto de salida `FirstName1`.

Funciones de consolidación basada en filas

Utilice las funciones de consolidación basada en filas para seleccionar un registro en un grupo de registros. Debe utilizar funciones de consolidación basada en filas en las declaraciones `IF-THEN-ELSE`.

CONSOL_GETROWFIELD

Lee la fila identificada por una función de consolidación basada en filas y devuelve el valor del puerto que especifique. Para especificar un puerto se utiliza un argumento numérico.

Debe utilizar la función `CONSOL_GETROWFIELD` junto con una de las siguientes funciones de consolidación basada en filas:

- `CONSOL_MODALLEXACT`
- `CONSOL_MOSTDATA`
- `CONSOL_MOSTFILLED`

Para cada puerto de entrada de una función de consolidación basada en filas debe utilizar una instancia de la función `CONSOL_GETROWFIELD`.

Sintaxis

```
CONSOL_GETROWFIELD(value)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/ opcional	Descripción
<i>value</i>	Obligatorio	El número que indica un puerto de entrada en la función de consolidación basada en filas. Utilice "0" para especificar el puerto situado en el extremo izquierdo de la función. Utilice los números posteriores para indicar otros puertos.

Valor de retorno

El valor del puerto que especifique. La función lee este valor desde una fila identificada por una función de consolidación basada en filas.

Ejemplo

La siguiente expresión usa la función `CONSOL_GETROWFIELD` junto con la función `CONSOL_MOSTDATA`:

```
IF (CONSOL_MOSTDATA(First_Name, Last_Name, GroupKey, GroupKey))
THEN
First_Name1 := CONSOL_GETROWFIELD(0)
Last_Name1 := CONSOL_GETROWFIELD(1)
GroupKey1 := CONSOL_GETROWFIELD(2)
ELSE
First_Name1 := First_Name
Last_Name1 := Last Name
GroupKey1 := GroupKey
ENDIF
```

En esta expresión, la función `CONSOL_MOSTDATA` analiza las filas de un grupo de registros e identifica una única fila. Las funciones `CONSOL_GETROWFIELD` utilizan números consecutivos para leer los valores de puerto de esa fila y escribir los valores en los puertos de salida.

CONSOL_MODALExact

Identifica la fila con el mayor número de valores más frecuentes.

Por ejemplo, imaginemos una fila que tiene tres puertos que contienen a su vez los valores más frecuentes del grupo de registros. El número de valores más frecuentes de esa fila es "3".

Debe usar esta función junto con la función `CONSOL_GETROWFIELD`. `CONSOL_GETROWFIELD` devuelve los valores de la fila que identifica la función `CONSOL_MODALExact`.

Sintaxis

```
CONSOL_MODALExact(string1, [string2, ..., stringN,]  
group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

TRUE para la fila que presenta el mayor número de campos más frecuentes, FALSE para todas las demás filas.

Ejemplo

La siguiente expresión usa la función `CONSOL_MODALExact` para buscar la fila que contiene el mayor número de campos más frecuentes:

```
IF (CONSOL_MODALExact(First_Name,Last_Name,GroupKey,GroupKey))  
THEN  
First_Name1 := CONSOL_GETROWFIELD(0)  
Last_Name1 := CONSOL_GETROWFIELD(1)  
GroupKey1 := CONSOL_GETROWFIELD(2)  
ELSE  
First_Name1 := First_Name  
Last_Name1 := Last_Name  
GroupKey1 := GroupKey  
ENDIF
```

En esta expresión, la función `CONSOL_MODALExact` analiza las filas de un grupo de registros e identifica una única fila. Las funciones `CONSOL_GETROWFIELD` utilizan números consecutivos para leer los valores de puerto de esa fila y escribir los valores en los puertos de salida.

CONSOL_MOSTDATA

Identifica la fila que contiene el mayor número de caracteres de todos los puertos

Debe usar esta función junto con la función `CONSOL_GETROWFIELD`. La función `CONSOL_GETROWFIELD` devuelve los valores de la fila que identifica la función `CONSOL_MOSTDATA`.

Sintaxis

```
CONSOL_MOSTDATA(string1, [string2, ..., stringN,]  
group by)
```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

TRUE para la fila que contiene el mayor número de caracteres de todos los puertos, FALSE para todas las demás filas.

Ejemplo

La siguiente expresión usa la función `CONSOL_MOSTDATA` para encontrar la fila que contiene el mayor número de caracteres:

```

IF (CONSOL_MOSTDATA(First_Name, Last_Name, GroupKey, GroupKey))
THEN
First_Name1 := CONSOL_GETROWFIELD(0)
Last_Name1 := CONSOL_GETROWFIELD(1)
GroupKey1 := CONSOL_GETROWFIELD(2)
ELSE
First_Name1 := First_Name
Last_Name1 := Last_Name
GroupKey1 := GroupKey
ENDIF

```

En esta expresión, la función `CONSOL_MOSTDATA` analiza las filas de un grupo de registros e identifica una única fila. Las funciones `CONSOL_GETROWFIELD` utilizan números consecutivos para leer los valores de puerto de esa fila y escribir los valores en los puertos de salida.

CONSOL_MOSTFILLED

Identifica la fila que contiene el número más alto de campos que no están en blanco.

Debe usar esta función junto con la función `CONSOL_GETROWFIELD`. `CONSOL_GETROWFIELD` devuelve los valores de la fila que identifica la función `CONSOL_MOSTFILLED`.

Sintaxis

```

CONSOL_MOSTFILLED(string1, [string2, ..., stringN,]
group by)

```

En la siguiente tabla se describen los argumentos de este comando:

Argumento	Obligatorio/opcional	Descripción
<i>string</i>	Obligatorio	Nombre del puerto de entrada.
<i>agrupar por</i>	Obligatorio	Nombre del puerto de entrada que contiene el identificador de grupo.

Valor de retorno

TRUE para la fila que contiene el número más alto de campos que no están en blanco, FALSE para todas las demás filas.

Ejemplo

La siguiente expresión usa la función `CONSOL_MOSTFILLED` para buscar la fila que contiene el mayor número de caracteres:

```
IF (CONSOL_MOSTFILLED(First_Name,Last_Name,GroupKey,GroupKey))
THEN
First_Name1 := CONSOL_GETROWFIELD(0)
Last_Name1 := CONSOL_GETROWFIELD(1)
GroupKey1 := CONSOL_GETROWFIELD(2)
ELSE
First_Name1 := First_Name
Last_Name1 := Last_Name
GroupKey1 := GroupKey
ENDIF
```

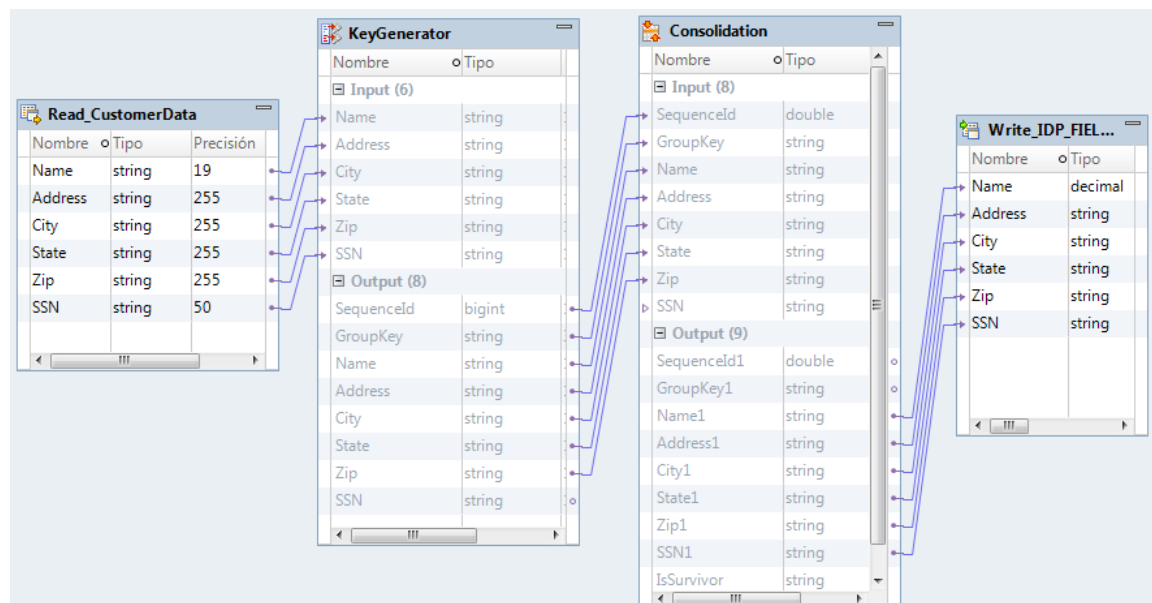
En esta expresión, la función `CONSOL_MOSTFILLED` analiza las filas de un grupo de registros e identifica una única fila. Las funciones `CONSOL_GETROWFIELD` utilizan números consecutivos para leer los valores de puerto de esa fila y escribir los valores en los puertos de salida.

Ejemplo de asignación de consolidación

Su organización necesita consolidar registros de clientes duplicados. Para consolidar los registros de clientes, debe agrupar los datos con una transformación de generador de claves y utilizar la transformación de consolidación para consolidar los registros.

Una asignación se crea con un origen de datos que contenga los registros de clientes, una transformación de generador de claves, una transformación de consolidación y un destino de datos. Esta asignación agrupa los registros de clientes, consolida los grupos y escribe un único registro consolidado.

La siguiente figura muestra la asignación:



Datos de entrada

Los datos de entrada que desea analizar contienen información de clientes.

La siguiente tabla contiene los datos de entrada para este ejemplo:

Nombre	Dirección	Ciudad	Estado	Código postal	SSN
Dennis Jones	100 All Saints Ave	New York	NY	10547	987-65-4320
Dennis Jones	1000 Alberta Rd	New York	NY	10547	987-65-4320
D Jones	100 All Saints Ave	New York	NY	10547-1521	

Transformación de generador de claves

Utilice la transformación de generador de claves para agrupar los datos de entrada con base en el puerto de código postal.

La transformación devuelve los siguientes datos:

SequenceId	GroupKey	Nombre	Dirección	Ciudad	Estado	Código postal	SSN
1	10547	Dennis Jones	100 All Saints Ave	New York	NY	10547	987-65-4320
2	10547	Dennis Jones	1000 Alberta Rd	New York	NY	10547	
3	10547	D Jones	100 All Saints Ave	New York	NY	10547-1521	987-65-4320

Transformación de consolidación

Utilice la transformación de consolidación para generar el registro consolidado.

Configure la transformación de consolidación para utilizar el tipo de estrategia basada en filas. Seleccione la estrategia Exacto modal para elegir la fila con el mayor número de valores más frecuentes. La estrategia Exacto modal utiliza los valores de la fila para generar un registro consolidado. El registro consolidado es el registro con el valor "Y" en el puerto IsSurvivor.

La transformación devuelve los siguientes datos:

GroupKey	Nombre	Dirección	Ciudad	Estado	Código postal	SSN	IsSurvivor
10547	Dennis Jones	100 All Saints Ave	New York	NY	10547	987-65-4320	N
10547	Dennis Jones	1000 Alberta Rd	New York	NY	10547		N
10547	D Jones	100 All Saints Ave	New York	NY	10547-1521	987-65-4320	N
10547	D Jones	100 All Saints Ave	New York	NY	10547-1521	987-65-4320	Y

Salida de la asignación de consolidación

Configure la transformación de consolidación de modo que la salida de asignación contenga solamente registros consolidados.

En este ejemplo, tiene certeza suficiente de que los valores más frecuentes seleccionados por la estrategia Exacto modal son los valores de puerto correctos. Para escribir solo registros consolidados en el destino de la asignación, seleccione la vista **Avanzadas** y establezca el modo de salida en "Solo superviviente".

Cuando se ejecuta la asignación, la salida de la asignación solo contiene registros consolidados.

Cómo configurar una transformación de consolidación

Cuando configura la transformación de consolidación, debe elegir los tipos de estrategia, elegir las estrategias o escribir expresiones, seleccionar un puerto de agrupación y configurar opciones avanzadas.

1. Seleccione la vista **Consolidación**.
2. Elija un tipo de estrategia.
3. Configure la estrategia.
 - Para el tipo de estrategia de simple, seleccione una estrategia para cada puerto.
 - Para el tipo de estrategia basada en filas, seleccione una estrategia.
 - Para el tipo de estrategia de avanzada, cree una expresión que utilice las funciones de consolidación.
4. En el campo Agrupar por, seleccione el puerto que contiene el identificador de grupo.
5. Habilite la ordenación en la vista **Avanzada** si los datos de entrada no están ordenados.
6. Configure la salida para que contenga los registros consolidados o todos los registros.

Transformación de consolidación en un entorno no nativo

El procesamiento de transformaciones de consolidación en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de consolidación en el motor de Blaze

La transformación de consolidación puede procesar datos de forma diferente en el entorno nativo y en un entorno no nativo.

La transformación es compatible con las siguientes restricciones:

- La transformación puede procesar registros en un orden diferente en cada entorno.
- La transformación puede identificar un registro diferente como superviviente en cada entorno.

Transformación de consolidación en el motor de Spark

La transformación de consolidación puede procesar datos de forma diferente en el entorno nativo y en un entorno no nativo.

La transformación es compatible con las siguientes restricciones:

- La transformación puede procesar registros en un orden diferente en cada entorno.

- La transformación puede identificar un registro diferente como superviviente en cada entorno.

Transformación de consolidación en el motor de Databricks Spark

La transformación de consolidación puede procesar datos de forma diferente en el entorno nativo y en un entorno no nativo.

La transformación es compatible con las siguientes restricciones:

- La transformación puede procesar registros en un orden diferente en cada entorno.
- La transformación puede identificar un registro diferente como superviviente en cada entorno.

CAPÍTULO 12

Transformación de enmascaramiento de datos

Este capítulo incluye los siguientes temas:

- [Resumen de transformación de enmascaramiento de datos, 208](#)
- [Técnicas de enmascaramiento, 209](#)
- [Reglas de enmascaramiento, 222](#)
- [Formatos de máscara especiales, 226](#)
- [Archivo de valores predeterminados, 230](#)
- [Configuración de la transformación de enmascaramiento de datos, 231](#)
- [Propiedades de tiempo de ejecución de la transformación de enmascaramiento de datos, 233](#)
- [Ejemplo de enmascaramiento de datos, 234](#)
- [Propiedades avanzadas de Transformación de enmascaramiento de datos, 237](#)
- [Transformación de enmascaramiento de datos en un entorno no nativo, 237](#)

Resumen de transformación de enmascaramiento de datos

La transformación de enmascaramiento de datos cambia datos de producción confidenciales a datos de prueba realistas para entornos que no son de producción. La transformación de enmascaramiento de datos modifica los datos de origen en función de las técnicas de enmascaramiento configuradas para cada columna.

Cree datos enmascarados para el desarrollo, la prueba, la formación y la minería de datos de software. Puede conservar las relaciones de datos en los datos enmascarados y conservar la integridad referencial entre tablas de bases de datos.

La transformación de enmascaramiento de datos proporciona reglas de enmascaramiento basadas en los tipos de datos de origen y la técnica de enmascaramiento que configure para una columna. Para las cadenas, puede restringir los caracteres que se reemplazarán en una cadena. Puede restringir los caracteres que se aplicarán en una máscara. Para los números y fechas, puede proporcionar un intervalo de números para los datos enmascarados. Puede configurar un intervalo que sea fijo o una variación de porcentaje del número original. El Servicio de integración de datos reemplaza los caracteres en función de la configuración regional que establezca para la transformación.

Técnicas de enmascaramiento

La técnica de enmascaramiento es el tipo de enmascaramiento de datos que se aplica a la columna seleccionada.

Puede seleccionar una de las siguientes técnicas de enmascaramiento para una columna de entrada:

Aleatoria

Produce resultados aleatorios no repetibles para los mismos datos de origen y reglas de enmascaramiento. Puede enmascarar los tipos de datos de fecha, numéricos y de cadena. El enmascaramiento aleatorio no requiere un valor de inicialización. Los resultados del enmascaramiento aleatorio no son deterministas.

Expresión

Aplica una expresión a una columna de origen para crear o enmascarar datos. Puede enmascarar todos los tipos de datos.

De claves

Reemplaza los datos de origen con valores repetibles. La transformación de enmascaramiento de datos produce resultados deterministas para los mismos datos de origen, reglas de enmascaramiento y valor de inicialización. Puede enmascarar tipos de datos de fecha, numéricos y de cadena.

Sustitución

Reemplaza una columna de datos con datos similares pero no relacionados de un diccionario. Puede enmascarar el tipo de datos de cadena.

Dependiente

Reemplaza los valores de una columna de origen según los valores de otra columna de origen. Puede enmascarar el tipo de datos de cadena.

Tokenization

Reemplaza los datos de origen con datos generados según criterios de enmascaramiento personalizados. La transformación de enmascaramiento de datos aplica reglas especificadas en un algoritmo personalizado. Puede enmascarar el tipo de datos de cadena.

Cifrado

Reemplaza los datos de origen con valores cifrados según los criterios de cifrado que configure en la transformación. Puede cifrar el tipo de datos de cadena.

Formatos de máscara especiales

Número de tarjeta de crédito, dirección de correo electrónico, dirección IP, número de teléfono, SSN, SIN o dirección URL. La transformación de enmascaramiento de datos aplica reglas integradas para enmascarar de manera inteligente estos tipos comunes de datos confidenciales.

Sin máscara

La transformación de enmascaramiento de datos no cambia los datos de origen.

La opción predeterminada es Sin máscara.

Enmascaramiento aleatorio

El enmascaramiento aleatorio genera datos enmascarados aleatorios no deterministas. La transformación de enmascaramiento de datos devuelve valores diferentes cuando el mismo valor de origen aparece en filas diferentes. Puede definir reglas de enmascaramiento que afecten al formato de los datos que devuelve la

transformación de enmascaramiento de datos. Enmascare valores numéricos, de cadena y de datos con el enmascaramiento aleatorio.

Enmascaramiento de valores de cadena

Configure el enmascaramiento aleatorio para generar una salida aleatoria para columnas de cadena. Para configurar las limitaciones para cada carácter de la cadena de salida, configure un formato de máscara. Configure los caracteres de filtro para definir qué caracteres de origen se van a enmascarar y los caracteres con los que se van a enmascarar.

Puede aplicar las siguientes reglas de enmascaramiento para un puerto de cadena:

Intervalo

Configure la longitud de cadena mínima y máxima. La transformación de enmascaramiento de datos devuelve una cadena de caracteres aleatorios con una longitud de cadena que se encuentra entre los valores mínimo y máximo.

Formato de máscara

Defina el tipo de carácter de sustitución para cada carácter de los datos de entrada. Puede limitar cada carácter a un tipo de carácter alfabético, numérico o alfanumérico.

Caracteres de cadena de origen

Defina los caracteres de la cadena de origen que quiera enmascarar. Por ejemplo enmascare el carácter del signo de número (#) siempre que aparezca en los datos de entrada. La transformación de enmascaramiento de datos enmascara todos los caracteres de entrada cuando la opción Caracteres de cadena de origen está en blanco.

Caracteres de reemplazo de cadena de resultado

Sustituya los caracteres de la cadena de destino por los caracteres que defina en la opción Caracteres de cadena de resultado. Por ejemplo, introduzca los siguientes caracteres para configurar cada máscara y que contenga caracteres alfabéticos en mayúscula de la A a la Z:

ABCDEFGHIJKLMN OPQRSTUVWXYZ

Enmascaramiento de valores numéricos

Cuando se enmascaran datos numéricos, se puede configurar un intervalo de valores de salida para una columna. La transformación de enmascaramiento de datos devuelve un valor que se encuentra entre los valores mínimo y máximo del intervalo en función de la precisión del puerto. Para definir el intervalo, configure los intervalos mínimo y máximo o configure un intervalo de desenfoque en función de una variación del valor de origen inicial.

Puede configurar los siguientes parámetros de enmascaramiento para datos numéricos:

Intervalo

Defina un intervalo de valores de salida. La transformación de enmascaramiento de datos devuelve datos numéricos entre los valores mínimo y máximo.

Intervalo de desenfoque

Defina un intervalo de valores de salida que se encuentren dentro de una variación fija o una variación de porcentaje de los datos de origen. La transformación de enmascaramiento de datos devuelve datos numéricos que están cerca del valor de los datos de origen. Puede configurar un intervalo y un intervalo de desenfoque.

Enmascaramiento de valores de fecha

Para enmascarar valores de fecha con enmascaramiento aleatorio, puede configurar un intervalo de fechas de salida o seleccionar una variación. Cuando se configura una variación, se selecciona una parte de la fecha para desenfocarla. Seleccione el año, el mes, el día, la hora, los minutos o los segundos. La transformación de enmascaramiento de datos devuelve una fecha que está dentro del intervalo configurado.

Puede configurar las siguientes reglas de enmascaramiento cuando se enmascara un valor de fecha y hora:

Intervalo

Establece los valores mínimo y máximo que se van a devolver para el valor de fecha y hora seleccionado.

Desenfoque

Enmascara una fecha en función de una variación que aplique a una unidad de la fecha. La transformación de enmascaramiento de datos devuelve una fecha que está dentro de la variación. Puede desenfocar el año, el mes, el día, la hora, los minutos o los segundos. Seleccione una variación baja o alta para aplicarla.

Enmascaramiento de expresiones

El enmascaramiento de expresiones aplica una expresión a un puerto para cambiar los datos o crear nuevos datos. Al configurar máscaras de expresión, cree una expresión en el Editor de expresiones. Seleccione puertos de entrada y salida, funciones, variables y operadores para generar expresiones.

Puede concatenar datos de varios puertos para crear un valor para otro puerto. Por ejemplo, debe crear un nombre de inicio de sesión. El origen tiene las columnas de nombre de pila y apellido. Enmascare el nombre de pila y el apellido desde los archivos de búsqueda. En la transformación de enmascaramiento de datos, cree otro puerto llamado Inicio de sesión. Para el puerto Inicio de sesión, configure una expresión para concatenar la primera letra del nombre de pila con el apellido:

```
SUBSTR (FIRSTNM, 1, 1) || LASTNM
```

Seleccione funciones, puertos, variables y operadores desde la interfaz de apuntar y hacer clic para reducir al mínimo el número de errores en la construcción de expresiones.

El Editor de expresiones muestra los puertos de salida que no se han configurado para el enmascaramiento de expresiones. No puede utilizar la salida desde una expresión como entrada a otra expresión. Si añade el manualmente el nombre del puerto de salida a la expresión, puede obtener resultados inesperados.

Cuando crea una expresión, compruebe que la expresión devuelve un valor que coincide con el tipo de datos del puerto. La transformación de enmascaramiento de datos devuelve cero si el tipo de datos del puerto de expresión es numérico y el tipo de datos de la expresión no es el mismo. La transformación de enmascaramiento de datos devuelve valores nulos si el tipo de datos del puerto de expresión es una cadena y el tipo de datos de la expresión no es el mismo.

Enmascaramiento de expresiones repetible

Configure una asignación de expresión repetible cuando una columna de origen se encuentra en más de una tabla y tiene que enmascarar la columna para cada tabla con el mismo valor.

Al configurar máscaras de expresión repetibles, la transformación de enmascaramiento de datos guarda los resultados de una expresión en una tabla de almacenamiento. Si la columna se encuentra en otra tabla de origen, la transformación de enmascaramiento de datos devuelve el valor enmascarado de la tabla de almacenamiento, en lugar de partir de la expresión.

Nombre del diccionario

Al configurar máscaras de expresión repetibles debe especificar un nombre de diccionario. El nombre del diccionario es una clave que permite varias transformaciones de enmascaramiento de datos para generar los mismos valores enmascarados desde los mismos valores de origen. Defina el mismo nombre de diccionario en cada transformación de enmascaramiento de datos. El nombre del diccionario puede ser cualquier texto.

Tabla de almacenamiento

La tabla de almacenamiento contiene los resultados del enmascaramiento de expresiones repetible entre sesiones. Una fila de la tabla de almacenamiento contiene la columna de origen y un par de valores enmascarados. La tabla de almacenamiento para el enmascaramiento de expresiones es una tabla independiente de la tabla de almacenamiento para la máscara de sustitución.

Cada vez que la transformación de enmascaramiento de datos enmascara un valor con una expresión repetible, busca en la tabla de almacenamiento por el nombre de diccionario, la configuración regional, el nombre de columna y el valor de entrada. Si encuentra una fila en la tabla de almacenamiento, devuelve el valor enmascarado de la tabla de almacenamiento. Si la transformación de enmascaramiento de datos no encuentra una fila, genera un valor enmascarado de la expresión para la columna.

Es necesario cifrar las tablas de almacenamiento para el enmascaramiento de expresiones cuando tiene datos no cifrados en el almacenamiento y utiliza el mismo nombre de diccionario como clave.

Cifrado de tablas de almacenamiento para el enmascaramiento de expresiones

Puede utilizar funciones de codificación del idioma de transformación para cifrar tablas de almacenamiento. Es necesario que cifre las tablas de almacenamiento cuando el cifrado de almacenamiento está habilitado.

1. Cree una asignación con la tabla de almacenamiento IDM_EXPRESSION_STORAGE como origen.
2. Cree una transformación de enmascaramiento de datos.
3. Aplique la técnica de enmascaramiento de expresión en los puertos con valor enmascarado.
4. Utilice la siguiente expresión en el puerto MASKEDVALUE:

```
Enc_Base64(AES_Encrypt(MASKEDVALUE, Key))
```

5. Vincule los puertos con el destino.

Ejemplo

Por ejemplo, una tabla Employees contiene las siguientes columnas:

```
FirstName  
LastName  
LoginID
```

En la transformación de enmascaramiento de datos, enmascare LoginID con una expresión que combine FirstName y LastName. Configure la máscara de expresión para que sea repetible. Especifique un nombre de diccionario como clave para el enmascaramiento repetible.

La tabla Computer_Users contiene LoginID, pero no contiene las columnas FirstName o LastName:

```
Dept  
LoginID  
Password
```

Para enmascarar LoginID en Computer_Users con el mismo LoginID que Employees, configure el enmascaramiento de expresiones para la columna LoginID. Habilite el enmascaramiento repetible y

especifique el mismo nombre de diccionario que ha definido para la tabla LoginID Employees. El servicio de integración recupera los valores de LoginID desde la tabla de almacenamiento.

Cree una expresión predeterminada para utilizar cuando el servicio de integración no pueda encontrar una fila en la tabla de almacenamiento para LoginID. La tabla Computer_Users no tiene las columnas FirstName o LastName, de modo que la expresión crea un LoginID con menos sentido.

Scripts de la tabla de almacenamiento

Informatica proporciona los scripts que puede ejecutar para crear la tabla de almacenamiento. Los scripts están en la ubicación siguiente:

```
<PowerCenter installation directory>\client\bin\Extensions\DataMasking
```

El directorio contiene un script para las bases de datos de Sybase, Microsoft SQL Server, IBM DB2 y Oracle. Cada script se llama <Expression_<database type>.

Reglas y directrices para el enmascaramiento de expresiones

Utilice las siguientes reglas y directrices para el enmascaramiento de expresiones:

- No puede utilizar la salida desde una expresión como entrada a otra expresión. Si añade el manualmente el nombre del puerto de salida a la expresión, puede obtener resultados inesperados.
- Utilice el método de apuntar y hacer clic para generar expresiones. Seleccione funciones, puertos, variables y operadores desde la interfaz de apuntar y hacer clic para reducir al mínimo el número de errores en la construcción de expresiones.
- Si la transformación de enmascaramiento de datos está configurada para el enmascaramiento repetible y la tabla de almacenamiento no existe, el servicio de integración sustituye los datos de origen con los valores predeterminados.

Enmascaramiento de claves

Una columna configurada para el enmascaramiento de claves devuelve datos enmascarados deterministas siempre que el valor de origen y el valor de inicialización son los mismos. La transformación de enmascaramiento de datos devuelve valores únicos para la columna.

Cuando se configura una columna para el enmascaramiento de claves, la transformación de enmascaramiento de datos crea un valor de inicialización para la columna. Puede modificar el valor de inicialización para producir datos repetibles entre diferentes transformaciones de enmascaramiento de datos. Por ejemplo, configure el enmascaramiento de claves para aplicar la integridad referencial. Utilice el mismo valor de inicialización para enmascarar una clave principal en una tabla y el valor de clave externa en otra tabla.

Puede definir reglas de enmascaramiento que afecten al formato de los datos que devuelve la transformación de enmascaramiento de datos. Enmascare los valores de cadena y numéricos con el enmascaramiento de claves.

Enmascaramiento de valores de cadena

Puede configurar el enmascaramiento de claves para generar salidas repetibles para cadenas. Configure un formato de máscara que defina las limitaciones de cada carácter en la cadena de salida. Configure los caracteres de cadena de origen que definan qué caracteres de origen enmascarar. Configure los caracteres de reemplazo de cadena de resultado para limitar los datos enmascarados a determinados caracteres.

Puede configurar las siguientes reglas de enmascaramiento para cadenas de enmascaramiento de claves:

Semilla

Aplice un valor de inicialización para generar datos enmascarados deterministas para una columna. Puede introducir un número entre 1 y 1.000.

Formato de máscara

Defina el tipo de carácter de sustitución para cada carácter de los datos de entrada. Puede limitar cada carácter a un tipo de carácter alfabético, numérico o alfanumérico.

Caracteres de cadena de origen

Defina los caracteres de la cadena de origen que quiera enmascarar. Por ejemplo enmascare el carácter del signo de número (#) siempre que aparezca en los datos de entrada. La transformación de enmascaramiento de datos enmascara todos los caracteres de entrada cuando la opción Caracteres de cadena de origen está en blanco. La transformación de enmascaramiento de datos no siempre devuelve datos únicos si el número de caracteres de cadena de origen es inferior al número de caracteres de cadena de resultado.

Caracteres de cadena de resultado

Sustituya los caracteres de la cadena de destino por los caracteres que defina en la opción Caracteres de cadena de resultado. Por ejemplo, introduzca los siguientes caracteres para configurar cada máscara y que contenga todos los caracteres alfabéticos en mayúsculas:

ABCDEFGHIJKLMN OPQRSTUVWXYZ

Enmascaramiento de valores numéricos

Configure el enmascaramiento de claves de datos de origen numéricos para generar una salida determinista. Cuando se configura en una columna el enmascaramiento de claves numérico, se asigna un valor de inicialización aleatorio a la columna. Cuando la transformación de enmascaramiento de datos enmascara los datos de origen, aplica un algoritmo de enmascaramiento que requiere una semilla.

Puede cambiar el valor de inicialización de una columna para obtener resultados repetibles si el mismo valor de origen aparece en una columna diferente. Por ejemplo, si quiere conservar una relación de claves principal-externa entre dos tablas. En cada transformación de enmascaramiento de datos, introduzca el mismo valor de inicialización para la columna de claves primarias que el valor de inicialización para la columna de claves externas. La transformación de enmascaramiento de datos proporciona resultados deterministas para los mismo valores numéricos. La integridad referencial se mantiene entre las tablas.

Enmascaramiento de valores de fecha y hora

Si puede configurar el enmascaramiento de claves para valores de fecha y hora, la transformación de enmascaramiento de datos requiere un número aleatorio como semilla. Puede cambiar la semilla para que coincida con el valor de inicialización de otra columna y así que se devuelvan valores de fecha y hora repetibles entre las columnas.

La transformación de enmascaramiento de datos puede enmascarar fechas entre 1753 y 2400 con enmascaramiento de claves. Si el año de origen es un año bisiesto, la transformación de enmascaramiento de datos devuelve un año que es también un año bisiesto. Si el mes de origen tiene 31 días, la transformación de enmascaramiento de datos devuelve un mes que tiene también 31 días. Si el mes de origen es febrero, la transformación de enmascaramiento de datos devuelve el mes de febrero.

La transformación de enmascaramiento de datos siempre genera fechas válidas.

Máscara de sustitución

La máscara de sustitución reemplaza una columna de datos con datos similares pero no relacionados. Utilice la máscara de sustitución para reemplazar datos de producción con datos de prueba realistas. Al configurar la máscara de sustitución, defina el diccionario que contiene los valores de sustitución.

La transformación de enmascaramiento de datos realiza una búsqueda en el diccionario que haya configurado. La transformación de enmascaramiento de datos reemplaza datos de origen con datos del diccionario. Los archivos de diccionario pueden contener datos de cadena, valores de fecha y hora, enteros y números de coma flotante. Especifique los valores de fecha y hora en el siguiente formato:

mm/dd/yyyy

Puede sustituir datos con valores repetibles o no repetibles. Si selecciona valores repetibles, la transformación de enmascaramiento de datos genera resultados deterministas para los mismos datos de origen y valor de semilla. Debe configurar un valor de semilla para sustituir datos con resultados deterministas. El servicio de integración mantiene una tabla de almacenamiento de origen y valores enmascarados para el enmascaramiento repetible.

Puede sustituir más de una columna de datos con valores enmascarados desde la misma fila del diccionario. Configure la máscara de sustitución para una columna de entrada. Configure el enmascaramiento de datos dependiente para otras columnas que reciben datos enmascarados de la fila del diccionario.

Diccionarios

Un diccionario es una tabla de referencia que contiene los datos de sustitución y un número de serie para cada fila de la tabla. Cree una tabla de referencia para la máscara de sustitución desde un archivo sin formato o una tabla relacional que importe al repositorio de modelos.

La transformación de enmascaramiento de datos genera un número para recuperar una fila del diccionario por el número de serie. La transformación de enmascaramiento de datos genera una clave hash para la máscara de sustitución repetible o un número aleatorio para la máscara no repetible. Puede configurar una condición de búsqueda adicional si configura la máscara de sustitución repetible.

Puede configurar un diccionario para enmascarar más de un puerto en la transformación de enmascaramiento de datos.

Cuando la transformación de enmascaramiento de datos recupera los datos de sustitución desde un diccionario, la transformación no comprueba si el valor de los datos de sustitución es el mismo que el valor original. Por ejemplo, la transformación de enmascaramiento de datos puede sustituir el nombre John con el mismo nombre (John) desde un archivo de diccionario.

El siguiente ejemplo muestra una tabla de diccionario que contiene el nombre de pila y el sexo:

SNO	SEX	FIRSTNAME
1	M	Adam
2	M	Adeel
3	M	Adil
4	F	Alice
5	F	Alison

En esta diccionario, el primer campo en la fila es el número de serie y el segundo campo es el sexo. El servicio de integración siempre busca un registro de diccionario por el número de serie. Puede añadir el sexo

como una condición de búsqueda si configura el enmascaramiento repetible. El servicio de integración recupera una fila desde el diccionario mediante una clave hash y encuentra una fila con el sexo que coincide con el sexo de los datos de origen.

Utilice las siguientes reglas y directrices cuando cree una tabla de referencia:

- Cada registro de la tabla debe tener un número de serie.
- Los números de serie son enteros secuenciales que empieza en el uno. No puede faltar un número en la secuencia de los número de serie.
- La columna de número de serie puede estar en cualquier lugar de la fila de una tabla. Puede tener cualquier etiqueta.

Si utiliza una tabla de archivo sin formato para crear la tabla de referencia, utilice las siguientes reglas y directrices:

- La primera fila de la tabla del archivo sin formato debe tener etiquetas de columnas para identificar los campos de cada registro. Los campos están separados por comas. Si la primera fila no contiene etiquetas de columnas, el servicio de integración toma los valores de los campos de la primera fila como los nombres de columna.
- Si crea una tabla de archivo sin formato en Windows y la copia en un equipo UNIX, compruebe que el formato de archivo es correcto para UNIX. Por ejemplo, Windows y UNIX utilizan diferentes caracteres para el final del marcador de línea.

Tablas de almacenamiento

La transformación de enmascaramiento de datos mantiene tablas de almacenamiento para la sustitución repetible entre sesiones. Una fila de la tabla de almacenamiento contiene la columna de origen y un par de valores enmascarados. Cada vez que la transformación de enmascaramiento de datos enmascara un valor con una sustitución repetible, busca en la tabla de almacenamiento por el nombre de diccionario, la configuración regional, el nombre de columna, el valor de entrada y la semilla. Si encuentra una fila, devuelve el valor enmascarado de la tabla de almacenamiento. Si la transformación de enmascaramiento de datos no encuentra una fila, recupera una del diccionario con una clave hash.

El formato del nombre de diccionario en la tabla de almacenamiento es diferente para un diccionario de archivo sin formato y un diccionario relacional. Un nombre de diccionario de archivo sin formato se identifica por el nombre de archivo. El nombre del diccionario relacional tiene la siguiente sintaxis:

```
<Connection object>_<dictionary table name>
```

Informatica proporciona los scripts que puede ejecutar para crear una tabla de almacenamiento relacional. Los scripts están en la ubicación siguiente:

```
<PowerCenter Client installation directory>\client\bin\Extensions\DataMasking
```

El directorio contiene un script para las bases de datos de Sybase, Microsoft SQL Server, IBM DB2 y Oracle. Cada script se llama Substitution_<database type>. Puede crear una tabla en una base de datos diferente si configura las instrucciones de SQL y la clave principal se restringe.

Es necesario para cifrar las tablas de almacenamiento para la máscara de sustitución cuando tiene datos no cifrados en el almacenamiento y utiliza el mismo valor de semilla y diccionario para cifrar las mismas columnas.

Cifrado de tablas de almacenamiento para la máscara de sustitución

Puede utilizar funciones de codificación del idioma de transformación para cifrar tablas de almacenamiento. Es necesario que cifre las tablas de almacenamiento cuando el cifrado de almacenamiento está habilitado.

1. Cree una asignación con la tabla de almacenamiento IDM_SUBSTITUTION_STORAGE como origen.

2. Cree una transformación de enmascaramiento de datos.
3. Aplique la técnica de máscara de sustitución en el valor de entrada y en los puertos de valores enmascarados.
4. Utilice la siguiente expresión en el puerto INPUTVALUE:


```
Enc_Base64(AES_Encrypt(INPUTVALUE, Key))
```
5. Utilice la siguiente expresión en el puerto MASKEDVALUE:


```
Enc_Base64(AES_Encrypt(MASKEDVALUE, Key))
```
6. Vincule los puertos con el destino.

Máscara de sustitución Propiedades

Puede configurar las siguientes reglas de enmascaramiento para la máscara de sustitución:

- **Salida repetible.** Devuelve resultados deterministas entre sesiones. La transformación de enmascaramiento de datos almacena valores enmascarados en la tabla de almacenamiento.
- **Valor de semilla.** Aplique un valor de inicialización para generar datos enmascarados deterministas para una columna. Especifique un número entre 1 y 1.000.
- **Salida única.** Fuerce la transformación de enmascaramiento de datos para crear valores de salida únicos para valores de entrada únicos. No se puede enmascarar dos valores de entrada en el mismo valor de salida. El diccionario debe tener suficientes filas únicas para habilitar salidas únicas. Cuando deshabilita la salida única, es posible que la transformación de enmascaramiento de datos no enmascare valores de entrada en valores de salida únicos. El diccionario puede contener menos filas.
- **Puerto único.** El puerto utilizado para identificar los registros únicos para la máscara de sustitución. Por ejemplo, desea enmascarar los nombres de pila en una tabla llamada Cliente. Si selecciona la columna de tabla que contiene los nombres de pila como el puerto único, la transformación de enmascaramiento de datos reemplaza los nombres de pila duplicados por el mismo valor enmascarado. Si selecciona la columna Customer_ID como el puerto único, la transformación de enmascaramiento de datos reemplaza cada nombre de pila con un valor único.
- **Optimizar uso del diccionario.**

Aplicable si selecciona la opción **Salida repetible**. Aumenta el uso de valores enmascarados de un diccionario.
- **Información del diccionario.** Configure la tabla de referencia que contiene los valores de los datos de sustitución. Haga clic en **Seleccionar origen** para seleccionar una tabla de referencia.
 - **Nombre del diccionario.** Muestra el nombre de la tabla de referencia que seleccione.
 - **Conexión de diccionario.** Muestra el nombre de la conexión que contiene el diccionario.
 - **Columna de número de serie.**- Seleccione la columna para volver a la transformación de enmascaramiento de datos.
 - **Columna de ordenación.** La columna del diccionario por la que desea ordenar las entradas. Especifique una columna de ordenación para generar resultados deterministas incluso si el orden de las entradas en el diccionario cambia. Por ejemplo, si mueve un diccionario relacional y el orden de las entradas cambia, ordene por la columna del número de serie para enmascarar los datos de manera coherente.

Nota: La columna que elija debe contener valores únicos. No puede utilizar una columna que pueda contener valores duplicados para ordenar los datos.

 - **Columna de salida.** Seleccione la columna para volver a la transformación de enmascaramiento de datos.
- **Condición de búsqueda.** Configure una condición de búsqueda para cualificar qué fila del diccionario se utilizará para el enmascaramiento de sustitución. La condición de búsqueda es similar a la cláusula

WHERE de una consulta SQL. Cuando configure una condición de búsqueda compare el valor de una columna en el origen con una columna en el diccionario.

Por ejemplo, desea enmascarar el nombre de pila. Los datos de origen y el diccionario tienen una columna de nombre de pila y una columna de sexo. Puede añadir una condición para que cada nombre de pila de mujer sea reemplazado con un nombre de mujer del diccionario. La condición de búsqueda compara el sexo del origen con el sexo del diccionario.

- **Puerto de entrada.**- Columna de datos de origen que utilizar en la búsqueda.

- **Columna de diccionario.**- Columna del diccionario que comparar con el puerto de entrada.

Reglas y directrices de la máscara de sustitución

Utilice las siguientes reglas y directrices para la máscara de sustitución:

- Si una tabla de almacenamiento no existe para una única máscara de sustitución repetible, la sesión genera un error.
- Si el diccionario no contiene filas, la transformación de enmascaramiento de datos devuelve un mensaje de error.
- Cuando la transformación de enmascaramiento de datos encuentra un valor de entrada con la configuración regional, el diccionario y la semilla en la tabla de almacenamiento, recupera el valor enmascarado, incluso si la fila ya no está en el diccionario.
- Si elimina un objeto de conexión o modifica el diccionario, debe truncar la tabla de almacenamiento. De lo contrario, puede obtener resultados inesperados.
- Si el número de valores en el diccionario es inferior al número de valores únicos de los datos de origen, la transformación de enmascaramiento de datos no puede enmascarar los datos con valores únicos repetibles. La transformación de enmascaramiento de datos devuelve un mensaje de error.

Enmascaramiento dependiente

El enmascaramiento dependiente sustituye varias columnas de datos de origen con los datos de la misma fila del diccionario.

Cuando la transformación de enmascaramiento de datos realiza la máscara de sustitución para varias columnas, los datos enmascarados pueden contener combinaciones no realistas de los campos. Puede configurar el enmascaramiento dependiente para sustituir datos por varias columnas de entrada desde la misma fila del diccionario. Los datos enmascarados reciben combinaciones válidas como "Nueva York, Nueva York" o "Chicago, Illinois".

Al configurar el enmascaramiento dependiente, primero debe configurar una columna de entrada para la máscara de sustitución. Configure otras columnas de entrada para que dependan de la columna de sustitución. Por ejemplo, seleccione la columna de código postal para la máscara de sustitución y seleccione las columnas de ciudad y estado para que dependan de la columna de código postal. El enmascaramiento dependiente garantiza que los valores de ciudad y estado sustituidos son válidos para valor del código postal sustituido.

Nota: No se puede configurar una columna para el enmascaramiento dependiente sin configurar primero una columna para la máscara de sustitución.

Configure las siguientes reglas de enmascaramiento cuando configure una columna para el enmascaramiento de dependientes:

Columna de dependientes

El nombre de la columna de entrada que ha configurado para la máscara de sustitución. La transformación de enmascaramiento de datos recupera los datos de sustitución de un diccionario

utilizando las reglas de enmascaramiento de esa columna. La columna que ha configurado para la máscara de sustitución se convierte en la columna de clave para recuperar los datos enmascarados del diccionario.

Columna de salida

El nombre de la columna del diccionario que contiene el valor para la columna que está configurando con el enmascaramiento dependiente.

Ejemplo de enmascaramiento dependiente

Un diccionario de enmascaramiento de datos puede contener filas de dirección con los siguientes valores:

SNO	CALLE	CIUDAD	ESTADO	CÓDIGO POSTAL	PAÍS
1	32 Apple Lane	Chicago	IL	61523	ES
2	776 Ash Street	Dallas	TX	75240	ES
3	2229 Big Square	Atleeville	TN	38057	ES
4	6698 Cowboy Street	Houston	TX	77001	ES

Es necesario enmascarar datos de origen con combinaciones válidas de la ciudad, estado y código postal del diccionario de direcciones.

Configure el puerto de código postal para la máscara de sustitución. Especifique las siguientes reglas de enmascaramiento para el puerto de código postal:

Regla	Valor
Nombre del diccionario	Dirección
Columna de número de serie	SNO
Columna de salida	CÓDIGO POSTAL

Configure el puerto de ciudad para el enmascaramiento dependiente. Especifique las siguientes reglas de enmascaramiento para el puerto de ciudad:

Regla	Valor
Columna de dependientes	CÓDIGO POSTAL
Columna de salida	Ciudad

Configure el puerto de estado para el enmascaramiento dependiente. Especifique las siguientes reglas de enmascaramiento para el puerto de estado:

Regla	Valor
Columna de dependientes	CÓDIGO POSTAL
Columna de salida	Estado

Cuando la transformación de enmascaramiento de datos enmascara el código postal, devuelve la ciudad y el estado correctos para el código postal de la fila de diccionario.

Enmascaramiento de Tokenization

Utilice la técnica de enmascaramiento de tokenization para enmascarar datos de cadena de origen en función de los criterios que especifique en un algoritmo. Por ejemplo, puede crear un algoritmo que contiene una dirección de correo electrónico falsa para reemplazar las entradas de los campos en los datos de origen.

Puede configurar el formato de los datos enmascarados mediante el enmascaramiento de tokenization. Debe asignar un nombre de tokenizer al algoritmo de enmascaramiento para poder utilizarlo. El nombre de tokenizer hace referencia al algoritmo de enmascaramiento (JAR) que se usa. Especifique el nombre de tokenizer cuando aplique la técnica de enmascaramiento de tokenization.

Configuración del enmascaramiento de Tokenization

Realice las tareas siguientes antes de utilizar la técnica de enmascaramiento de tokenization:

1. Busque el directorio `tokenprovider` en la ruta de acceso: `<Informatica_home>\services\shared`.
2. Abra el siguiente archivo XML: `com.informatica.products.ilm.tx-tokenizerprovider.xml`.
3. Añada el nombre de tokenizer y el nombre totalmente cualificado del archivo de clase para cada tokenizer que desee utilizar. Implemente la clase de tokenizer dentro de la clase `com.informatica.products.ilm.tx-tokenprovider-<Build-Number>.jar` en el directorio `tokenprovider`. Para cada tokenizer, introduzca la información en el archivo XML como en el siguiente ejemplo:

```
<TokenizerProvider>
  <Tokenizer Name="CCTokenizer"
  ClassName="com.informatica.tokenprovider.CCTokenizer"/>
</TokenizerProvider>
```

Donde:

- Nombre de tokenizer es nombre definido por el usuario entre comillas.
- ClassName es el nombre definido por el usuario para el atributo CLASSNAME. Implemente esto desde `com.informatica.products.ilm.tx-tokenprovider-<Build-Number>.jar`.

Después de la configuración, puede utilizar técnica de enmascaramiento de tokenization. Introduzca el nombre de tokenizer para especificar el algoritmo que utilizar cuando se crea una asignación.

Cifrado

El enmascaramiento de cifrado aplica algoritmos de cifrado para enmascarar los datos de origen.

Enmascare cualquier tipo de datos de cadena con el enmascaramiento de cifrado.

Puede elegir entre preservar el formato y la longitud de los datos de origen o solo la longitud de los datos de origen. También puede optar por cambiar el formato y la longitud de los datos de origen después de cifrarlos.

Puede elegir los caracteres que no quiera cifrar.

Después de cifrar los datos de origen, también puede descifrarlos para recuperar los datos originales. Para descifrar los datos, debe crear y ejecutar una asignación que utilice la misma técnica de cifrado con la misma frase de contraseña que utilizó para cifrar los datos de origen. Establezca el modo en Descifrado.

Nota: Si los datos de origen contienen caracteres UTF-8 de cuatro bytes, no podrá usar el cifrado para enmascarar los datos.

Seleccione una de las siguientes técnicas de cifrado:

Preservar formato y metadatos

Use la opción de cifrado Preservar formato y metadatos para preservar el formato y la longitud de los datos de origen. Cuando decide preservar el formato y los metadatos, todos los caracteres en mayúsculas se reemplazan por caracteres en mayúsculas, los caracteres en minúsculas se reemplazan por caracteres en minúsculas, los números se reemplazan por números y los caracteres especiales se reemplazan por caracteres especiales tras el cifrado. Así, por ejemplo, la dirección de correo electrónico `Abc123@xyz.com` podría pasar a ser `Mpz849#dje!kuw`. Si hubiéramos configurado los caracteres "@" y "." de este ejemplo como No cifrar los caracteres, la dirección de correo electrónico sería `Mpz849@dje.kuw`.

Preservar metadatos

Use la opción de cifrado Preservar metadatos para preservar solo la longitud de los datos de origen. Si decide preservar los metadatos, la longitud de los datos seguirá siendo la misma tras el cifrado. Así, por ejemplo, el nombre `Alexander` sería `jl6#HB91v`; es decir, la longitud es la misma que la de los datos de origen.

Cambiar los metadatos

Use la opción de cifrado Cambiar los metadatos para modificar la longitud de los datos de origen tras el cifrado. Si decide cambiar los metadatos, los datos cifrados no conservarán ni la longitud ni el formato de los datos de origen. Así, por ejemplo, un nombre de ciudad como `Lisboa` podría convertirse en `Xuep@8f5, fmch529` o `6ky#ke33h*we`.

Nota: Antes de usar la opción de cifrado Cambiar los metadatos, debe cambiar la precisión de la columna de la base de datos a la que quiera aplicar el cifrado.

Use la siguiente fórmula para calcular la precisión y redondear el valor al alza al siguiente entero de valor superior:

$$\text{Required Precision} = (1.33 * \text{Original Precision}) + 24$$

Tras cambiar la precisión de la columna en la base de datos, hay que actualizar la precisión de la columna en la asignación. Para actualizar la precisión de la columna, puede volver a importar los metadatos desde la base de datos actualizada o cambiar manualmente la precisión de la columna en cada transformación de la asignación.

Reglas de enmascaramiento

Las reglas de enmascaramiento son opciones que puede configurar después de seleccionar la técnica de enmascaramiento.

Si selecciona una técnica de enmascaramiento aleatoria o de claves, puede configurar el formato de máscara, los caracteres de cadena de origen y los caracteres de cadena de resultado. Puede configurar el intervalo o el desenfoque con el enmascaramiento aleatorio.

En la siguiente tabla se describen las reglas de enmascaramiento que puede configurar para cada técnica de enmascaramiento:

Regla de enmascaramiento	Descripción	Técnica de enmascaramiento	Tipo de datos de origen
Formato de máscara	Máscara que limita cada carácter de una cadena de salida a un carácter alfabético, numérico o alfanumérico.	Aleatoria y de claves	Cadena
Caracteres de cadena de origen	Conjunto de caracteres de origen que se van a enmascarar o excluir del enmascaramiento.	Aleatoria y de claves	Cadena
Caracteres de reemplazo de cadena de resultado	Conjunto de caracteres que se van a incluir o excluir de la máscara.	Aleatoria y de claves	Cadena
Intervalo	Un intervalo de valores de salida. <ul style="list-style-type: none"> - Numeric. La transformación de enmascaramiento de datos devuelve datos numéricos que se encuentran entre los valores mínimo y máximo. - Cadena. Devuelve una cadena de caracteres aleatorios entre la longitud de cadena mínima y la máxima. - Date/Time. Devuelve una fecha y una hora que están entre los valores de fecha y hora mínimo y máximo. 	Aleatoria	Numérico Cadena Fecha/Hora
Desenfoque	Intervalo de valores de salida con una variación fija o de porcentaje de los datos de origen. La transformación de enmascaramiento de datos devuelve datos que están cerca del valor de los datos de origen. Las columnas de fecha y hora requieren una variación fija. las columnas requieren una variación fija.	Aleatoria	Numérico Fecha/Hora

Formato de máscara

Configure un formato de máscara para limitar cada carácter de la columna de salida a un carácter alfabético, numérico o alfanumérico. Utilice los siguientes caracteres para definir un formato de máscara:

A, D, N, X, +, R

Nota: El formato de máscara contiene caracteres en mayúscula. Si introduce un carácter de máscara en minúscula, la transformación de enmascaramiento de datos convierte este carácter a mayúsculas.

En la siguiente tabla se describen los caracteres de formato de máscara:

Carácter	Descripción
A	Caracteres alfabéticos. Por ejemplo, caracteres ASCII de la "a" a la "z" y de la "A" a la "Z".
D	Dígitos de 0 a 9. La transformación de enmascaramiento de datos devuelve una "X" para los caracteres distintos de los dígitos de 0 a 9.
N	Caracteres alfanuméricos. Por ejemplo, caracteres ASCII de la "a" a la "z", de la "A" a la "Z" y del 0 al 9.
X	Cualquier carácter. Por ejemplo, un carácter alfanumérico o un símbolo.
+	Sin máscara.
R	Caracteres restantes. R especifica que los caracteres restantes de la cadena pueden ser de cualquier tipo. R debe aparecer como último carácter de la máscara.

Por ejemplo, el nombre de un departamento tiene el siguiente formato:

```
nnn-<department_name>
```

Puede configurar una máscara para hacer que los tres primeros caracteres sean numéricos, el nombre del departamento sea alfabético y el guión se quede en la salida. Configure el siguiente formato de máscara:

```
DDD+AAAAAAAAAAAAAAAA
```

La transformación de enmascaramiento de datos reemplaza los tres primeros caracteres por caracteres numéricos. No reemplaza el cuarto carácter. La transformación de enmascaramiento de datos reemplaza los caracteres restantes por caracteres alfabéticos.

Si no define un formato de máscara, la transformación de enmascaramiento de datos reemplaza cada uno de los caracteres de origen por un carácter. Si el formato de la máscara es más largo que la cadena de entrada, la transformación de enmascaramiento de datos ignora los caracteres extra del formato de máscara. Si el formato de máscara es más corto que la cadena de origen, la transformación de enmascaramiento de datos enmascara los caracteres restantes con el formato R.

Nota: No puede configurar un formato de máscara con la opción de intervalo.

Caracteres de cadena de origen

Los caracteres de cadena de origen son caracteres de origen que selecciona para enmascararlos o no enmascararlos. La posición de los caracteres en la cadena de origen es indiferente. Los caracteres de origen distinguen mayúsculas de minúsculas.

Puede configurar tantos caracteres como desee. Cuando la opción Caracteres está en blanco, la transformación de enmascaramiento de datos reemplaza todos los caracteres de origen de la columna.

Seleccione una de las siguientes opciones para los caracteres de cadena de origen:

Enmascarar solo

La transformación de enmascaramiento de datos enmascara los caracteres del origen que haya configurado como caracteres de cadena de origen. Por ejemplo, si introduce los caracteres A, B y c, la transformación de enmascaramiento de datos reemplaza A, B o c por un carácter diferente siempre que aparezca en los datos de origen. Cualquier carácter de origen que no sea A, B o c no cambiará. La máscara distingue mayúsculas de minúsculas.

Enmascarar todos excepto

Enmascara todos los caracteres exceptuando los caracteres de cadena de origen que aparezcan en la cadena de origen. Por ejemplo, si introduce el carácter de origen de filtro "-" y selecciona Enmascarar todos excepto, la transformación de enmascaramiento de datos no reemplaza el carácter "-" cuando aparece en los datos de origen. El resto de caracteres de origen cambiará.

Ejemplo de cadena de origen

Un archivo de origen tiene una columna que se llama Dependientes. La columna Dependientes contiene más de un nombre separados por comas. Tiene que enmascarar la columna Dependientes y conservar la coma en los datos de prueba para delimitar los nombres.

En la columna Dependientes, seleccione Caracteres de cadena de origen. Seleccione No enmascarar y escriba "," como el carácter de origen que hay que ignorar. No escriba las comillas.

La transformación de enmascaramiento de datos reemplaza a todos los caracteres de la cadena de origen exceptuando las comas.

Caracteres de reemplazo de cadena de resultado

Los caracteres de reemplazo de cadena de resultado son caracteres que selecciona como caracteres de sustitución en los datos enmascarados. Si configura caracteres de reemplazo de cadena de resultado, la transformación de enmascaramiento de datos reemplaza los caracteres de la cadena de origen por los caracteres de reemplazo de cadena de resultado. Para evitar que se genere la misma salida para valores de entrada diferentes, configure un intervalo amplio de caracteres de sustitución, o enmascare solo unos cuantos caracteres de origen. La posición de cada carácter en la cadena es indiferente.

Seleccione una de las siguientes opciones para los caracteres de reemplazo de cadena de resultado:

Utilizar solo

Enmascare el origen solo con los caracteres que defina como caracteres de reemplazo de cadena de resultado. Por ejemplo, si introduce los caracteres A, B y c, la transformación de enmascaramiento de datos reemplaza cada carácter de la columna de origen con una A, B o c. La palabra "horse" se podría reemplazar por "BAcBA".

Utilizar todos excepto

Enmascare el origen con cualquier carácter exceptuando los caracteres que defina como caracteres de reemplazo de cadena de resultado. Por ejemplo, si introduce A, B y c como caracteres de reemplazo de cadena de resultado, los datos enmascarados no tendrán nunca los caracteres A, B o c.

Ejemplo de caracteres de reemplazo de cadena de resultado

Para reemplazar todas las comas de la columna Dependientes por punto y coma, realice las siguientes tareas:

1. Configure la coma como un carácter de cadena de origen y seleccione Enmascarar solo.
La transformación de enmascaramiento de datos enmascara solo la coma cada vez que aparece en la columna Dependientes.
2. Configure el punto y coma como un carácter de reemplazo de cadena de resultado y seleccione Utilizar solo.
La transformación de enmascaramiento de datos reemplaza las comas de la columna Dependientes por un punto y coma.

Intervalo

Establezca un intervalo para los datos numéricos, de fecha o de cadena. Al definir un intervalo para los valores numéricos o de fecha, la transformación de enmascaramiento de datos enmascara los datos de origen con un valor comprendido entre los valores mínimo y máximo. Al configurar un intervalo para una cadena se está estableciendo un intervalo de longitudes de cadena.

Intervalo de cadenas

Al configurar el enmascaramiento de cadenas aleatorias, la transformación de enmascaramiento de datos genera cadenas cuya longitud no es la misma que la de la cadena de origen. De forma opcional, puede configurar una anchura mínima y máxima. Estos valores de anchura deben ser números positivos y enteros. Cada anchura debe ser menor o igual que la precisión del puerto.

Intervalo numérico

Establezca los valores mínimo y máximo de una columna numérica. El valor máximo debe ser menor o igual que la precisión del puerto. El intervalo predeterminado va del uno a la longitud de la precisión del puerto.

Intervalo de fechas

Establezca los valores mínimo y máximo para un valor de fecha y hora. Los campos mínimo y máximo contienen las fechas mínima y máxima predeterminadas. El formato de la fecha y hora predeterminadas es MM/DD/AAAA HH24:MI:SS. La fecha y hora máxima debe ser posterior a la mínima.

Desenfoque

El desenfoque crea un valor de salida de una variación fija o de porcentaje del valor de fecha de origen. Configure el desenfoque para devolver un valor aleatorio próximo al valor original. Es posible desenfocar valores numéricos y de fecha.

Desenfoque de valores numéricos

Seleccione una variación fija o de porcentaje para desenfocar un valor numérico de origen. El valor de desenfoque inferior es una variación por debajo del valor de origen. El valor de desenfoque superior es una variación por encima del valor de origen. Tanto el valor inferior como el superior deben ser igual o mayor que cero. Cuando la transformación de enmascaramiento de datos devuelva una fecha enmascarada, la fecha numérica estará dentro del intervalo establecido.

En la siguiente tabla se indican los resultados de enmascaramiento para los valores del intervalo de desenfoque cuando el valor de origen de entrada es 66:

Tipo de desenfoque	Bajo	Alto	Resultado
Fijo	0	10	Entre 66 y 76
Fijo	10	0	Entre 56 y 66
Fijo	10	10	Entre 56 y 76
Porcentaje	0	50	Entre 66 y 99

Tipo de desenfoque	Bajo	Alto	Resultado
Porcentaje	50	0	Entre 33 y 66
Porcentaje	50	50	Entre 33 y 99

Desenfoque de valores de fecha

Configure el desenfoque para enmascarar una fecha como una variación de la fecha de origen. Seleccione una unidad de la fecha en la que desee aplicar la variación. Puede seleccionar el año, el mes, el día o la hora. Introduzca los límites inferior y superior para establecer una variación tanto por arriba como por abajo de la unidad de la fecha de origen. La transformación de enmascaramiento de datos aplica la variación y devuelve una fecha que esté incluida en ella.

Por ejemplo, para restringir la fecha enmascarada a una fecha incluida en los dos años siguientes a la fecha de origen, seleccione el año como la unidad. Introduzca dos fechas como límites inferior y superior. Si una fecha de origen es el 02/02/2006, la transformación de enmascaramiento de datos devolverá una fecha comprendida entre el 02/02/2004 y el 02/02/2008.

De forma predeterminada, la unidad de desenfoque es el año.

Formatos de máscara especiales

Los formatos de máscara especiales son máscaras que puede aplicar a los tipos de datos comunes. Con un formato de máscara especial, la transformación de enmascaramiento de datos devuelve un valor enmascarado con un formato realista, pero que no es un valor válido.

Por ejemplo, cuando enmascara un SSN, la transformación de enmascaramiento de datos devuelve un SSN que tiene el formato correcto pero no es válido. Puede configurar máscaras repetibles para números de la seguridad social.

Configure máscaras especiales para los siguientes tipos de datos:

- Números de la seguridad social
- Números de tarjeta de crédito
- Números de teléfono
- Direcciones URL
- Direcciones de correo electrónico
- Direcciones IP
- Números de la seguridad social

Cuando el formato de datos de origen o el tipo de datos no es válido para una máscara, el servicio de integración de datos aplica una máscara predeterminada a los datos. El servicio de integración aplica valores enmascarados desde el archivo de valores predeterminados. Puede editar el archivo de valores predeterminados y modificar estos valores predeterminados.

Enmascaramiento del número de la tarjeta de crédito

La transformación de enmascaramiento de datos genera lógicamente un número de tarjeta de crédito válido cuando enmascara un número de tarjeta de crédito válido. La longitud del número de tarjeta de crédito de origen debe encontrarse entre 13 y 19 dígitos. El número de tarjeta de crédito de entrada debe tener una suma de comprobación válida en función de las reglas del sector de tarjetas de crédito.

El número de tarjeta de crédito de origen puede contener números, espacios y guiones. Si la tarjeta de crédito tiene caracteres incorrectos, o una longitud que no sea la adecuada, el servicio de integración escribe un error para el registro de la sesión. El servicio de integración aplica una máscara de números de tarjetas de crédito predeterminada cuando los datos de origen no son válidos.

La transformación de enmascaramiento de datos no enmascara el número de identificación bancaria (BIN) de seis dígitos. Por ejemplo, la transformación de enmascaramiento de datos puede enmascarar el número de tarjeta de crédito 4539 1596 8210 2773 como 4539 1516 0556 7067. La transformación de enmascaramiento de datos crea un número enmascarado con una suma de comprobación válida.

Enmascaramiento de la dirección de correo electrónico

Utilice la transformación de enmascaramiento de datos para enmascarar la dirección de correo electrónico que contiene el valor de cadena. La transformación de enmascaramiento de datos puede enmascarar una dirección de correo electrónico con caracteres ASCII aleatorios o reemplazar la dirección de correo electrónico con una dirección de correo electrónico realista.

Puede aplicar los siguientes tipos de enmascaramiento con la dirección de correo electrónico:

Enmascaramiento de correo electrónico estándar

La transformación de enmascaramiento de datos devuelve caracteres ASCII aleatorios cuando enmascara una dirección de correo electrónico. Por ejemplo, la transformación de enmascaramiento de datos puede enmascarar `Georgesmith@yahoo.com` como `KtrlupQAPyk@vdSKh.BIC`. El valor predeterminado es estándar.

Enmascaramiento de correo electrónico avanzado

La transformación de enmascaramiento de datos enmascara la dirección de correo electrónico con otra dirección de correo electrónico realista derivada de los puertos de salida de la transformación o de las columnas del diccionario.

Enmascaramiento de correo electrónico avanzado

Con el tipo de enmascaramiento de correo electrónico avanzado, puede enmascarar la dirección de correo electrónico con otra dirección de correo electrónico realista. La transformación de enmascaramiento de datos crea la dirección de correo electrónico a partir de las columnas del diccionario o de los puertos de salida de la transformación.

Puede crear la parte local en la dirección de correo electrónico a partir de los puertos de salida de asignación. O puede crear la parte local de la dirección de correo electrónico a partir de las columnas de la tabla relacional o del archivo sin formato.

La transformación de enmascaramiento de datos puede crear el nombre de dominio para la dirección de correo electrónico a partir de un valor constante o de un valor aleatorio en el diccionario del dominio.

Puede crear un enmascaramiento de correo electrónico avanzado basado en las siguientes opciones:

Dirección de correo electrónico basada en puertos dependientes

Puede crear una dirección de correo electrónico basada en los puertos de salida de la transformación de enmascaramiento de datos. Seleccione los puertos de salida de la transformación para las columnas de nombre y de apellido. La transformación de enmascaramiento de datos enmascara el nombre, el apellido, o ambos en función de los valores que especifique para la longitud del nombre y el apellido.

Dirección de correo electrónico basada en un diccionario

Puede crear una dirección de correo electrónico basada en las columnas de un diccionario. Seleccione una tabla de referencia como el origen para el diccionario.

Seleccione las columnas del diccionario para el nombre y el apellido. La transformación de enmascaramiento de datos enmascara el nombre, el apellido, o ambos en función de los valores que especifique para la longitud del nombre y el apellido.

Parámetros de configuración para un tipo de enmascaramiento de direcciones de correo electrónico avanzado

Especifique los parámetros de configuración cuando configure el enmascaramiento de direcciones de correo electrónico avanzado.

Puede especificar los siguientes parámetros de configuración:

Delimitador

Puede seleccionar un delimitador, tales como un punto, guión, o guión bajo, para separar el nombre y el apellido en la dirección de correo electrónico. Si no desea separar el nombre y el apellido en la dirección de correo electrónico, deje en blanco el delimitador.

Columna FirstName

Seleccione un puerto de salida de transformación de enmascaramiento de datos o una columna de diccionario para enmascarar el nombre en la dirección de correo electrónico.

Columna LastName

Seleccione un puerto de salida de transformación de enmascaramiento de datos o una columna de diccionario para enmascarar el apellido en la dirección de correo electrónico.

Longitud de las columnas FirstName o LastName

Restringe el número de caracteres para enmascarar las columnas de nombre y apellido. Por ejemplo, los datos de entrada son Timothy para el nombre y Smith para el apellido. Seleccione 5 como la longitud de la columna de nombre. Seleccione 1 como la longitud de la columna de apellido con un punto como delimitador. La transformación de enmascaramiento de datos genera la siguiente dirección de correo electrónico:

```
timot.s@<nombre_de_dominio>
```

DomainName

Puede utilizar un valor constante, como gmail.com, para el nombre de dominio. También puede especificar otro archivo de diccionario que contenga una lista de nombres de dominio. El diccionario de dominio puede ser un archivo sin formato o una tabla relacional.

Enmascaramiento de direcciones IP

La transformación de enmascaramiento de datos enmascara una dirección IP como otra dirección IP dividiéndola en cuatro números, separados por un punto. El primer número es la red. La transformación de enmascaramiento de datos enmascara el número de la red dentro de un intervalo de red.

La transformación de enmascaramiento de datos enmascara una dirección IP de Clase A como una dirección IP de Clase A y una dirección 10.x.x.x como una dirección 10.x.x.x. La transformación de enmascaramiento de datos no enmascara la dirección de red privada y de clase. Por ejemplo, la transformación de enmascaramiento de datos puede enmascarar 11.12.23.34 como 75.32.42.52 y 10.23.24.32 como 10.61.74.84.

Nota: Cuando enmascara muchas direcciones IP, la transformación de enmascaramiento de datos puede devolver valores no exclusivos porque no enmascara la clase o red privada de las direcciones IP.

Enmascaramiento del número de teléfono

La transformación de enmascaramiento de datos enmascara un número de teléfono sin cambiar el formato del número de teléfono original. Por ejemplo, la transformación de enmascaramiento de datos puede enmascarar el número (607)382 0658 como (408)256 3106.

Los datos de origen pueden contener números, espacios, guiones y paréntesis. El servicio de integración no enmascara caracteres alfabéticos o especiales.

La transformación de enmascaramiento de datos puede enmascarar datos de cadenas, enteros y bigint.

Enmascaramiento del número de seguridad social

La transformación de enmascaramiento de datos genera un número de la seguridad social que no es válido según la lista de grupo alto de la administración de la seguridad social. La lista de grupo alto contiene números válidos que ha emitido la Administración de la Seguridad Social.

La lista de grupo alto predeterminada es un archivo de texto en la ubicación siguiente:

```
<Installation Directory>\infa_shared\SrcFiles\highgroup.txt
```

Para utilizar la el archivo de la lista de grupo alto en los flujos de trabajo, copie el archivo de texto en el directorio de origen que configure para el servicio de integración de datos.

La transformación de enmascaramiento de datos genera números de SSN que no están en la lista de grupo alto. La Administración de la Seguridad Social actualiza la lista de grupo alto cada mes. Descargue la última versión de la lista en la siguiente ubicación:

```
http://www.socialsecurity.gov/employer/ssns/highgroup.txt
```

Formato de números de la seguridad social

La transformación de enmascaramiento de datos acepta cualquier formato de SSN que contenga nueve dígitos. Los dígitos se pueden limitar a cualquier conjunto de caracteres. Por ejemplo, la transformación de enmascaramiento de datos acepta el siguiente formato: +=54-*9944\$#789-, * () ”.

Requisito de código de área

La transformación de enmascaramiento de datos devuelve un número de la seguridad social que no es válido con el mismo formato que el origen. Los primeros tres dígitos del SSN indican el código de área. La transformación de enmascaramiento de datos no enmascara el código de área. Sí lo hace con el número de grupo y el número de serie. El SSN de origen debe contener un código de área válido. La transformación de enmascaramiento de datos posiciona el código de área en la lista de grupos altos y establece un intervalo de números sin utilizar que se pueden aplicar como datos enmascarados. Si el SSN no es válido, la transformación de enmascaramiento de datos no enmascarará los datos de origen.

Enmascaramiento de números de la seguridad social repetibles

La transformación de enmascaramiento de datos devolverá números de seguridad social deterministas con enmascaramiento repetible. La transformación de enmascaramiento de datos no puede devolver todos los números de seguridad social exclusivos porque es incapaz de devolver los números de seguridad social válidos que la administración de la seguridad social ha emitido.

Enmascaramiento de direcciones URL

La transformación de enmascaramiento de datos analiza una URL buscando la cadena "://" y analizando la subcadena situada a la derecha de la misma. La dirección URL de origen debe contener la cadena "://". La dirección URL de origen puede contener números y caracteres alfabéticos.

La transformación de enmascaramiento de datos no enmascara el protocolo de la dirección URL. Por ejemplo, si la dirección URL es `http://www.yahoo.com`, la transformación de enmascaramiento de datos puede devolver `http://MgLaHjCa.VsD/`. La transformación de enmascaramiento de datos puede generar una URL que no es válida.

Nota: La transformación de enmascaramiento de datos siempre devuelve caracteres ASCII para una dirección URL.

Enmascaramiento del número de seguro social

La transformación de enmascaramiento de datos enmascara un número de la seguridad social que tiene nueve dígitos. Los dígitos se pueden limitar a cualquier conjunto de caracteres.

Si el número no contiene delimitadores, el número enmascarado no contiene delimitadores tampoco. En caso contrario, el número enmascarado tiene el siguiente formato:

```
xxx-xxx-xxx
```

Números de SIN repetibles

Puede configurar la transformación de enmascaramiento de datos para que devuelva valores de SIN repetibles. Si configura un puerto para el enmascaramiento de SIN repetible, la transformación de enmascaramiento de datos devolverá datos enmascarados deterministas cada vez que el valor de SIN de origen y el valor de inicialización sean iguales.

Para obtener números de SIN repetibles, habilite **Valores repetibles** e introduzca un número de semilla. La transformación de enmascaramiento de datos devuelve valores únicos para cada SIN.

Dígito de inicio de SIN

Puede definir el primer dígito del SIN enmascarado.

Habilite **Dígito de inicio** e introduzca el dígito. La transformación de enmascaramiento de datos crea números de SIN enmascarados que empiezan por el número que introduzca.

Archivo de valores predeterminados

Cuando el formato de datos de origen o el tipo de datos no es válido para una máscara, el servicio de integración de datos aplica una máscara predeterminada a los datos. El servicio de integración aplica valores enmascarados desde el archivo de valores predeterminados. Puede editar el archivo de valores predeterminados y modificar estos valores predeterminados.

El archivo de valores predeterminados es un archivo XML que se encuentra en la siguiente ubicación:

```
<Directorio de instalación>\infa_shared\SrcFiles\defaultValue.xml
```

Para utilizar el archivo de valores predeterminados en los flujos de trabajo, copie el archivo de valores predeterminados en el directorio de origen que haya configurado para el servicio de integración de datos.

El archivo defaultvalue.xml contiene los siguientes pares nombre-valor:

```
<?xml version="1.0" standalone="yes" ?>
<defaultValue
  default_char = "X"
  default_digit = "9"
  default_date = "11/11/1111 00:00:00"
  default_email = "abc@xyz.com"
  default_ip = "99.99.9.999"
  default_url = "http://www.xyz.com"
  default_phone = "999 999 999 9999"
  default_ssn = "999-99-9999"
  default_cc = "9999 9999 9999 9999"
  default_sin = "999-999-999"
  default_seed = "500"/>
```

TEMAS RELACIONADOS

- [“Configuración del Servicio de integración de datos” en la página 231](#)

Configuración de la transformación de enmascaramiento de datos

Lleve a cabo los pasos siguientes para configurar la transformación de enmascaramiento de datos.

1. Configure las opciones de ejecución para el servicio de integración de datos.
2. Cree la transformación.
3. Defina los puertos de entrada.
4. Configure las reglas de enmascaramiento de cada puerto que desee cambiar.
5. Obtenga una vista previa de los datos para comprobar los resultados.

Configuración del Servicio de integración de datos

Puede configurar opciones de ejecución para el Servicio de integración de datos en Informatica Administrator (Herramienta del administrador).

Configure las opciones de ejecución para establecer los siguientes directorios predeterminados:

- Directorio de inicio. Contiene el directorio de origen y el directorio de la memoria caché.
- Directorio de origen. Contiene los archivos de origen para los flujos de trabajo. Por ejemplo, el directorio de origen puede contener los archivos highgrp.txt y defaultvalue.xml.
- Directorio de la memoria caché. Contiene los archivos de memoria caché para la máscara de sustitución.

Para definir valores para las opciones de ejecución, abra la Herramienta del administrador y seleccione el Servicio de integración de datos en el **Navegador de dominio**. Haga clic en la vista **Propiedades** y, a continuación, haga clic en **Editar** en la sección **Opciones de ejecución**.

TEMAS RELACIONADOS

- [“Archivo de valores predeterminados” en la página 230](#)

Cómo crear una transformación de enmascaramiento de datos

Cree una transformación de enmascaramiento de datos en Developer Tool.

Antes de crear la transformación de enmascaramiento de datos, cree el origen. Importe un archivo sin formato o una tabla de base de datos relacional como objeto de datos físicos.

1. Seleccione un proyecto o una carpeta en la vista **Object Explorer**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de enmascaramiento de datos.
4. Haga clic en **Siguiente**.
5. Especifique un nombre para la transformación.
6. Haga clic en **Finalizar**.

La transformación aparece en el editor.

Cómo definir los puertos

Añada los puertos de entrada de enmascaramiento de datos en la vista **Resumen**. Cuando se crea un puerto de entrada, Developer Tool crea un puerto de salida correspondiente de forma predeterminada. El puerto de salida tiene el mismo nombre que el puerto de entrada.

1. En la vista **Resumen**, haga clic en **Nuevo** para añadir un puerto.
2. Configure el tipo de datos, la precisión y la escala para la columna.
Tiene que configurar el tipo de datos de la columna antes de definir las reglas de enmascaramiento para la columna.
3. Para configurar el enmascaramiento de datos para el puerto, haga clic en la flecha de la columna de tipo de enmascaramiento en la vista **Resumen**.

Cómo configurar un enmascaramiento de datos para cada puerto

Seleccione una técnica de enmascaramiento y las reglas de enmascaramiento correspondientes para un puerto en el cuadro de diálogo Enmascaramiento de datos. El cuadro de diálogo Enmascaramiento de datos se muestra al hacer clic en la columna de enmascaramiento de datos de la ficha **Puertos**.

1. Habilite **Aplicar enmascaramiento** para configurar el enmascaramiento para el puerto seleccionado.
Developer Tool muestra una lista de técnicas de enmascaramiento que puede utilizar en función de los tipos de datos del puerto que vaya a enmascarar.
2. Seleccione una técnica de enmascaramiento de la lista.
Developer Tool muestra diferentes reglas de enmascaramiento en función de la técnica de enmascaramiento que seleccione. Algunos formatos de máscara especiales no tienen reglas de enmascaramiento que se puedan configurar.
3. Configure las reglas de enmascaramiento.
4. Haga clic en **Aceptar** para aplicar la configuración de enmascaramiento de datos al puerto.
Cuando se define un enmascaramiento de datos para un puerto, Developer Tool crea un puerto de salida que se llama **out-<nombre del puerto>**. El **<nombre del puerto>** es el mismo que el del puerto de entrada. La transformación de enmascaramiento de datos devuelve los datos enmascarados en el puerto **out-<nombre del puerto>**.

Vista previa de los datos enmascarados

Puede comparar los datos enmascarados con los datos originales si visualiza los resultados de la transformación de enmascaramiento de datos en el **Visor de datos**.

1. Después de configurar los puertos de transformación de enmascaramiento de datos, cree una asignación que incluya el origen del objeto de datos físicos y la transformación de enmascaramiento de datos.
2. Conecte el origen a la transformación de enmascaramiento de datos.
3. Compruebe que el origen tiene datos en una ubicación compartida a la que pueda acceder el servicio de integración de datos.
4. Haga clic en la transformación de enmascaramiento de datos para seleccionarla en la asignación.
5. Haga clic en **Visor de datos** y en **Ejecutar**.

Developer Tool muestra datos de todos los puertos de salida de transformación de enmascaramiento de datos. Los puertos que tienen el prefijo de **salida** contienen datos enmascarados. Puede comparar los datos enmascarados con los datos originales en la vista **Visor de datos**.

Propiedades de tiempo de ejecución de la transformación de enmascaramiento de datos

Puede configurar las propiedades de tiempo de ejecución de la transformación de enmascaramiento de datos para aumentar el rendimiento.

Configure las siguientes propiedades de tiempo de ejecución:

Nombre de conexión de almacenamiento

Obligatorio para los tipos de enmascaramiento que utilizan una conexión de almacenamiento. El nombre de la conexión que usa como conexión de almacenamiento.

Tamaño de la memoria caché

El tamaño de la memoria caché del diccionario en la memoria principal. Aumente el tamaño de memoria para mejorar el rendimiento. El tamaño mínimo recomendado es 32 MB para 100.000 registros. El valor predeterminado es 8 MB.

Directorio de la memoria caché

La ubicación de la memoria caché del diccionario. Debe tener permisos de escritura para el directorio. El valor predeterminado es CacheDir.

Frase de contraseña

Obligatoria para el enmascaramiento de cifrado. La frase de contraseña genera una clave con la que los datos se cifran o descifran. En las asignaciones que crea en Test Data Management, la frase de contraseña está cifrada. Para las asignaciones que crea en Developer tool, puede usar la clave del sitio para cifrar la frase de contraseña e introducir el valor cifrado en el cuadro de texto.

Modo

Obligatoria para el enmascaramiento de cifrado. Determina si la transformación de enmascaramiento de datos realiza el cifrado o descifrado de los datos. Establezca el valor en Cifrado para cifrar los datos de origen. Para descifrar los datos enmascarados y devolver los datos de origen originales, ejecute una asignación con la misma configuración de técnica de cifrado y frase de contraseña, con el modo establecido en Descifrado.

Usar SoftHSM

Obligatorio para el cifrado. Elija si desea utilizar SoftHSM durante el cifrado. SoftHSM es más seguro, pero podría advertir diferencias en el rendimiento.

Tabla de almacenamiento compartido

Habilita la compartición de la tabla de almacenamiento entre las instancias de transformación de enmascaramiento de datos. Habilite la tabla de almacenamiento compartido cuando dos instancias de transformación de enmascaramiento de datos usen la misma columna del diccionario para la conexión de base de datos, valor de semilla y configuración regional. También puede habilitar la tabla de almacenamiento compartido cuando dos puertos en la misma transformación de enmascaramiento de datos usen la misma columna del diccionario para la conexión, la semilla y la configuración regional. Deshabilite la tabla de almacenamiento compartido cuando los puertos o las transformaciones de enmascaramiento de datos no compartan la columna del diccionario. El valor predeterminado es deshabilitada.

Intervalo de confirmación de la tabla de almacenamiento

El número de filas para confirmar al mismo tiempo para la tabla de almacenamiento. Aumente el valor para aumentar el rendimiento. Configure el intervalo de confirmación si no configura la tabla de almacenamiento compartido. El valor predeterminado es 100.000.

Almacenamiento cifrado

Cifra las tablas de almacenamiento, tales como `IDM_SUBSTITUTION_STORAGE` y `IDM_EXPRESSION_STORAGE`. Compruebe que dispone de datos cifrados en las tablas de almacenamiento antes de habilitar la propiedad de almacenamiento cifrado. Borre esta opción si no desea cifrar las tablas de almacenamiento. El valor predeterminado es deshabilitado.

Clave de cifrado de almacenamiento

La transformación de enmascaramiento de datos cifra el almacenamiento en función de la clave de cifrado de almacenamiento. Utilice la misma clave de cifrado para cada ejecución de sesión de la misma instancia de transformación de enmascaramiento de datos.

Nombre del propietario del diccionario de sustitución

Nombre del propietario de la tabla del diccionario de sustitución cuando se selecciona el tipo de máscara de sustitución. Si el usuario de la base de datos especificado en la conexión de la base de datos no es el propietario de la tabla del diccionario de sustitución en una sesión, debe especificar el propietario de la tabla.

Nombre del propietario del almacenamiento

Nombre del propietario de tabla para `IDM_SUBSTITUTION_STORAGE` o `IDM_EXPRESSION_STORAGE` cuando selecciona el tipo de máscara de sustitución repetible única o de expresión repetible.

Ejemplo de enmascaramiento de datos

El desarrollador tiene que crear datos de prueba para las aplicaciones del cliente. Los datos deben contener datos del cliente que sean realistas y a los que otros desarrolladores puedan acceder en el entorno de desarrollo de la compañía.

El desarrollador crea un servicio de datos que devuelve datos de cliente enmascarados como el ID, el número de tarjeta de crédito y los ingresos del cliente. La asignación incluye una transformación de enmascaramiento de datos que transforma los datos del cliente.

La siguiente figura muestra la asignación:



La asignación tiene las siguientes transformaciones:

- Read_Customer_Data. Contiene la información de la tarjeta de crédito y los ingresos del cliente.
- Transformación de Customer_Data_Masking. Enmascara todas las columnas excepto FirstName y LastName. La transformación de enmascaramiento de datos pasa las columnas enmascaradas al destino.
- Customer_TestData. La transformación de salida que recibe los datos enmascarados del cliente.

Read_Customer Data

Los datos de cliente contienen las siguientes columnas:

Column	Tipo de datos
CustomerID	Entero
LastName	Cadena
FirstName	Cadena
CreditCard	Cadena
Income	Entero
Join_Date	Fecha y hora (DD/MM/AAAA)

En la siguiente tabla aparecen datos de cliente de ejemplo.

CustomerID	LastName	FirstName	CreditCard	Income	JoinDate
0095	Bergeron	Barbara	4539-1686-3069-3957	12000	31/12/1999
0102	Brosseau	Derrick	5545-4091-5232-8948	4000	03/03/2011
0105	Anderson	Lauren	1234-5678-9012-3456	5000	04/03/2009
0106	Boonstra	Pauline	4217-9981-5613-6588	2000	07/07/2007
0107	Chan	Brian	4533-3156-8865-3156	4500	06/18/1995

Transformación de enmascaramiento de datos de clientes

La transformación de enmascaramiento de datos enmascara todas las columnas de la fila del cliente excepto el nombre y el apellido.

La transformación de enmascaramiento de datos realiza los siguientes tipos de enmascaramiento:

- Enmascaramiento de claves
- Enmascaramiento aleatorio
- Enmascaramiento de tarjeta de crédito

En la tabla siguiente se muestran las reglas de enmascaramiento para cada puerto en la transformación de enmascaramiento de datos:

Puerto de entrada	Tipo de enmascaramiento	Reglas de enmascaramiento	Descripción
CustomerID	De claves	La semilla es 934. El ID de cliente no tiene ningún formato de máscara. Los caracteres de reemplazo de la cadena de resultado son 1234567890.	La máscara CustomerID es determinista. El ID del cliente enmascarado contiene números.
LastName	Sin máscara	-	-
FirstName	Sin máscara	-	-
CreditCard	CreditCard	-	La transformación de enmascaramiento de datos enmascara el número de la tarjeta de crédito con otro número que tiene una suma de comprobación válida.
Income	Aleatoria	Desenfoco Porcentaje Límite inferior = 1 Límite superior = 10	El ingreso enmascarado se encuentra dentro del diez por ciento del ingreso de origen.
JoinDate	Aleatoria	Desenfoco Unidad = Año Límite inferior = 5 Límite superior = 5	La fecha enmascarada se encuentra en el periodo de 5 años posterior a la fecha original.

Resultados de datos de prueba de cliente

La transformación Customer_TestData recibe datos de clientes realistas provenientes de la transformación de enmascaramiento de datos.

El destino Customer_TestData recibe los siguientes datos:

out-CustomerID	out-LastName	outFirstName	out-CreditCard	out-Income	out-JoinDate
3954	Bergeron	Barbara	4539-1625-5074-4106	11500	03/22/2001
3962	Brosseau	Derrick	5545-4042-8767-5974	4300	04/17/2007
3964	Anderson	Lauren	1234-5687-2487-9053	5433	09/13/2006
3965	Boonstra	Pauline	4217-9935-7437-4879	1820	02/03/2010
3966	Chan	Brian	4533-3143-4061-8001	4811	10/30/2000

El ingreso está dentro del diez por ciento del ingreso original. La fecha de unión se encuentra en el periodo de 5 años posterior a la fecha original.

Propiedades avanzadas de Transformación de enmascaramiento de datos

Configure propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de enmascaramiento de datos.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de enmascaramiento de datos en un entorno no nativo

El procesamiento de transformaciones de enmascaramiento de datos en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. No compatible.

Transformación de enmascaramiento de datos en el motor de Blaze

La transformación de enmascaramiento de datos es compatible con las siguientes restricciones:

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación está configurada para el enmascaramiento de expresiones repetible.
- La transformación está configurada para un enmascaramiento de sustitución repetible único.

Se pueden usar las siguientes técnicas de enmascaramiento en este motor:

Tarjeta de crédito
Correo electrónico
Expresión
Dirección IP
Clave
Teléfono
Aleatorio
SIN
SSN

Tokenización
URL
Sustitución aleatoria
Sustitución repetible
Dependiente con sustitución aleatoria
Dependiente con sustitución repetible

Transformación de enmascaramiento de datos en el motor de Spark

La transformación de enmascaramiento de datos es compatible con las siguientes restricciones:

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación está configurada para el enmascaramiento de expresiones repetible.
- La transformación está configurada para un enmascaramiento de sustitución repetible único.

Se pueden usar las siguientes técnicas de enmascaramiento en este motor:

Tarjeta de crédito
Correo electrónico
Expresión
Dirección IP
Clave
Teléfono
Aleatorio
SIN
SSN
Tokenización
URL
Sustitución aleatoria
Sustitución repetible
Dependiente con sustitución aleatoria
Dependiente con sustitución repetible

Para optimizar el rendimiento de la transformación de enmascaramiento de datos, configure las siguientes propiedades de configuración del motor de Spark en la conexión de Hadoop:

spark.executor.cores

Indica el número de núcleos que utiliza cada proceso de ejecutor para ejecutar los tasklets en el motor de Spark.

Establezca el valor en `spark.executor.cores=1`

spark.executor.instances

Indica el número de instancias que utiliza cada proceso de ejecutor para ejecutar los tasklets en el motor de Spark.

Establezca el valor en: `spark.executor.cores=1`

spark.executor.memory

Indica la cantidad de memoria que utiliza cada proceso de ejecutor para ejecutar los tasklets en el motor de Spark.

Establezca el valor en: `spark.executor.memory=3G`

Transformación de enmascaramiento de datos en una asignación de transmisión

Las asignaciones de transmisión tienen las mismas reglas de procesamiento que las asignaciones por lotes en el motor de Spark.

CAPÍTULO 13

Transformación del Procesador de datos

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación del Procesador de datos, 240](#)
- [Vistas de la transformación del Procesador de datos, 241](#)
- [Puertos de la transformación del Procesador de datos, 242](#)
- [Componente de inicio, 244](#)
- [Referencias, 245](#)
- [Configuración de la transformación del Procesador de datos, 246](#)
- [Eventos, 253](#)
- [Registros, 254](#)
- [Desarrollo de la transformación del Procesador de datos, 257](#)
- [Exportación e importación de transformaciones del Procesador de datos, 261](#)
- [Transformación del procesador de datos Validación, 264](#)
- [Transformación del procesador de datos en un entorno no nativo, 265](#)

Resumen de la transformación del Procesador de datos

La transformación del Procesador de datos procesa formatos de archivo sin estructura y semiestructurados en una asignación. Configure la transformación para procesar formatos de mensajería, páginas HTML, XML, JSON y documentos PDF. También puede convertir formatos estructurados como ACORD, HIPAA, HL7, EDI-X12, EDIFACT y SWIFT.

Una asignación utiliza una transformación del Procesador de datos para cambiar documentos de un formato a otro. La transformación del Procesador de datos procesa archivos de cualquier formato en una asignación. Al crear una transformación del Procesador de datos, puede definir componentes que convierten los datos.

Una transformación del Procesador de datos puede contener varios componentes para procesar los datos. Cada componente puede contener otros componentes.

Por ejemplo, puede recibir facturas del cliente en archivos de Microsoft Word. Puede configurar una transformación del Procesador de datos para analizar los datos de cada archivo de Word. Extraiga los datos de clientes a una tabla Clientes. Extraiga información de pedidos a una tabla Pedidos.

Al crear una transformación del Procesador de datos, define un XMap, script o biblioteca. Un XMap convierte un archivo jerárquico de entrada en un archivo jerárquico de salida de otra estructura. Una biblioteca convierte un tipo de mensajería de industria en un documento XML con una estructura de jerarquías, o un formato XML en un formato estándar de la industria. Un script puede analizar documentos de origen a formato jerárquico, convertir el formato jerárquico en otros formatos de archivo o asignar un documento jerárquico a otro formato jerárquico.

Defina los scripts en el editor de IntelliScript de transformación del Procesador de datos. Puede definir los siguientes tipos de scripts:

- **Analizador.** Convierte documentos de origen a XML. La salida de un analizador siempre es XML. La entrada puede tener cualquier formato, como texto, HTML, Word, PDF o HL7.
- **Serializador.** Convierte un archivo XML en un documento de salida de cualquier formato. La salida de un serializador puede tener cualquier formato, como documento de texto, documento HTML o PDF.
- **Asignador.** Convierte un documento de origen XML en otra estructura o esquema XML. Puede convertir los mismos documentos XML como en un XMap.
- **Transformador.** Modifica los datos a cualquier formato. Añade, quita, convierte o modifica texto. Utilice Transformadores con un analizador, un Asignador o un Serializador. También puede ejecutar un transformador como componente autónomo.
- **Transmisión por secuencias.** Divide documentos de entrada de gran tamaño, como secuencias de datos de varios gigabytes, en segmentos. La transmisión por secuencias procesa los documentos que contienen varios mensajes o registros, como archivos HIPAA o EDI.

Vistas de la transformación del Procesador de datos

La transformación del Procesador de datos tiene varias vistas a las que puede acceder al configurar la transformación y ejecutarla en la Herramienta del desarrollador.

Algunas de las vistas de la transformación del Procesador de datos no aparecen en la Herramienta del desarrollador de forma predeterminada. Para cambiar las vistas para la transformación, haga clic en **Ventana > Mostrar vista > Otro > Informatica**. Seleccione las vistas que desee ver.

La transformación del Procesador de datos tiene las siguientes vistas fijas:

Vista Resumen

Configure los puertos y defina el componente de inicio.

Vista de referencias

Añada o elimine esquemas desde la transformación.

Vista de configuración

Configure la configuración de transformación para la codificación, el control de salida y la generación de XML.

Vista Objetos

Añada, cambie o elimine objetos de script, biblioteca o XMap desde la transformación.

También puede acceder a las siguientes vistas para la transformación del Procesador de datos:

Vista del Procesador de datos de origen hexadecimal

Muestra un documento de entrada en formato hexadecimal.

Vista Eventos del procesador de datos

Muestra información sobre los eventos que ocurren cuando ejecuta la transformación en la Herramienta del desarrollador. Muestra la inicialización, la ejecución y los eventos de resumen.

Vista de ayuda de script

Muestra la ayuda dependiente del contexto para el editor de scripts.

Vista del Visor de datos

Vea los datos de entrada de ejemplo, ejecute la transformación y vea resultados de salida.

Puertos de la transformación del Procesador de datos

Defina los puertos de transformación del Procesador de datos en la vista **Resumen** de la transformación.

Una transformación del Procesador de datos puede leer entradas desde un archivo, un búfer o un búfer de transmisión desde un lector de archivos complejos. Puede utilizar un lector de archivos sin formato como un búfer para leer todo un archivo de una vez. También puede leer un archivo de entrada desde una base de datos.

Los puertos de salida que puede crear dependen de si desea devolver una cadena, archivos complejos o filas de datos relacionales desde la transformación.

Puertos de entrada de la transformación del Procesador de datos

Cuando se crea una transformación del Procesador de datos, la Herramienta del desarrollador crea un puerto de entrada predeterminado. Cuando se define un puerto de entrada adicional en un componente de inicio de script, la Herramienta del desarrollador crea un puerto de entrada adicional en la transformación.

El tipo de entrada determina el tipo de datos que el servicio de integración de datos pasa a la transformación del Procesador de datos. El tipo de entrada determina si la entrada está formada por datos o si es la ruta de acceso de un archivo de origen.

Configure uno de los siguientes tipos de entrada:

Búfer

La transformación del Procesador de datos recibe filas de datos de origen en el puerto Entrada. Utilice el tipo de entrada de búfer cuando configure la transformación para recibir datos desde un archivo sin formato o desde un transformaton de Informatica.

Archivo

La transformación del Procesador de datos recibe la ruta de acceso al archivo de origen en el puerto Entrada. El componente de inicio del Procesador de datos abre el archivo de origen. Utilice el tipo de entrada de archivo para analizar archivos binarios como los de Microsoft Excel o Microsoft Word.

También puede utilizar el tipo de entrada de archivo para archivos de gran tamaño que puede requerir un lote de memoria del sistema para el proceso con un puerto de entrada de búfer.

Parámetro de servicio

La transformación del Procesador de datos recibe valores para aplicarlos a variables en los puertos de parámetros de servicio. Cuando se seleccionan las variables para recibir datos de entrada, la Herramienta del desarrollador crea un puerto de parámetros de servicio para cada variable.

Output_Filename

Cuando se configura el puerto de salida predeterminado para devolver un nombre de archivo en lugar de datos sin formato, la Herramienta del desarrollador crea un puerto Output_Filename. Puede pasar un nombre de archivo al puerto Output_Filename desde una asignación.

Cuando se define un puerto de entrada, puede definir la ubicación del archivo de entrada de ejemplo para el puerto. Un archivo de entrada de ejemplo es un pequeño ejemplo del archivo de entrada. Haga referencia a un archivo de entrada de ejemplo al crear scripts. También puede utilizar el archivo de entrada de ejemplo cuando pruebe la transformación en la vista **Visor de datos**. Defina el archivo de entrada de ejemplo en el campo **Ubicación de entrada**.

Puertos de parámetros de servicio

Puede crear puertos de entrada que reciben valores de variables. Las variables pueden contener cualquier tipo de datos como una cadena, una fecha o un número. Una variable también puede contener una ubicación de un documento de origen. Se puede hacer referencia a las variables de un componente del Procesador de datos.

Cuando se crea un puerto de entrada para una variable, la Herramienta del desarrollador muestra una lista de variables entre las que puede elegir.

Crear puertos de parámetros de servicio

Puede crear puertos de entrada que reciben valores de variables. También puede quitar los puertos que se crean a partir de variables.

1. Abra la vista **Resumen** de la transformación del Procesador de datos.
2. Haga clic en **Elegir**.
La Herramienta del desarrollador muestra una lista de las variables y indica que ya tiene puertos variables.
3. Seleccione una o más variables.
La Herramienta del desarrollador crea un puerto de entrada de búfer para cada variable que seleccione. No se puede modificar el puerto.
4. Para eliminar un puerto que se crea a partir de una variable, deshabilite la selección de la lista de variable. Cuando deshabilite la selección, la Herramienta del desarrollador elimina el puerto de entrada.

Puertos de salida de la transformación del Procesador de datos

La transformación del Procesador de datos tiene un puerto de salida de forma predeterminada. Si se definen los puertos de salida adicionales en un script, la Herramienta del desarrollador añade los puertos a la transformación del Procesador de datos. Puede crear grupos de puertos si configura la transformación para devolver datos relacionales. También puede crear puertos de parámetros de servicio y puertos de transferencia.

Puerto de salida predeterminado

La transformación del Procesador de datos tiene un puerto de salida de forma predeterminada. Cuando crea una salida relacional, puede definir grupos de puertos de salida relacionados en lugar del puerto de salida predeterminado. Cuando se define un puerto de salida adicional en un componente de script, la Herramienta del desarrollador añade un puerto de salida adicional para la transformación.

Configure uno de los siguientes tipos de salida para un puerto de salida predeterminado:

Búfer

La transformación del Procesador de datos devuelve XML a través del puerto Salida. Elija el tipo de archivo de búfer cuando analice documentos o cuando asigne XML a otros documentos XML en la transformación del Procesador de datos.

Archivo

El servicio de integración de datos devuelve un nombre de archivo de salida en el puerto de salida para cada instancia de origen o de fila. El componente de transformación del Procesador de datos escribe el archivo de salida en lugar de devolver datos mediante los puertos de salida de transformación del Procesador de datos.

Cuando seleccione un puerto de salida Archivo, la Herramienta del desarrollador crea un puerto de entrada Output_Filename. Puede pasar un nombre de archivo al puerto de nombre de archivo Salida. La transformación del Procesador de datos crea el archivo de salida con un nombre que recibe en este puerto.

Si el nombre del archivo de salida está vacío, el servicio de integración de datos devuelve un error de fila. Cuando se produce un error, el servicio de integración de datos escribe un valor nulo en el puerto de salida y devuelve un error de fila.

Elija el tipo de salida Archivo al transformar XML a un archivo de datos binarios como un archivo PDF o un archivo de Microsoft Excel.

Puertos de transferencia

Puede configurar puertos de transferencia para cualquier transformación del Procesador de datos. Los puertos de transferencia son puertos de entrada y salida que reciben datos de entrada y devuelven los mismos datos a una asignación sin modificarlos.

Puede configurar puertos de transferencia en una instancia de transformación del Procesador de datos que esté en una asignación.

Para añadir un puerto de transferencia, arrastre un puerto desde otra transformación en la asignación. También puede añadir puertos en la ficha **Puertos** de la vista **Propiedades**. Haga clic en **Nuevo** para añadir un puerto de transferencia.

Nota: Cuando añada puertos de transferencia a una transformación del Procesador de datos con entrada relacional y salida jerárquica, añada los puertos al grupo raíz de la estructura relacional.

Las transformaciones del Procesador de datos pueden incluir puertos de transferencia con tipos de datos personalizados.

Componente de inicio

Un componente de inicio define el componente que inicia el proceso en la transformación del Procesador de datos. Configure el componente de inicio en la vista **Resumen**.

Una transformación del Procesador de datos puede contener varios componentes para procesar los datos. Cada componente puede contener otros componentes. Se debe identificar qué componente es el punto de entrada para la transformación.

Cuando configure el componente de inicio en una transformación del Procesador de datos, puede elegir un componente XMap, de biblioteca o de script como el componente de inicio. En términos de scripts, puede seleccionar uno de los siguientes tipos de componentes:

- Analizador. Convierte documentos de origen a XML. La entrada puede tener cualquier formato, como texto, HTML, Word, PDF o HL7.
- Asignador. Convierte un documento de origen XML en otra estructura o esquema XML.
- Serializador. Convierte un archivo XML en un documento de salida de cualquier formato.
- Transmisión por secuencias. Divide documentos de entrada de gran tamaño, como secuencias de datos de varios gigabytes, en segmentos.
- Transformador. Modifica los datos a cualquier formato. Añade, quita, convierte o modifica texto. Utilice Transformadores con un analizador, un Asignador o un Serializador. También puede ejecutar un transformador como componente autónomo.

Nota: Si el componente de inicio no es un XMap o una biblioteca, también puede configurar el componente de inicio en un script en lugar de en la vista **Resumen**.

Referencias

Puede definir las referencias de la transformación, como las referencias a esquemas o mapplets, seleccionando un esquema o un mapplet para que funcionen como referencia. Algunas transformaciones del Procesador de datos requieren un esquema jerárquico para definir la jerarquía de las entradas o salidas de los componentes relevantes en la transformación. Para utilizar el esquema en la transformación, debe definir una referencia al esquema para la transformación. También puede utilizar una acción especializada llamada RunMapplet para invocar un mapplet desde una transformación del Procesador de datos. Para llamar a un mapplet, primero debe definir una referencia al mapplet para la transformación.

Puede definir las referencias de la transformación, como las referencias a esquemas o mapplets, en la vista **Referencias** de la transformación.

Referencias a esquemas

La transformación del Procesador de datos hace referencia a objetos de esquema en el repositorio de modelos. Los objetos de esquema pueden existir en el repositorio antes de crear la transformación. También puede importar esquemas desde la vista **Referencias** de la transformación.

La codificación de esquemas debe coincidir con la codificación de entradas para los objetos Serializador o Asignador. La codificación de esquemas debe coincidir con la codificación de salidas para los objetos Analizador o Asignador. Configure la codificación de trabajo en la vista de transformación **Configuración**.

Un esquema puede hacer referencia a esquemas adicionales. La Herramienta del desarrollador muestra el espacio de nombres y el prefijo para cada esquema al que hace referencia la transformación del Procesador de datos. Cuando se hace referencia a varios esquemas con espacios de nombres vacíos, la transformación no es válida.

Referencias a mapplets

Puede llamar a un mapplet desde una transformación del Procesador de datos con la acción RunMapplet. Antes de añadir la acción RunMapplet a un componente de transformación del Procesador de datos, primero debe definir una referencia al mapplet que desee llamar.

Configuración de la transformación del Procesador de datos

Configure las páginas de códigos, las opciones de procesamiento de XML y la configuración de registro en la vista **Configuración** de la transformación del procesador de datos.

Codificación de caracteres

Una codificación de caracteres es una asignación de los caracteres de un idioma o grupo de idiomas con el código hexadecimal.

Cuando se diseña un script, debe definir la codificación de documentos de entrada y la codificación de documentos de salida. Defina la codificación de trabajo para definir cómo el editor de IntelliScript muestra caracteres y cómo la transformación del Procesador de datos procesa los caracteres.

Codificación de trabajo

La codificación de trabajo es la página de códigos para los datos en la memoria y la página de códigos para los datos que aparecen en la interfaz de usuario y los archivos de trabajo. Debe seleccionar una codificación de trabajo que sea compatible con la codificación de los esquemas a los que se hace referencia en la transformación del Procesador de datos.

La siguiente tabla muestra la configuración de codificación del trabajo:

Valor	Descripción
Utilizar la página de códigos predeterminada del Procesador de datos	Utiliza la codificación predeterminada de la transformación del Procesador de datos.
Otras	Seleccione la codificación de la lista.
Codificación de caracteres especiales XML	<p>Determina la representación de caracteres especiales XML. Puede seleccionar Ninguno o XML.</p> <ul style="list-style-type: none">- Ninguno. Dejar como <code>&amp;</code>, <code>&lt;</code>, <code>&gt;</code>, <code>&quot;</code>, <code>&apos;</code>; Las referencias de entidad para caracteres especiales XML se interpretan como texto. Por ejemplo, el carácter <code>></code> aparece como <code>&gt;</code>; El valor predeterminado es Ninguno.- XML. Convertir a <code>&lt;</code> y <code>&gt;</code> ' Las referencias de entidad para caracteres especiales XML se interpretan como caracteres normales. Por ejemplo, <code>&gt;</code> aparece como el siguiente carácter: <code>></code>

Codificación de entrada

La codificación de entrada determina cómo los datos de caracteres se codifican en documentos de entrada. Puede configurar la codificación para puertos de entrada adicionales en un script.

La siguiente tabla describe la configuración de codificación en el área **Entrada**:

Valor	Descripción
Utilizar la codificación especificada en documento de entrada	Utilice la página de códigos que el documento de origen define, tal como el atributo de codificación de un documento XML. Si el documento de origen no tiene una especificación de codificación, la transformación del Procesador de datos utiliza la configuración de la codificación de la vista Configuración .
Utilizar codificación de trabajo	Utilice la misma codificación que la codificación de trabajo.
Otras	Seleccione la codificación de entrada de una lista desplegable.
Codificación de caracteres especiales XML	Determina la representación de caracteres especiales XML. Puede seleccionar Ninguno o XML . - Ninguno. Dejar como & < > " ' Las referencias de entidad para los caracteres especiales XML se interpretan como texto, por ejemplo, el carácter > aparece como > El valor predeterminado es Ninguno. - XML. Convertir a & < > " ' Las referencias de entidad para caracteres especiales XML se interpretan como caracteres normales. Por ejemplo, > aparece como el siguiente carácter: >
Orden de bytes	Describe cómo aparecen los caracteres multibyte en el documento de entrada. Puede seleccionar las opciones siguientes: - Little-endian. El byte menos significativo aparece primero. Predeterminado. - Big-endian. El byte más significativo aparece primero. - No hay conversión binaria.

Codificación de salida

La codificación de salida determina cómo los datos de caracteres se codifican en el documento de salida principal..

La siguiente tabla describe la configuración de codificación en el área **Salida**:

Valor	Descripción
Utilizar codificación de trabajo	La codificación de salida es la misma que la codificación de trabajo.
Otras	El usuario selecciona la codificación de salida de la lista.

Valor	Descripción
Codificación de caracteres especiales XML	<p>Determina la representación de caracteres especiales XML. Puede seleccionar Ninguno o XML.</p> <ul style="list-style-type: none"> - Ninguno. Dejar como <code>&lt;</code> <code>&gt;</code> <code>&quot;</code> <code>&apos;</code> Las referencias de entidad para los caracteres especiales XML se interpretan como texto, por ejemplo, el carácter <code>></code> aparece como <code>&gt;</code> Predeterminado. - XML. Convertir a <code><</code> <code>></code> <code>"</code> <code>'</code> Las referencias de entidad para caracteres especiales XML se interpretan como caracteres normales. Por ejemplo, <code>&gt;</code> aparece como el siguiente carácter: <code>></code>
Igual que la codificación de entrada	La codificación de salida es la misma que la codificación de entrada.
Orden de bytes	<p>Describe cómo aparecen los caracteres multibyte en el documento de entrada. Puede seleccionar las opciones siguientes:</p> <ul style="list-style-type: none"> - Little-endian. El byte menos significativo aparece primero. Predeterminado. - Big-endian. El byte más significativo aparece primero. - No hay conversión binaria.

Reglas y directrices de codificación de caracteres

Utilice las siguientes reglas y directrices cuando configure codificaciones:

- Para aumentar el rendimiento, establezca la misma codificación que utilice el documento de salida.
- Establezca la codificación de entrada que utilice el documento de entrada.
- Establezca la codificación de salida que utilice el documento de salida.
- Para los idiomas que utilicen caracteres de varios bytes, establezca la codificación en UTF-8. Para la codificación de entrada y salida, puede utilizar una codificación Unicode como UTF-8 o una página de codificación de bytes dobles, como Big5 o Shift_JIS.

Configuración de salida

Configurar la configuración de control de salida para controlar si la transformación del Procesador de datos crea registros de eventos y guarda documentos de salida.

Puede controlar los tipos de mensajes que la transformación Procesador de datos escribe en el registro de eventos de tiempo de diseño. Si desea guardar los documentos de entrada analizados con los registros de eventos, puede ver el contexto donde se produjo el error en la vista **Event**.

La siguiente tabla describe la configuración en el área **Eventos de tiempo de diseño**:

Valor	Descripción
Registrar eventos en tiempo de diseño	Determina si desea crear un registro de eventos de tiempo de diseño. De forma predeterminada, la transformación del Procesador de datos registra notificaciones, advertencias y fallos en el registro de eventos de tiempo de diseño. Se pueden excluir los siguientes tipos de eventos: <ul style="list-style-type: none"> - Notificaciones - Avisos - Fallos
Guardar documentos analizados	Determina si la transformación del Procesador de datos guarda un documento de entrada analizado. Puede seleccionar las opciones siguientes: <ul style="list-style-type: none"> - always. - Nunca - En fallo El valor predeterminado es always.

La siguiente tabla describe la configuración en el área **Eventos de tiempo de ejecución**:

Valor	Descripción
Eventos de registro en tiempo de ejecución	Determina si se ha creado un evento de registro cuando se ejecuta la transformación de una asignación. <ul style="list-style-type: none"> - Never. - En fallo El valor predeterminado es Never.

La siguiente tabla describe los ajustes del área **Salida**:

Valor	Descripción
Deshabilitar salida automática	Determina si la transformación Procesador de datos escribe la salida en el archivo de salida estándar. Deshabilite la salida estándar en las siguientes situaciones: <ul style="list-style-type: none"> - Si pasa la salida de un analizador a la entrada de otro componente antes de que la transformación cree un archivo de salida. - Si utiliza una acción WriteValue para escribir los datos directamente en la salida desde un script en lugar de pasarlos a través de los puertos de salida.
Deshabilitar compresión de valor	Determina si la transformación Procesador de datos utiliza la compresión del valor para optimizar el uso de memoria. <p>Importante: no deshabilite la compresión del valor, a menos que el servicio internacional de atención al cliente de Informatica le indique que puede hacerlo.</p>

En la siguiente tabla se describen las opciones del área **Modo de colección del puerto de salida binario**. Puede seleccionar cualquiera de estas opciones de salida binaria de una transformación de jerárquico a

relacional con una salida XML, de Avro o de Parquet, o bien de un analizador de transformación de Procesador de datos con una salida de Avro o de Parquet.

Valor	Descripción
Recopilar las filas de entrada en una sola salida	Establece si la transformación de Procesador de datos acumula la entrada relacional en un solo puerto de salida binario.
Dividir salida cuando el tamaño supere	Establece si la transformación de Procesador de datos divide la salida en fragmentos según un tamaño de salida máximo definido.
Fila de salida para cada fila (no recopilar)	Establece si la transformación de Procesador de datos pasa la salida en filas separadas.

Configuración de procesamiento

La configuración de procesamiento define cómo procesa la transformación Procesador de datos un elemento sin un tipo de datos definido. La configuración afecta a los scripts. La configuración no afecta a los elementos que procesa un XMap.

La siguiente tabla describe la configuración de procesamiento que afecta al procesamiento XML en los scripts:

Valor	Descripción
Tratar como xs:string	La transformación del Procesador de datos trata un elemento sin tipo como una cadena. En el cuadro de diálogo Elegir XPath , el elemento o atributo aparece como un único nodo.
Tratar como xs:anyType	La transformación Procesador de datos trata un elemento sin tipo como anyType. En el cuadro de diálogo Elegir XPath , el elemento o atributo aparece como un árbol de nodos. Un nodo es de tipo xs:string y todos los tipos de datos complejos con nombre aparecen como nodos de árbol.

La siguiente tabla describe una configuración de procesamiento que influye en el procesamiento de transmisión por secuencias:

Valor	Descripción
Tamaño de fragmento de transmisión por secuencias	Esta configuración define la cantidad de datos que la transmisión por secuencias lee cada vez desde una secuencia de archivos de entrada. La transformación del Procesador de datos aplica esta configuración a una transmisión por secuencias con un archivo de entrada.

La siguiente tabla describe una configuración de procesamiento que afecta al procesamiento de transformación de jerárquico a relacional:

Valor	Descripción
Aplicar validación estricta	Esta opción establece si la transformación del Procesador de datos efectúa una validación estricta de la entrada jerárquica. Cuando se aplica una validación estricta, el archivo de entrada jerárquica debe ajustarse estrictamente a su esquema. Esta opción se puede aplicar cuando el modo de Procesador de datos está configurado como Asignación de salida , lo que crea puertos de salida para la salida relacional. Esta opción no se aplica a las asignaciones con entrada JSON de las versiones anteriores a 10.2.1.
Normalizar entrada de XML	Esta configuración determina si la transformación del Procesador de datos normaliza la entrada XML. De forma predeterminada, la transformación lleva a cabo la normalización para la entrada XML. En algunos casos, puede optar por omitir la normalización automática para aumentar el rendimiento.

Configuración de XMap

La configuración de XMap define cómo procesa la transformación del Procesador de datos los elementos de entrada de XMap que no se transforman en elementos de salida. Los elementos no leídos se transfieren a un puerto dedicado llamado **XMap_Unread_Input_Values**. La configuración se aplica solo cuando se selecciona el XMap como componente de inicio. La configuración no afecta a los elementos que el XMap procesa.

Para transferir elementos de XMap no leídos a un puerto dedicado, habilite el ajuste **Escriba los elementos no leídos en un puerto de salida adicional**.

Configuración de salida XML

La configuración de generación XML define las características de los documentos XML de salida.

La siguiente tabla describe la configuración de generación XML en el área de **Título de esquema**:

Valor	Descripción
Ubicación del esquema	Define el schemaLocation para el elemento raíz del documento de salida principal.
No hay ubicación de esquema de espacio de nombres	Define el atributo xsi:noNamespaceSchemaLocation del elemento raíz del documento de salida principal.

Configure los ajustes del modo de salida XML para determinar cómo la transformación del Procesador de datos maneja los elementos o los atributos que faltan en el documento XML de entrada. La siguiente tabla describe la configuración de generación XML en el área de **Modo de salida XML**:

Valor	Descripción
Tal cual	No se añaden ni se quitan elementos vacíos. Esta propiedad está habilitada de manera predeterminada.
Completo	Todos los elementos necesarios y opcionales definidos en el esquema de salida se escriben en la salida. Los elementos sin contenido se escriben como elementos vacíos.
Compacto	Quita elementos vacíos de la salida. Si Añadir para elementos está habilitado, la transformación del Procesador de datos quita solamente los elementos opcionales. Si Añadir para elementos está deshabilitado, la transformación del Procesador de datos quita todos los elementos vacíos. Puede que la salida XML no sea válida.

La siguiente tabla describe la configuración de generación XML en el área de **Valores predeterminados para nodos requeridos**:

Valor	Descripción
Añadir para elementos	Cuando el esquema de salida define un valor predeterminado para un elemento necesario, la salida incluye el elemento con un valor predeterminado. Esta propiedad está habilitada de manera predeterminada.
Añadir para atributos	Cuando el esquema de salida define un valor predeterminado para un atributo necesario, la salida incluye el atributo con su valor predeterminado. Esta propiedad está habilitada de manera predeterminada.
Validar valores agregados	Determina si la transformación del Procesador de datos valida los elementos vacíos que se han añadido por la salida del modo Completo. El valor predeterminado es deshabilitada. Si Validar valores agregados está habilitado y el esquema no admite elementos vacíos, la salida XML podría no ser válida.

La siguiente tabla describe la configuración de generación XML en el área de **Instrucciones de proceso**:

Valor	Descripción
Añadir instrucciones de procesamiento XML	Define la codificación de caracteres y la versión XML del documento de salida. El valor predeterminado es seleccionado.
Versión XML	Define la versión XML. La configuración de la versión XML tiene las siguientes opciones: - 1.0 - 1.1 El valor predeterminado es 1.0.

Valor	Descripción
Codificación	Define la codificación de caracteres que se ha especificado en la instrucción de proceso. La configuración de Codificación tiene las siguientes opciones: <ul style="list-style-type: none"> - Igual que la codificación de salida. La codificación de salida de las instrucciones de proceso es la misma que para la codificación de salida definida en la configuración de la transformación del Procesador de datos. - Personalizada. Define la codificación de salida en la instrucción de proceso. El usuario escribe el valor en el campo.
Añadir instrucciones de procesamiento personalizadas	Añade otras instrucciones de proceso para el documento de salida. Introduzca la instrucción de proceso tal como aparece en el documento de salida. El valor predeterminado es Deshabilitado.

La siguiente tabla describe la configuración de generación XML en el área de **Raíz XML**:

Valor	Descripción
Añadir elemento raíz XML	Añade un elemento raíz para el documento de salida. Utilice esta opción cuando el documento de salida contenga más de una ocurrencia del elemento raíz definida en el esquema de salida. El valor predeterminado es Deshabilitado.
Nombre de elemento raíz	Define un nombre del elemento raíz para añadirlo al documento de salida.

Eventos

Un evento es un registro de un paso de procesamiento de un componente en la transformación del Procesador de datos. En un script o biblioteca, cada anclaje, acción o transformador genera un evento. En un XMap, cada instrucción de asignación genera un evento.

Puede ver los eventos en la vista **Eventos del procesador de datos**.

Tipos de evento

La transformación del Procesador de datos escribe eventos en los archivos de registro. Cada evento tiene un tipo de evento que indica si el evento se ha realizado correctamente, el evento ha generado un error o si el evento se ha ejecutado con errores.

Un componente puede generar uno o más eventos. El componente puede pasar o generar un error según si los eventos tienen éxito o generan un error. Si un evento falla, un componente genera un error.

La siguiente tabla describe los tipos de eventos que la transformación del Procesador de datos genera:

Tipo de evento	Descripción
Notificación	Operación normal.
Advertencia	La transformación del Procesador de datos se ha ejecutado, pero se ha producido una condición inesperada. Por ejemplo, la transformación del Procesador de datos ha escrito datos para el mismo elemento varias veces. Cada vez que el elemento se sobrescribe, la transformación del Procesador de datos genera una advertencia.
Anomalía	La transformación del Procesador de datos se ha ejecutado, pero un componente ha generado un error. Por ejemplo, un elemento de entrada necesario estaba vacío.
Error opcional	La transformación del Procesador de datos se ha ejecutado, pero un componente opcional ha generado un error. Por ejemplo, un anclaje opcional faltaba en el documento de origen.
Error fatal	La transformación del Procesador de datos ha generado un error debido a un problema grave. Por ejemplo, el documento de entrada no existe.

Vista Eventos del procesador de datos

La vista **Eventos del procesador de datos** muestra eventos cuando se ejecuta una transformación del Procesador de datos desde la Herramienta del desarrollador.

La vista **Eventos del procesador de datos** tiene un panel **Navegación** y un panel **Detalles**. El panel de navegación contiene un árbol de navegación. El árbol de navegación enumera los componentes que la transformación ha ejecutado en orden cronológico. Cada nodo tiene un icono que representa el evento más importante debajo de él en el árbol. Cuando seleccione un nodo en el panel **Navegación**, aparecen eventos en el panel **Detalles**.

El árbol de navegación contiene los siguientes nodos de nivel superior:

- Inicialización del servicio. Describe los archivos y las variables que la transformación del Procesador de datos inicializa.
- Ejecución. Enumera los componentes que ha ejecutado el script, la biblioteca o XMap.
- Resumen. Muestra las estadísticas sobre el procesamiento.

Cuando se ejecuta un XMap, cada nombre de nodo del panel de navegación tiene un número entre corchetes, tal como [5]. Para identificar la instrucción que ha generado los eventos para el nodo, haga clic con el botón derecho en la malla de instrucciones y seleccione Ir al número de fila. Introduzca el número de nodo.

Cuando ejecute un script y haga doble clic en un evento en el panel **Navegación** o en el panel **Detalles**, el editor de scripts resaltará el componente de script que ha generado el evento. El panel **Entrada** de la vista **Visor de datos** resalta la parte del documento de origen de ejemplo que ha generado el evento.

Registros

Un registro contiene un registro de la transformación del Procesador de datos. La transformación del Procesador de datos escribe eventos en los registros.

La transformación del Procesador de datos crea los siguientes tipos de registros:

Registro de eventos en tiempo de diseño

El registro de eventos de tiempo de diseño contiene los eventos que ocurren cuando se ejecuta la transformación del Procesador de datos en la vista **Visor de datos** . Ver el registro en tiempo de diseño en la vista **Eventos**.

Registro de eventos de tiempo de ejecución

El registro de eventos de tiempo de ejecución contiene los eventos que ocurren cuando se ejecuta la transformación del Procesador de datos en una asignación. Puede ver el registro de eventos de tiempo de ejecución en un editor de texto o puede arrastrar un evento de tiempo de ejecución en la vista **Eventos** de la transformación del Procesador de datos.

Registro de usuario

El registro de usuario contiene eventos que se configuran para los componentes en un script. La transformación del Procesador de datos escribe en el registro de usuarios cuando se ejecuta desde la vista **Visor de datos** y cuando se ejecuta en una asignación. Puede ver el registro de usuario en un editor de texto.

Registro de eventos en tiempo de diseño

El registro de eventos de tiempo de diseño contiene los eventos generados cuando se ejecuta la transformación del Procesador de datos desde el **Visor de datos** en la Herramienta del desarrollador.

Cuando se ejecuta una transformación del Procesador de datos desde la vista **Visor de datos**, el registro de eventos de tiempo de diseño aparece en la vista **Eventos del procesador de datos**. De forma predeterminada, el registro de eventos de tiempo de diseño contiene notificaciones, advertencias y errores. En la configuración de transformación, puede configurar la transformación del Procesador de datos para excluir uno o más tipos de eventos del registro.

Cuando se guardan los documentos de entrada con el registro, puede hacer clic en un evento en la vista **Eventos del procesador de datos** para encontrar la ubicación en el documento de entrada que ha generado el evento. Cuando configure la configuración de la transformación del Procesador de datos, puede elegir guardar los archivos de entrada para cada ejecución o solo en caso de error.

El registro de eventos de tiempo de diseño se llama `events.cme`. Puede encontrar el registro de eventos de tiempo de diseño de la última ejecución de la transformación del Procesador de datos en el directorio siguiente:

```
C:\<Installation_directory>\clients\DT\CMReports\Init\events.cme
```

La transformación del Procesador de datos sobrescribe el registro de eventos de tiempo de diseño cada vez que se ejecuta la transformación en el **Visor de datos**. Cambie el nombre del registro de eventos de tiempo de diseño si desea verlo después de una posterior ejecución de la transformación, o si desea comparar los registros de distintas ejecuciones. Cuando se cierra la Herramienta del desarrollador, el desarrollador no guarda ningún archivo en el

Registro de eventos de tiempo de ejecución

El registro de eventos en tiempo de ejecución registra los eventos que ocurren cuando se ejecuta la transformación del Procesador de datos en una asignación.

Si la transformación del Procesador de datos finaliza la ejecución sin fallos, no escribe un registro de eventos. Si hay errores en la ejecución, la transformación del Procesador de datos se ejecuta por segunda vez y escribe un registro de eventos durante la segunda ejecución. El registro de eventos en tiempo de ejecución se llama `events.cme`.

En un equipo Windows, el registro de eventos en tiempo de ejecución se encuentra en el directorio siguiente:

```
C:<Installation_Directory>\clients\DT\CMReports\Tmp\
```

En un equipo UNIX o Linux, el registro de eventos en tiempo de ejecución se encuentra en el directorio siguiente:

```
/root/<Installation_Directory>/clients/DT/CMReports/Tmp
```

En un equipo UNIX o Linux, puede encontrar el registro de eventos en tiempo de ejecución para un usuario que no sea raíz en el directorio siguiente:

```
/home/[UserName]/<Installation_Directory>/DT/CMReports/Tmp
```

Utilice el editor de configuración para cambiar la ubicación del registro de eventos en tiempo de ejecución.

Visualizar un registro de eventos en la vista Eventos del Procesador de datos

Utilice la vista **Eventos del procesador de datos** para ver un registro de eventos de tiempo de diseño o un registro de eventos de tiempo de ejecución.

Abra el explorador de Windows y, a continuación, busque el archivo de registro de eventos que desee visualizar. Arrastre el registro desde el explorador de Windows a la vista **Eventos del procesador de datos**. Haga clic con el botón derecho en la vista **Eventos del procesador de datos** y, a continuación, seleccione **Buscar** para buscar el registro.

Nota: Para volver a cargar el último registro de eventos de tiempo de diseño, haga clic con el botón derecho en la vista **Eventos del procesador de datos** y, a continuación, seleccione **Volver a cargar los eventos del proyecto**.

Registro de usuario

El registro de usuario contiene mensajes personalizados que se configuran con fallos de componentes de un script.

La transformación del Procesador de datos escribe mensajes en el registro de usuario cuando se ejecuta un script desde la vista del **Visor de datos** y cuando se ejecuta en una asignación.

Cuando un componente del script tiene la propiedad **on_fail**, puede configurarlo para que escriba un mensaje en el registro de usuario si genera un error. En el script, configure la propiedad **on_fail** con uno de los siguientes valores:

- LogInfo
- LogWarning
- LogError

Cada ejecución del script genera un nuevo registro de usuario. El nombre del archivo de registro de usuario contiene el nombre de la transformación con un GUID único:

```
<Transformation_Name>_<GUID>.log
```

Por ejemplo, CalculateValue_Aa93a9d14-a01f-442a-b9cb-c9ba5541b538.log

En un equipo Windows, puede encontrar el registro de usuario en el directorio siguiente:

```
c:\Users\[UserName]\AppData\Roaming\Informatica\DataTransformation\UserLogs
```

En un equipo Linux o UNIX, puede encontrar el registro de usuario para el usuario raíz en el directorio siguiente:

```
/<Installation_Directory>/DataTransformation/UserLogs
```


En un equipo Linux o UNIX, puede encontrar el registro de usuario para un usuario que no sea raíz en el directorio siguiente:

```
home/<Installation_Dirctory>/DataTransformation/UserLogs
```

Desarrollo de la transformación del Procesador de datos

Utilice el asistente para nueva transformación para generar automáticamente una transformación del Procesador de datos o para crear una transformación del Procesador de datos en blanco y configurarla más adelante. Si crea una transformación del Procesador de datos en blanco, debe seleccionar un objeto Script, XMap, Biblioteca o Reglas de validación para crearlo en la transformación. Un script puede analizar documentos de origen a formato jerárquico, convertir el formato jerárquico en otros formatos de archivo o asignar un documento jerárquico a otro formato jerárquico. Un XMap convierte un archivo jerárquico de entrada en un archivo jerárquico de salida de otra estructura. Una biblioteca convierte un tipo de mensajería de industria en un documento XML con una estructura de jerarquías, o un formato XML en un formato estándar de la industria. Seleccione los esquemas que definan las jerarquías de entrada o salida.

1. Cree la transformación en Developer tool.
2. Para una transformación del Procesador de datos en blanco, lleve a cabo los siguientes pasos adicionales:
 - a. Añada referencias de esquema que definan las jerarquías XML de entrada o salida.
 - b. Cree un objeto Script, XMap, Biblioteca o Reglas de validación.
3. Configure los puertos de entrada y salida.
4. Pruebe la transformación.

Crear la transformación del Procesador de datos

Cree una transformación del Procesador de datos en Developer tool. Si crea una transformación del Procesador de datos en blanco, debe crear a continuación un objeto Script, XMap, Biblioteca o Reglas de validación para crearlo en la transformación. También puede utilizar el asistente para nueva transformación para generar automáticamente una transformación del Procesador de datos.

1. En Developer tool, haga clic en **Archivo > Nuevo > Transformación**.
2. Seleccione la transformación del Procesador de datos y haga clic en **Siguiente**.
3. Introduzca un nombre para la transformación y busque la ubicación en el repositorio de modelos donde colocar la transformación.
4. Seleccione si desea crear la transformación del Procesador de datos mediante un asistente o crear una transformación del Procesador de datos en blanco.
5. Si ha seleccionado crear una transformación del Procesador de datos en blanco, haga clic en **Finalizar**. Developer tool crea la transformación vacía en el repositorio. La vista **Resumen** aparece en Developer tool.
6. Si ha seleccionado crear una transformación del Procesador de datos con un asistente, realice los pasos siguientes:
 - a. Haga clic en **Siguiente**.

- b. Seleccione un formato de entrada.
- c. Seleccione un esquema, un cuaderno, un archivo de ejemplo o un archivo de especificación si es necesario para determinados formatos de entrada, como COBOL o JSON.
- d. Seleccione un formato de salida.
- e. Seleccione un esquema, un cuaderno, un archivo de ejemplo o un archivo de especificación si es necesario para el formato de salida.
- f. Haga clic en **Finalizar**. El asistente crea la transformación en el repositorio.

La transformación puede contener un analizador, un serializador, un asignador o un objeto con componentes comunes. Si ha seleccionado un esquema, un cuaderno, un archivo de ejemplo o un archivo de especificación, el asistente también crea un esquema en el repositorio que es equivalente a la jerarquía del archivo.

Seleccionar los objetos de esquema

Seleccione los objetos de esquema que definen las jerarquías de entrada o de salida para cada componente de XMap o de script que vaya a crear.

Puede añadir referencias de esquema en la vista Referencias o puede añadir las referencias de esquema al crear objetos de Script o XMap. El objeto de esquema debe estar en el repositorio de modelos antes de que pueda hacer referencia a él en un Script o XMap.

1. En la vista **Referencias** de la transformación del Procesador de datos, haga clic en **Añadir**.
2. Si el objeto de esquema existe en el repositorio de modelos, busque y seleccione el esquema.
3. Si el esquema no existe en el repositorio de modelos, haga clic en **Crear un nuevo objeto de esquema** e importe un objeto de esquema desde un archivo de esquema jerárquico.
4. Haga clic en Finalizar para añadir la referencia de esquema a la transformación del Procesador de datos.

Crear objetos en una transformación del Procesador de datos en blanco

Cree un objeto Script, Biblioteca, XMap o Reglas de validación en la vista **Objetos** de la transformación del Procesador de datos. Después de crear el objeto, puede abrirlo desde la vista **Objetos** para configurarlo.

Crear un script

Cree un objeto de script y defina el tipo de componente de script que va a crear. Opcionalmente, puede definir una referencia de esquema y un archivo de origen de ejemplo.

1. En la vista **Objetos** de la transformación del Procesador de datos, haga clic en **Nuevo**.
2. Especifique un nombre para el script y haga clic en **Siguiente**.
3. Elija crear un Analizador o un Serializador. Seleccione Otro para crear un componente de asignador, de transformador o de transmisión por secuencias.
4. Especifique un nombre para el componente.
5. Si el componente es el primer componente para procesar datos en la transformación, habilite **Establecer como componente de inicio**.
6. Haga clic en **Siguiente** si desea especificar una referencia de esquema para este script. Haga clic en **Finalizar** si no desea especificar la referencia de esquema.

7. Si elige crear una referencia de esquema, seleccione **Agregar referencia al objeto de esquema** y busque el objeto de esquema en el repositorio de modelos. Haga clic en **Crear un nuevo objeto de esquema** para crear el objeto de esquema en el repositorio de modelos.
8. Haga clic en **Siguiente** para especificar una referencia de origen de ejemplo o para especificar un texto de ejemplo. Haga clic en **Finalizar** si no desea definir un origen de ejemplo.
Utilice un origen de ejemplo para definir datos de ejemplo y para probar el script.
9. Si elige seleccionar un origen de ejemplo, seleccione **Archivo** y busque el archivo de ejemplo.
También puede introducir texto de ejemplo en el área **Texto**. Developer tool utiliza el texto para probar un script.
10. Haga clic en **Finalizar**.
La vista **Script** aparece en el editor de Developer tool.

Crear un XMap

Cree un XMap en la vista **Objetos** de Data Transformation. Cuando se crea un XMap, debe tener un esquema que describa la entrada y la salida de documentos jerárquicos. Seleccione el elemento del esquema que sea el elemento raíz para la jerarquía de entrada.

1. En la vista **Objetos** de la transformación del Procesador de datos, haga clic en **Nuevo**.
2. Seleccione XMap y haga clic en **Siguiente**.
3. Especifique un nombre para el XMap.
4. Si el componente XMap es el primer componente para procesar datos en la transformación, habilite **Establecer como componente de inicio**.
Haga clic en **Siguiente**.
5. Si elige crear una referencia de esquema, seleccione **Agregar referencia al objeto de esquema** y busque el objeto de esquema en el repositorio de modelos.
Para importar un nuevo objeto de esquema, haga clic en **Crear un nuevo objeto de esquema**.
6. Si tiene un archivo jerárquico de ejemplo que puede usar para probarlo con XMap, busque y seleccione el archivo en el sistema de archivos.
Puede cambiar el archivo jerárquico de ejemplo.
7. Elija la raíz de la jerarquía de entrada.
En el cuadro de diálogo **Selección del elemento raíz**, seleccione el elemento en el esquema que sea el elemento raíz del archivo jerárquico de entrada. Puede buscar un elemento en el esquema. Puede utilizar búsqueda de patrones. Introduzca `*<string>` para que coincida con cualquier número de caracteres en la cadena. Intro `?<character>` para que coincide con un solo carácter.
8. Haga clic en **Finalizar**.
La Herramienta del desarrollador crea una vista para cada XMap que cree. Haga clic en la vista para configurar la asignación.

Crear una biblioteca

Cree un objeto de biblioteca en la vista **Objetos** de Data Transformation. Seleccione el tipo de mensaje, el componente y el nombre. También puede definir un archivo de origen de tipo de mensaje de muestra que puede utilizar para probar el objeto de la biblioteca.

Antes de crear una biblioteca en la transformación del Procesador de datos, instale el paquete de software de la biblioteca en su equipo.

1. En la vista **Objetos** de la transformación del Procesador de datos, haga clic en **Nuevo**.
2. Seleccione la biblioteca y haga clic en **Siguiente**.
3. Busque el tipo de mensaje para seleccionarlo.
4. Elija crear un analizador o un serializador.
Cree un analizador si la entrada del objeto de la biblioteca es un tipo de mensaje y la salida es XML. Cree un serializador si la entrada del objeto de la biblioteca es XML y la salida es un tipo de mensaje.
5. Si la biblioteca es el primer componente para procesar datos en la transformación del Procesador de datos, habilite **Establecer como componente de inicio**.
Haga clic en **Siguiente**.
6. Si tiene un archivo de origen de tipo de mensaje de ejemplo que puede usar para probarlo con la biblioteca, busque y seleccione el archivo en el sistema de archivos.
Puede cambiar el archivo de ejemplo.
7. Haga clic en **Finalizar**.
La herramienta del desarrollador crea una vista para cada tipo de mensaje que cree. Haga clic en la vista para acceder a la asignación.

Crear reglas de validación

Cree un objeto Reglas de validación en la vista **Objetos** de la transformación del Procesador de datos.

1. En la vista **Objetos** de la transformación del Procesador de datos, haga clic en **Nuevo**.
2. Seleccione Reglas de validación y haga clic en **Siguiente**.
3. Especifique un nombre para las Reglas de validación.
4. Si tiene un archivo XML de ejemplo que puede usar para probarlo con las Reglas de validación, busque y seleccione el archivo en el sistema de archivos.
Puede cambiar el archivo XML de ejemplo.
5. Haga clic en **Finalizar**.
La Herramienta del desarrollador crea un objeto Reglas de validación y lo abre en el editor de Reglas de validación.

Cómo añadir un origen de ejemplo

Seleccione el origen de ejemplo para probar el script, XMap, la biblioteca o las reglas de validación que vaya a crear.

Puede añadir un origen de ejemplo al crear un script, XMap, una biblioteca o reglas de validación. Una vez seleccionado, el origen de ejemplo se añade al repositorio de modelos. Debido a las limitaciones del repositorio de modelos, el tamaño del archivo de origen de ejemplo está limitado a 5 MB.

Puede cambiar el origen de ejemplo.

Crear los puertos

Configure los puertos de entrada y salida en la vista **Resumen**.

Cuando configure puertos de entrada o de salida adicionales en un script, Developer tool añade puertos de entrada y de salida adicionales a la transformación de manera predeterminada. No puede añadir puertos de entrada en la vista **Resumen**.

1. Si desea que se devuelvan filas de datos de salida en lugar de XML, habilite **Salida relacional**.
Al habilitar la salida relacional, Developer tool elimina el puerto de salida predeterminado.
2. Seleccione el tipo de datos, el tipo de puerto, la precisión y la escala del puerto de entrada.
3. Si no define los puertos de salida relacional, defina el tipo de datos, el tipo de puerto, la precisión y la escala del puerto de salida.
4. Si un Script tiene puertos de entrada adicionales, puede definir la ubicación del archivo de entrada de ejemplo para los puertos. Haga clic en el botón **Abrir** del campo **Ubicación de entrada** para buscar el archivo.
5. Si ha habilitado la salida relacional, haga clic en **Salida de asignación** para crear los puertos de salida.
6. En la vista Puertos, asigne nodos desde la zona de **Salida jerárquica** a los campos de la zona **Puertos relacionales**.

Probar la transformación

Pruebe la transformación del Procesador de datos en la vista **Visor de datos**.

Antes de probar la transformación, compruebe que ha definido el componente de inicio. Puede definir el componente de inicio en un script o puede seleccionar el componente de inicio en la pestaña **Resumen**. También debe haber elegido un archivo de entrada de ejemplo con el que probarlo.

1. Abra la vista **Visor de datos**.
2. Haga clic en **Ejecutar**.
Developer tool valida la transformación. Si no hay errores, Developer tool muestra el archivo de ejemplo en la zona de **Entrada**. Los resultados de salida aparecen en el panel de Salida.
3. Haga clic en **Mostrar eventos** para mostrar la vista **Eventos del procesador de datos**.
4. Haga doble clic en un evento de la vista **Eventos del procesador de datos** para depurar los eventos del editor del script.
5. Haga clic en **Sincronizar con el editor** para cambiar el archivo de entrada cuando pruebe varios componentes, cada uno con un archivo de entrada de ejemplo diferente.
Si modifica el contenido del archivo de ejemplo en el sistema de archivos, los cambios aparecen en la zona de **Entrada**.

Exportación e importación de transformaciones del Procesador de datos

Puede exportar una transformación del Procesador de datos como servicio y ejecutarla desde un repositorio de Data Transformation. También puede importar un servicio de Data Transformation a la Herramienta del

desarrollador. Cuando se importa un servicio de Data Transformation, la Herramienta del desarrollador crea una transformación del Procesador de datos desde el servicio.

Nota: Cuando se importa un servicio de Data Transformation al repositorio de modelos, la Developer Tool importa los esquemas asociados con el repositorio. Si modifica el esquema en el repositorio, a veces los cambios no aparecen en las referencias del esquema de transformación. Puede cerrar y abrir la conexión del repositorio de modelos, o bien cerrar y abrir Developer tool para que los cambios en el esquema aparezcan en la transformación.

Exportar la transformación del Procesador de datos como un servicio

Puede exportar la transformación del Procesador de datos como un servicio de Data Transformation. Exporte el servicio al repositorio del sistema de archivos de la máquina donde quiera ejecutar el servicio. Puede ejecutar el servicio con PowerCenter, aplicaciones definidas por el usuario o el comando CM_console de Data Transformation.

1. En la vista **Explorador de objetos**, haga clic con el botón derecho en la transformación del Procesador de datos que desea exportar y seleccione **Exportar**.
A continuación, aparece el cuadro de diálogo **Exportar**.
2. Seleccione **Informática > Exportar transformación del Procesador de datos** y haga clic en **Siguiente**.
Aparece la página **Seleccionar**.
3. Haga clic en **Siguiente**.
Aparece la página **Seleccione el nombre de servicio y la carpeta de destino**.
4. Elija una carpeta de destino:
 - Para exportar el servicio en el equipo que aloja la Herramienta del desarrollador, haga clic en **Carpeta de servicio**.
 - Para implementar el servicio en otro equipo, haga clic en **Carpeta**. Busque en el directorio `\ServiceDB` del equipo donde desee implementar el servicio.
5. Haga clic en **Finalizar**.

Importar varios servicios de Data Transformation

Puede importar un directorio de servicios de Data Transformation desde el equipo donde ese directorio esté guardado. Cuando se importan servicios de Data Transformation al repositorio de modelos de Developer, Developer tool importa las transformaciones, los esquemas y los datos de ejemplo junto con los archivos .cmw. Si necesita importar una gran cantidad de servicios, importe un directorio de servicios en lugar de hacerlo uno a uno.

1. Haga clic en **Archivo > Importar**.
Aparece el cuadro de diálogo **Importar**.
2. Seleccione **Informática Importar servicio de Data Transformation (carpeta)** y haga clic en **Siguiente**.
Aparece la página **Importar servicio de Data Transformation**.
3. Busque el directorio que desee importar.
4. Busque una ubicación en el repositorio donde desee guardar la transformación y haga clic en **Finalizar**.
Developer tool importa las transformaciones, los esquemas y los datos de ejemplo junto con el archivo .cmw.

Importar un servicio de Data Transformation

Puede importar un archivo .cmw de un servicio de Data Transformation al repositorio de modelos para crear una transformación de Procesador de datos. Developer tool importa la transformación, los esquemas y los datos de ejemplo con el archivo .cmw.

1. Haga clic en **Archivo > Importar**.
Aparece el cuadro de diálogo **Importar**.
2. Seleccione **Informatica Importar servicio de Data Transformation (único)** y haga clic en **Siguiente**.
Aparece la página **Importar servicio de Data Transformation**.
3. Busque el archivo .cmw del servicio que desee importar.
Developer tool denomina la transformación según el nombre de archivo del servicio. Puede cambiar el nombre.
4. Busque una ubicación en el repositorio donde desee guardar la transformación y, a continuación, haga clic en **Finalizar**.
Developer tool importa la transformación, los esquemas y los datos de ejemplo con el archivo .cmw.
5. Si desea editar la transformación, haga doble clic en la transformación en la vista **Explorador de objetos**.

Exportación de una asignación con una transformación del Procesador de datos a PowerCenter

Cuando exporta una asignación con una transformación del Procesador de datos a PowerCenter, podrá exportar los objetos a un archivo local y, a continuación, importar la asignación en PowerCenter. También es posible exportar la asignación directamente al repositorio de PowerCenter.

1. En la vista **Explorador de objetos**, seleccione la asignación que desea exportar. Haga clic con el botón derecho y seleccione **Exportar**.
A continuación, aparece el cuadro de diálogo **Exportar**.
2. Seleccione **Informatica > PowerCenter**.
3. Haga clic en **Siguiente**.
A continuación, aparece el cuadro de diálogo **Exportar a PowerCenter**.
4. Seleccione el proyecto.
5. Seleccione la versión de PowerCenter.
6. Elija la ubicación para la exportación, un archivo de importación XML a PowerCenter o un repositorio de PowerCenter.
7. Especifique las opciones de exportación.
8. Haga clic en **Siguiente**.
La Herramienta del desarrollador le pide que seleccione los objetos para exportarlos.
9. Seleccione los objetos para exportar y haga clic en **Finalizar**.
La Herramienta del desarrollador exporta los objetos a la ubicación seleccionada. Si ha exportado la asignación a una ubicación, la Herramienta del desarrollador también exporta las transformaciones del Procesador de datos de la asignación, como servicios, a una carpeta en la ubicación que haya especificado.
10. Si ha exportado la asignación a un repositorio de PowerCenter, los servicios se exportan a la siguiente ruta de acceso del directorio: %temp%\DTServiceExport2PC\

La función de exportación crea una carpeta independiente para cada servicio con el siguiente nombre:
<date><serviceFullName>

Si la transformación incluye la asignación relacional, se creará una carpeta para la asignación relacional a jerárquica y otra carpeta independiente para la asignación jerárquica a relacional.

11. Copie la carpeta o carpetas con los servicios de transformación del Procesador de datos desde la ubicación local donde ha exportado hasta la carpeta ServiceDB de PowerCenter.
12. Si ha exportado la asignación a un archivo XML de importación de PowerCenter, importe la asignación en PowerCenter. Para obtener más información sobre cómo importar un objeto en PowerCenter, consulte la *Guía del repositorio de PowerCenter 9.6.0*.

Transformación del procesador de datos Validación

Después de exportar una transformación del Procesador de datos como servicio, puede ejecutar validaciones de VRL en el servicio desde el repositorio de Data Transformation.

Puede utilizar un motor de Data Transformation con velocidad mejorada para las validaciones de VRL. El motor de Data Transformation con velocidad mejorada admite las siguientes funciones de VRL:

- dt:exist
- dt:empty
- dt:date-valid
- dt:next-sequence
- dt:all-equal
- dt:lookup
- dt:regex-match

El motor de Data Transformation con velocidad mejorada genera la salida **ValidateValue**. **ValidateValue** contiene la propiedad `max_error_count`, con un valor predeterminado de 200 errores. Cuando el número de errores excede el valor de `max_error_count`, la validación se detiene.

Nota: La sintaxis de VRL del motor de Data Transformation con velocidad mejorada no admite la etiqueta <list>.

Utilizar un motor de Data Transformation con velocidad mejorada para las validaciones de VRL

Después de exportar un servicio de Data Transformation, puede utilizar un motor de Data Transformation con velocidad mejorada para las validaciones de VRL con el servicio.

- Establezca la siguiente marca en el archivo `.cmw` del servicio `optimize_vrl`.

Añada la marca `optimize_vrl` a la instancia `ServiceConfigProf`, como se muestra en el siguiente ejemplo:

```
instance ServiceConfig = ServiceConfigProf<add_required_xml_elements,  
add_required_xml_attributes, optimize_vrl>
```


Transformación del procesador de datos en un entorno no nativo

El procesamiento de transformaciones del procesador de datos en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.*
- Motor de Databricks Spark. No compatible.

** Para obtener información sobre la compatibilidad de la transformación del procesador de datos en el motor de Spark, vea el [KB article](#).*

CAPÍTULO 14

Transformación de decisión

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de decisión, 266](#)
- [Funciones de transformación de decisión, 267](#)
- [Instrucciones condicionales de transformación de decisión , 269](#)
- [Operadores de transformación de decisión, 270](#)
- [Control de valores NULL en transformaciones de decisión, 271](#)
- [Cómo configurar una estrategia de decisión , 271](#)
- [Propiedades avanzadas de Transformación de decisión, 272](#)
- [Transformación de decisión en un entorno no nativo, 272](#)

Resumen de la transformación de decisión

La transformación de decisión es una transformación pasiva que evalúa condiciones en datos de entrada y crea una salida basada en los resultados de esas condiciones.

Configure una transformación de decisión para generar valores diferentes según los valores que se encuentran en los campos de entrada. Por ejemplo, si los ingresos de un cliente exceden un importe específico, puede añadir la cadena "Prioridad" al nombre del cliente.

Puede añadir varias estrategias de decisión a la transformación de decisión. Cada estrategia evalúa una instrucción condicional `IF-THEN-ELSE`. Dentro de esta instrucción, puede utilizar condiciones `ELSEIF` o anidar instrucciones `IF-THEN-ELSE` adicionales.

La transformación de decisión es similar a la transformación de expresión en cuanto que permite utilizar instrucciones condicionales y funciones para comprobar los datos de origen. Sin embargo, la transformación de decisión difiere de la transformación de expresión en los siguientes aspectos:

- La transformación de decisión utiliza instrucciones `IF-THEN-ELSE` para evaluar condiciones. La transformación de expresión utiliza instrucciones `IIF`.
- La transformación de decisión contiene funciones que no están disponibles en la transformación de expresión.
- Cada estrategia de decisión puede generar varias salidas.

Funciones de transformación de decisión

La transformación de decisión proporciona acceso a funciones predefinidas que se utilizan para definir estrategias de decisión.

El editor de transformaciones de decisión contiene una carpeta Decision. Esta carpeta contiene funciones específicas a la transformación de decisión. Asimismo, el editor contiene otras carpetas que proporcionan acceso a funciones de transformación de expresión.

Cuando se hace clic en una función del editor de expresiones, la transformación muestra el uso y los tipos de datos de la función, además de una descripción de lo que hace la función.

Nota: No todas las funciones de una transformación de expresión son compatibles con la transformación de decisión. La transformación de decisión solamente proporciona acceso a funciones compatibles de una transformación de expresión.

Lista de funciones de una transformación de decisión

- ABS
- ADD_TO_DATE
- ASCII
- CEIL
- CHOOSE
- CHR
- CHRCODE
- CONCAT
- CONTAINS
- CONVERT_BASE
- COS
- COSH
- CRC32
- CUME
- CURDATE
- CURTIME
- DATE_COMPARE
- DATE_DIFF
- DATECONVERT
- EXP
- FLOOR
- FV
- GET_DATE_PART
- GREATEST
- IN
- INDEXOF
- INITCAP
- INSTR

- IS_DATE
- IS_NUMBER
- ISNULL
- LAST_DAY
- LEAST
- LEFTSTR
- LENGTH
- LN
- LOG
- LOWER
- LPAD
- LTRIM
- MAKE_DATE_TIME
- MAX
- MD5
- METAPHONE
- MIN
- MOD
- MONTHCOMPARE
- MOVINGAVG
- MOVINGSUM
- NPER
- PMT
- POWER
- PV
- RAND
- RATE
- REG_EXTRACT
- REG_MATCH
- REG_REPLACE
- REPLACECHR
- REPLACESTR
- REVERSE
- RIGHTSTR
- ROUND
- RPAD
- RTRIM
- SET_DATE_PART
- SIGN

- SIN
- SINH
- SOUNDEX
- SQRT
- SUBSTR
- TAN
- TANH
- TIMECOMPARE
- TO_CHAR
- TO_DATE
- TO_FLOAT
- TO_INTEGER
- TRUNC
- UPPER
- XOR

Nota: Utilice un valor constante para definir el formato de fecha en las funciones CURDATE, DATE_COMPARE, DATECONVERT y MONTHCOMPARE de la transformación de decisión.

Instrucciones condicionales de transformación de decisión

La transformación de decisión utiliza instrucciones condicionales IF-THEN-ELSE para evaluar datos de entrada.

Dentro de estas instrucciones condicionales se pueden usar condiciones ELSEIF o anidar instrucciones IF-THEN-ELSE adicionales. Las instrucciones condicionales de una transformación de decisión utilizan el siguiente formato:

```
// Primary condition
IF <Boolean expression>
THEN <Rule Block>
// Optional - Multiple ELSEIF conditions
ELSEIF <Boolean expression>
THEN <Rule Block>
// Optional ELSE condition
ELSE <Rule Block>
ENDIF
```

Se pueden anidar instrucciones condicionales adicionales dentro de un bloque de reglas.

Operadores de transformación de decisión

Utilice operadores de transformación de decisión para definir estrategias de decisión.

En la siguiente tabla se describen los operadores de la transformación de decisión:

Tipo de operador	Operador	Descripción
Asignación	:=	Asigna un valor a un puerto.
Booleana	AND	Añade una condición lógica requerida. Para que la expresión booleana superior sea verdadera, todas las condiciones lógicas vinculadas mediante este operador deben ser verdaderas.
Booleana	OR	Añade una condición lógica. Para que la expresión booleana superior sea verdadera, por lo menos una condición lógica vinculada mediante este operador debe ser verdadera.
Booleana	NOT	Añade una condición lógica negativa. Para que la expresión booleana superior sea verdadera, la condición negativa especificada mediante este operador debe ser verdadera.
Decisión	=	Comprueba si los elementos comparados son iguales. Utilícese con tipos de datos de cadena o numéricos.
Decisión	<>	Comprueba si los elementos comparados no son iguales. Utilícese con tipos de datos de cadena o numéricos.
Decisión	<	Comprueba si un valor es menor que otro valor. Utilícese con tipos de datos numéricos.
Decisión	<=	Comprueba si un valor es menor o igual que otro valor. Utilícese con tipos de datos numéricos.
Decisión	>	Comprueba si un valor es mayor que otro valor. Utilícese con tipos de datos numéricos.
Decisión	>=	Comprueba si un valor es mayor o igual que otro valor. Utilícese con tipos de datos numéricos.
Numérica	-	Resta
Numérica	NEG	Negación
Numérica	+	Suma
Numérica	*	Multiplicación
Numérica	/	División
Numérica	%	Módulo. Devuelve el resto resultante de dividir un número entre otro.
Cadena		Concatena cadenas.

Control de valores NULL en transformaciones de decisión

El control de valores NULL determina cómo el servicio de integración de datos procesa datos de valores nulos en una transformación de decisión.

Cuando se habilita el control de valores NULL, la transformación de decisión conserva el formato de datos de entrada NULL. La transformación evalúa las funciones utilizando el valor de entrada NULL.

Al deshabilitar el control de valores NULL, la transformación de decisión asigna un valor predeterminado a los datos de entrada NULL. La transformación evalúa las funciones utilizando el valor predeterminado. Por ejemplo, si un campo de entrada de tipo entero tiene un valor nulo, la transformación de decisión asigna un valor de 0 a la entrada y evalúa las funciones utilizando un valor de entrada de 0.

De forma predeterminada, el control de valores NULL no está habilitado en la transformación de decisión. El control de valores NULL se habilita en la ficha **Estrategias**. Puede habilitar el control de valores NULL tras configurar una estrategia para la transformación.

Cómo configurar una estrategia de decisión

Para configurar una estrategia de decisión, conecte los datos de origen a la transformación de decisión y edite las propiedades en las vistas de la transformación.

1. Abra una transformación de decisión.
2. Compruebe que la transformación contiene puertos de entrada y de salida.
3. Seleccione la vista **Decisión**.
4. Haga clic en **Añadir**.
5. Especifique un nombre para la estrategia.
6. En el área **Expresión**, introduzca una instrucción condicional **IF-THEN-ELSE**.
7. Para añadir una función a la expresión, explore las funciones que aparecen en la ficha **Funciones** y haga doble clic en un nombre de función.

Sugerencia: Para introducir una función rápidamente, escriba las primeras letras del nombre de la función y seleccione **CTRL-Espacio**.

8. Para añadir un puerto a la expresión, explore los puertos que aparecen en la ficha **Puertos**. Haga doble clic en un nombre de puerto para añadirlo a la expresión. Si lo desea, haga clic en **Editar puertos de salida** para editar los ajustes de los puertos de salida o para añadir puertos de salida.
9. Si lo desea, añada líneas de comentario escribiendo **"/"** seguido de los comentarios.
10. Haga clic en **Validar** para determinar si la expresión de decisión es válida.
11. Haga clic en **Aceptar** para guardar la estrategia.
12. Si lo desea, añada estrategias adicionales. Cada estrategia debe utilizar puertos de salida únicos. Las estrategias no pueden compartir puertos de salida.

Propiedades avanzadas de Transformación de decisión

Configure propiedades que ayudan a determinar cómo el Servicio de integración de datos procesa los datos para la Transformación de decisión.

Configure las siguientes propiedades avanzadas para una transformación de decisión:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Admite partición

La transformación se puede procesar con varios subprocesos. Desactive esta opción si quiere que el Servicio de integración de datos utilice un subproceso para procesar la transformación. El Servicio de integración de datos puede usar varios subprocesos para procesar el resto de etapas del canal de la asignación.

Puede que desee desactivar la partición para una transformación de decisión cuando la transformación utiliza una de estas funciones numéricas: CUME, MOVINGSUM o MOVINGAVG. Estas funciones calculan los totales y los promedios en ejecución fila por fila. Una transformación particionada que utiliza las funciones CUME, MOVINGSUM o MOVINGAVG podría no devolver los mismos resultados calculados con cada ejecución de asignación.

Si la transformación no utiliza las funciones CUME, MOVINGSUM ni MOVINGAVG, habilite la partición para la transformación para optimizar el rendimiento.

Transformación de decisión en un entorno no nativo

El procesamiento de transformaciones de decisión en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de expresión en el motor de Spark

La transformación de decisión es compatible con las siguientes restricciones:

- Las propiedades de la transformación deben especificar que la transformación es divisible en particiones.

Transformación de decisión en el motor de Databricks Spark

La transformación de decisión es compatible con las siguientes restricciones:

- Las propiedades de la transformación deben especificar que la transformación es divisible en particiones.

CAPÍTULO 15

Transformación de excepción de registros duplicados

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de excepción de registros duplicados, 273](#)
- [Flujo del proceso de la excepción de registros duplicados, 274](#)
- [Excepciones de registros duplicados, 274](#)
- [Vista de configuración de excepción de registros duplicados , 275](#)
- [Puertos, 277](#)
- [Propiedades avanzadas de la transformación de excepción de registros duplicados, 279](#)
- [Ejemplo de asignación de excepción de registros duplicados, 280](#)
- [Creación de una transformación de excepción de registros duplicados, 286](#)

Resumen de la transformación de excepción de registros duplicados

La transformación de excepción de registros duplicados es una transformación activa que lee la salida de un proceso de calidad de datos e identifica registros duplicados que requieren la revisión manual. La transformación de excepción de registros duplicados es una transformación de varios grupos.

La transformación de excepción de registros duplicados recibe la entrada de otra transformación o de un objeto de datos en otra asignación. La entrada a la transformación de excepción debe contener un valor numérico de puntuación de coincidencia que la transformación pueda utilizar para determinar si el registro es un duplicado. Establezca un umbral de puntuación de coincidencia superior e inferior en la transformación de excepción de registros duplicados.

La transformación de excepción de registros duplicados realiza una de las siguientes acciones:

- Si la puntuación de coincidencia es mayor o igual al umbral superior, la transformación trata el registro como un duplicado y lo escribe en un destino de base de datos.
- Si la puntuación de coincidencia es menor que el umbral superior y mayor que el umbral inferior de la transformación trata el registro como un posible duplicado y lo escribe en el registro en un destino diferente para la revisión manual. Si el registro pertenece a un clúster, la transformación escribe todos los registros del clúster en el destino.

- Cuando un clúster tiene una puntuación de coincidencia menor que el umbral inferior, todos los registros del clúster van al grupo de salida de registros únicos. Los clústeres de tamaño 1 se distribuyen al grupo único, independientemente de la puntuación de coincidencia. De forma predeterminada, la transformación de excepción no escribe registros únicos en un destino. Puede configurar la transformación para devolver los registros únicos.
- Si una puntuación de coincidencia de un clúster no está en el intervalo de 0 - 100, la transformación de excepción ignora todas las filas del clúster. El Servicio de integración de datos registra un mensaje que incluye el clusterID.

Flujo del proceso de la excepción de registros duplicados

La transformación de excepción de registros duplicados analiza la salida de otras transformaciones de calidad de datos y crea las tablas que contienen los registros con distintos niveles de calidad de datos.

Puede configurar las transformaciones de calidad de datos en una única asignación individual o puede crear asignaciones en distintas etapas del proceso.

Puede utilizar la herramienta del analista para revisar y actualizar los registros duplicados que se deben revisar de forma manual.

Utilice Developer Tool para completar las siguientes tareas:

1. Cree una asignación que genera valores de puntuación para problemas de calidad de datos.
2. Utilice una transformación de coincidencia en modo de clústeres para generar valores de puntuación para excepciones de registros duplicados.
3. Configure la transformación de excepción de registros duplicados para leer la salida de transformación de coincidencia. Configure la transformación para escribir los registros en las tablas de base de datos según los valores de puntuación de coincidencia de los registros.
4. Configure los objetos de datos de destino para los registros de consolidación automática.
5. Haga clic en la opción **Generar tabla de registros duplicados** para crear la tabla de registros duplicados y añádala al lienzo de la asignación.
6. Añada la asignación a un flujo de trabajo.
7. Configure una tarea humana para asignar la revisión manual de posibles registros duplicados a los usuarios. Los usuarios pueden revisar y actualizar los registros en la herramienta del analista.

Excepciones de registros duplicados

Puede utilizar una transformación de excepción de registros duplicados para identificar clústeres de datos duplicados que necesitan revisión manual. Las puntuaciones de coincidencia de registros en clústeres determina los posibles duplicados. Puede configurar los umbrales superior e inferior para las puntuaciones de coincidencia de la transformación. El los umbrales superior e inferior defina el grado de similitud.

Un clúster contiene registros relacionados que una operación de coincidencia agrupa simultáneamente. La transformación de coincidencia crea clústeres con la operación de análisis de duplicados y la operación de resolución de identidad. Cada registro de un clúster tiene el mismo ID de clúster. Cuando la puntuación de

coincidencia menor de un clúster está entre los umbrales superior e inferior, la transformación de excepción de registros duplicados identifica el clúster como un clúster de excepción de registros duplicados. La transformación de coincidencia añade una columna de valor de ID de clúster todos los registros. Los registros duplicados reciben el mismo ID de clúster.

La puntuación de registro más baja determina el tipo de clúster. Un clúster podría tener 11 registros que tienen una puntuación de coincidencia de 0,95 y un registro con la puntuación de coincidencia de 0,79. Si el umbral superior es 0,9 y el umbral inferior es 0,8, la transformación de excepción escribe los registros en la tabla de registros únicos.

Vista de configuración de excepción de registros duplicados

Defina los umbrales de puntuación de coincidencia y configure dónde la transformación de excepción de registros duplicados escribirá los diferentes tipos de datos de salida.

La siguiente figura muestra las propiedades que puede configurar:

Umbrales de revisión manual

Umbral inferior : 0,80

Umbral superior : 0,95

Opciones de distribución de datos

Tipo	Salida estándar	Tabla de registro duplicado
Consolidación automática (sobre el umbral superior)	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Consolidación manual (entre umbrales)	<input type="checkbox"/>	<input type="checkbox"/>
Registros únicos (por debajo del umbral inferior)	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Crear un grupo de salida aparte para los registros únicos

Generar tabla de registros duplicados

Puede configurar las propiedades siguientes:

Umbral inferior

El límite inferior para el intervalo de puntuaciones de registros duplicados. La transformación procesa los registros con puntuaciones de coincidencia inferiores a este valor como registros únicos. El valor del umbral inferior es un número de 0 a 1.

Umbral superior

El límite superior para el intervalo de puntuaciones de registros duplicados. La transformación procesa los registros con puntuaciones de coincidencia mayores o iguales que el umbral superior como registros duplicados. El valor del umbral superior es un número mayor que el umbral inferior.

Consolidación automática

Clústeres en los que todos los registros tienen puntuaciones de coincidencia mayores que el umbral superior. Los clústeres de consolidación automática no necesitan revisión. Los registros son duplicados. Puede utilizar la transformación de consolidación para combinar los registros. De forma

predeterminada, la transformación de excepción de registros duplicados escribe clústeres de consolidación automática en puertos de salida estándar.

Consolidación manual

Clústeres en los que todos los registros tienen puntuaciones de coincidencia mayores o iguales que el umbral inferior y al menos un registro tiene una puntuación de coincidencia menor que el umbral superior. Debe realizar una revisión manual de los clústeres para determinar si contienen registros genuinos duplicados. De forma predeterminada, la transformación de excepción de registros duplicados escribe registros de consolidación manual en la tabla de registros duplicados.

Consolidación única

Clústeres con un tamaño de clúster igual a uno o clústeres en los que cualquier registro tiene una puntuación de coincidencia menor que el umbral inferior. Los clústeres de registro único no son duplicados. De forma predeterminada, la transformación de excepción de registros duplicados no escribe registros únicos en una tabla de salida.

Salida estándar

Los tipos de registro que la transformación escribe en los puertos de salida estándar.

El valor predeterminado es registros de consolidación automática.

Tabla de registros duplicados

Los tipos de registro que la transformación escribe en los puertos de salida de registros duplicados. El valor predeterminado es registros de consolidación manual.

Crear un grupo de salida aparte para los registros únicos

Crea un grupo de salida aparte para los registros únicos. Si no crea una tabla aparte para los registros únicos, puede configurar la transformación para escribir los registros únicos en uno de los otros grupos. O bien, puede omitir la escritura de registros únicos en una tabla de salida. De forma predeterminada esta opción está deshabilitada.

Generar tabla de registros duplicados

Crea un objeto de base de datos para contener los datos de clústeres de registros duplicados. Cuando se selecciona esta opción, Developer tool crea el objeto de base de datos. Developer tool añade el objeto en el repositorio de modelos, añade una instancia del objeto al lienzo de la asignación y vincula los puertos al objeto.

Cómo generar una tabla de registros duplicados

Puede generar una tabla de registros duplicados desde una instancia de transformación de excepción de registros duplicados en una asignación.

1. En la vista **Configuración**, haga clic en **Generar tabla de registros duplicados** para generar la tabla. A continuación, aparece el cuadro de diálogo **Crear objeto de datos relacionales**.
2. Busque y seleccione una conexión a la base de datos donde guardar la tabla.
3. Especifique un nombre para la tabla de registros duplicados de la base de datos.
4. Especifique un nombre para el objeto de la tabla de registros duplicados del repositorio de modelos.
5. Haga clic en **Finalizar**.
Developer Tool añade la nueva tabla al lienzo de la asignación.

Puertos

La transformación de excepción de registros duplicados tiene varios grupos de puertos de entrada y salida.

La siguiente figura muestra un ejemplo de los puertos de entrada y salida:

Puertos								
	Nombre	Tipo	Precisión	Escala	Entrada	Salida	Predeterminado	Descripción
[-] Entradas								
[-] Datos (3)								
1	Employee	decimal	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
2	Name	string	50	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
3	Addr1	string	50	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
[-] Control (3)								
1	Puntuación	double	15	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
2	Row_Identifier	string	25	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
3	Cluster_ID	integer	10	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
[-] Salidas								
[-] Salida estándar (6)								
1	Puntuación	double	15	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
2	Row_Identifier	string	25	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
3	Cluster_ID	integer	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
4	Employee	decimal	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
5	Name	string	50	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
6	Addr1	string	50	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
[-] Datos del clúster (9)								
1	Row_Identifier	bigint	19	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
2	Sequential_Cl...	bigint	19	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
3	Cluster_ID	integer	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		

Puertos de entrada de transformación de excepción de registros duplicados

Una transformación de excepción de registros duplicados tiene un grupo de datos y un grupo de control de puertos de entrada.

El grupo **Datos** contiene puertos definidos por el usuario que reciben los datos de origen.

Los puertos de **control** reciben los metadatos que la transformación de coincidencia añade a los datos de origen. La siguiente tabla describe los puertos de **control**:

Puerto	Descripción
Puntuación	Valor decimal entre 0 y 1. Identifica el grado de similitud con el registro que enlazaba el registro con el clúster.
Row_Identifier	Identificador único del registro.
Cluster_ID	Identificador del clúster de coincidencias al que pertenece el registro.

Puertos de salida de transformación de excepción de registros duplicados

La transformación de excepción de registros duplicados tiene dos grupos de salida múltiple. De forma predeterminada, la transformación escribe registros duplicados para el grupo **Salida estándar**. La transformación escribe coincidencias potenciales en el grupo **Datos del clúster**. Se puede añadir un grupo de salida para registros únicos.

Puede cambiar los tipos de registro que la transformación escribe en los puertos de salida cambiando la configuración predeterminada en la vista **Configuración**.

En la siguiente tabla se describen los puertos de salida para el grupo **Salida estándar**:

Puerto	Descripción
Puntuación	Valor decimal entre 0 y 1. Identifica el grado de similitud entre el registro y otro registro en un clúster.
Row_Identifier	Identificador único del registro.
Cluster_ID	ID de clúster que la transformación de coincidencia asignado al registro.
Puertos definidos por el usuario	Los campos de datos de origen.

La siguiente tabla describe los puertos de salida del grupo **Datos del clúster**:

Puerto	Descripción
Row_Identifier	Identificador único del registro.
Sequential_Cluster_ID	Identifica el clúster en una tarea humana. Un flujo de trabajo utiliza el ID de clúster secuencial para asignar un clúster a una instancia de una tarea humana.
Cluster_ID	Identifica el clúster al que pertenece el registro. La transformación de coincidencia asigna un ID de clúster a todos los registros.
Puntuación	Valor decimal de 0 a 1. Identifica el grado de similitud con el registro que enlazaba el registro con el clúster.
Is_Master	Valor de la cadena que indica si el registro es el registro preferido del clúster. De forma predeterminada, la primera fila del clúster es el registro de localidad. El valor es Y o N.
Workflow_ID	ID que identifica el flujo de trabajo para el registro en una tarea. Cuando se ejecuta la asignación fuera de un flujo de trabajo, el ID de flujo de trabajo es DummyWorkflowID.
Puertos definidos por el usuario	Los puertos de datos de origen.

Cómo crear puertos

Añada cada puerto de entrada al grupo de datos. Cuando se añade el puerto de entrada, Developer Tool añade un puerto de salida con el mismo nombre al grupo de salida estándar, al grupo de datos del clúster y al grupo de registros únicos.

1. Seleccione el grupo de entrada.
El grupo se resalta.
2. Haga clic en **Nuevo (Insertar)**.
Developer Tool añade un campo al grupo de datos, grupo de salida estándar, al grupo de datos del clúster y al grupo de registros únicos.
3. Cambie el nombre del campo según sea necesario.
Developer Tool cambia el nombre de los otros grupos.
4. Introduzca el resto de los puertos que tenga que añadir al origen de datos.

Propiedades avanzadas de la transformación de excepción de registros duplicados

La transformación de excepción de registro duplicado contiene propiedades avanzadas que determinan el comportamiento de ordenación y de la memoria caché y el nivel de seguimiento.

Puede configurar las siguientes propiedades avanzadas:

Ordenar

Determina si la transformación ordena las filas de entrada en el puerto de datos **ID de clúster**. La propiedad está habilitada de forma predeterminada.

Seleccione la propiedad si las filas de entrada no se han ordenado previamente.

Directorio de archivo de memoria caché

Especifica el directorio en el que el servicio de integración de datos escribe los datos temporales de la transformación actual. El servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos de entrada es mayor que la memoria disponible del sistema. El servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del servicio de integración de datos. El servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de archivo de memoria caché

Determina la cantidad de memoria del sistema que utiliza el servicio de integración de datos para ordenar los datos de entrada en la transformación. El valor predeterminado es 400.000 bytes.

Antes de ordenar los datos, el servicio de integración de datos asigna la cantidad de memoria que se haya especificado. Si la operación de ordenación genera una cantidad de datos mayor, el servicio de integración de datos escribirá los datos en exceso en el directorio de archivos de memoria caché. Si la operación de clasificación necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

Nota: Si especifica un valor de 65536 o superior, la transformación leerá el valor en bytes. Si especifica un valor inferior, la transformación leerá el valor en megabytes.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Ejemplo de asignación de excepción de registros duplicados

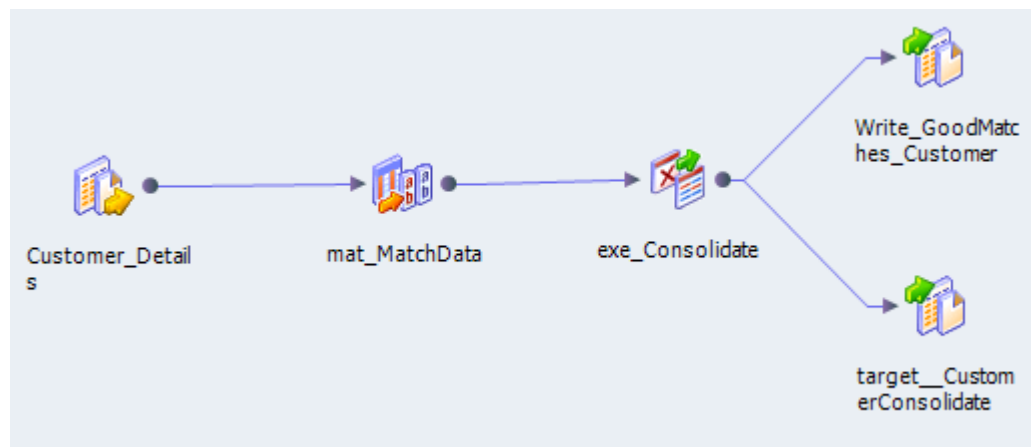
Una organización ejecuta un proyecto de datos para revisar los datos de clientes. La organización determina que los datos de clientes contienen varios registros duplicados. La organización necesita que se revisen manualmente algunos de los registros que pueden estar duplicados.

Cree una asignación de calidad de datos para identificar los registros de clientes duplicados. La asignación incluye la transformación de coincidencia. La transformación de excepción de registros duplicados recibe los resultados de la transformación de coincidencia. La transformación de excepción de registros duplicados escribe cada clúster del registro con un estado incierto en una tabla de base de datos. Un analista de datos revisa los datos en la herramienta del analista y determina los registros que son duplicados.

Asignación de excepción de registros duplicados

Configure una asignación de excepción de registros duplicados que examine los registros de los clientes y encuentra registros duplicados.

La siguiente figura muestra la asignación de excepción de registros duplicados:



La asignación contiene los siguientes objetos:

Customer_Details

El origen de datos que pueden contener los registros duplicados.

mat_MatchData

Una transformación de coincidencia que examina los datos de los clientes para determinar si los registros coinciden. La transformación de coincidencia crea una puntuación numérica que representa el grado de similitud entre dos valores de columna. Un algoritmo calcula una puntuación de coincidencia

como un valor decimal dentro del intervalo de 0 a 1. Un algoritmo asigna una puntuación de uno cuando dos valores de columna son idénticos.

exc_Consolidate

Una transformación de excepción de registros duplicados que determina qué registros son posibles clientes duplicados, clientes duplicados conocidos o registros de clientes únicos.

Tabla Write_GoodMatches_Customer

La tabla que recibe todos los registros que no requieren revisión manual. La transformación de excepción de registros duplicados escribe los registros duplicados y los registros únicos en esta tabla.

Tabla Target_CustomerConsolidate

La transformación de excepción escribe los posibles registros duplicados para la tabla Target_CustomerConsolidate. Los registros en esta tabla se deben revisar de forma manual en la herramienta del analista.

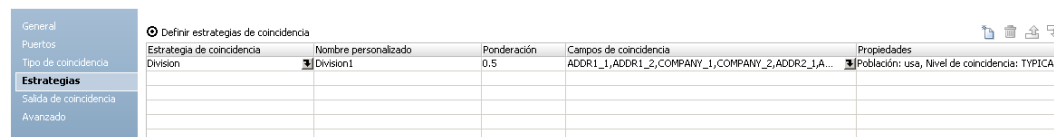
Transformación de coincidencia

La transformación de coincidencia recibe los datos de los clientes y realiza una coincidencia de identidad.

Configure la transformación de coincidencia para el tipo de salida Clústeres - Coincidir con todos. La transformación de coincidencia devuelve los registros de coincidencia de clústeres. Cada registro de un clúster debe coincidir con al menos otro registro del clúster con una puntuación mayor o igual que el umbral de coincidencia. El umbral de coincidencia es 0.75.

Seleccione la estrategia de coincidencia de divisiones en la ficha **Estrategias** de la transformación de coincidencia. La estrategia de divisiones es una estrategia de coincidencia predefinida que identifica una organización en función de los campos de dirección. En la ficha **Estrategias** de la transformación de coincidencia, elija los puertos de entrada que desea examinar en una coincidencia. Configure la ponderación de estrategia como 0,5.

La siguiente figura muestra la configuración de la estrategia de división para la transformación de coincidencia:



Estrategia de coincidencia	Nombre personalizado	Ponderación	Campos de coincidencia	Propiedades
Division	Division1	0.5	ADDR1_1,ADDR1_2,COMPANY_1,COMPANY_2,ADDR2_1A...	Población: usa, Nivel de coincidencia: TYPICAL

La transformación de coincidencia añade la información de los clústeres a cada registro de salida. La transformación también añade un RowID único a cada registro.

Grupos de entrada de excepción de registros duplicados

La transformación de excepción de registros duplicados tiene dos grupos de entrada. La transformación tiene un grupo de datos que recibe los datos de los clientes. La transformación tiene el grupo de controles que contiene la puntuación de coincidencia para la fila, el identificador de fila, y el ID de clúster.

La siguiente figura muestra los grupos de entrada de la transformación de excepción:

	Nombre	Tipo	Precisión	Escala	Entrada	Salida	Predeterminado	Descripción
[-] Entradas								
[-] Datos (11)								
1	..CUST_ID	decimal	20	0	▼	■		
2	..COMPANY	string	200	0	▼	■		
3	..CONTACT	string	200	0	▼	■		
4	..TITLE	string	200	0	▼	■		
5	..ADDR1	string	200	0	▼	■		
6	..ADDR2	string	100	0	▼	■		
7	..ADDR3	string	100	0	▼	■		
8	..ADDR4	string	50	0	▼	■		
9	..COUNTRY	string	50	0	▼	■		
10	..PHONE	string	100	0	▼	■		
11	..EMAIL	string	100	0	▼	■		
[-] Control (3)								
1	..Score	double	15	0	▼	■		
2	..Row_Identifier	string	25	0	▼	■		
3	..Cluster_ID	integer	10	0	▼	■		

El grupo de datos contiene los datos de los clientes. Los datos de los clientes incluyen el ID del cliente, el contacto, la actividad profesional y los campos de dirección. El grupo de controles contiene los metadatos adicionales que la transformación de coincidencia ha añadido para cada registro del cliente. El grupo de controles contiene la puntuación de coincidencia, el ID de fila y el ID de clúster.

Vista de configuración de ejemplo de excepción de registros duplicados

Defina los umbrales superior e inferior en la vista **Configuración**. Identifique dónde escribe la transformación los registros de clientes duplicados, los posibles registros duplicados y los registros de clientes únicos.

La siguiente figura muestra la vista **Configuración** de la transformación de excepción de registros duplicados:

Umbrales de revisión manual

Umbral inferior :

Umbral superior :

Opciones de distribución de datos

Tipo	Salida estándar	Tabla de registro duplicado
Consolidación automática (sobre el umbral superior)	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Consolidación manual (entre umbrales)	<input type="checkbox"/>	<input type="checkbox"/>
Registros únicos (por debajo del umbral inferior)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Crear un grupo de salida aparte para los registros únicos

La siguiente tabla describe las opciones de configuración:

Opción	Valor
Umbral inferior	.80
Umbral superior	.95
Consolidación automática	Tabla de salida estándar
Consolidación manual	Tabla de registros duplicados
Registros únicos	Tabla de salida estándar

Haga clic en **Generar tabla de registros duplicados** para crear la tabla de registros duplicados. No cree una tabla independiente para los registros únicos. La transformación escribe los registros únicos en la tabla de salida estándar.

Registros de tabla de salida estándar

La tabla de destino Write_GoodMatches_Customer recibe filas desde el grupo Salida estándar. La tabla recibe los registros únicos y los registros que son duplicados. Estos registros no requieren de revisión manual.

La siguiente figura muestra los registros de la salida estándar que la transformación de excepción devuelve:

Salida										
Nombre: exc Consoli.Good Records										
	Score	Row_Identifier	Cluster_ID	CUST_ID	COMPANY	CONTACT	TITLE	ADDR1	ADDR2	ADDR3
1	0,9166666666...	1 - 345	260	10111568	SUSQUEHAN...	BALA CYNWYD	United Parcel Serv...	"401 CITY LIN..."		"401 CITY LIN..."
2	0,9166666666...	1 - 835	1174	15951678	"ENOW	INC."	LOS ANGELES	US Postal Int...	"10435 SA..."	2ND"
3	0,9166666666...	1 - 440	260	10111949	TULLETT & T...	** SYSTEM TE...	Federal Express 2-...	** SYSTEM TE...	"200 OLD C..."	STE 215S"
4	0,9166666666...	1 - 487	260	10112102	NATIONAL Q...	NEW YORK	Federal Express 2-...	11 PENN PLA...	"200 OLD C..."	STE 215S"
5	0,9166666666...	1 - 533	260	10112292	PHELPS DOD...	TEMPE	Federal Express 2-...	1501 W. FOU...	"200 OLD C..."	STE 215S"
6	0,9583333333...	1 - 580	260	10112460	GLEN DOSHAY	RANCHO SA...	Federal Express 2-...	6279 VIA CA...	"200 OLD C..."	STE 215S"
7	0,9166666666...	1 - 627	260	10112646	HANI QATTAN	HOUSTON	United Parcel Serv...	5404 CEDAR...	"200 OLD C..."	STE 215S"
8	0,9166666666...	1 - 674	260	10112817	CENTRAL COL...	WESTERVILLE	Federal Express 2-...	975 S. SUNB...	"200 OLD C..."	STE 215S"
9	0,9583333333...	1 - 721	1174	15951251	EVERETT SEC...	MINNEAPOLIS	US Postal Internat...	2930 NORWE...	"200 OLD C..."	STE 215S"
10	0,9166666666...	1 - 768	1174	15951449	ROBERT G PE...	NEW YORK	Federal Express 2-...	86 TRINITY P...	"200 OLD C..."	STE 215S"
11	0,9166666666...	1 - 815	1174	15951607	YMCA RETIRE...	NEW YORK	United Parcel Serv...	140 BROAD...	"200 OLD C..."	STE 215S"
12	0,9166666666...	1 - 234	260	10111114	"NIPPON OIL...	LTD."	SAN RAMON	United Parcel...	"2680 BISH..."	SUITE 275"
13	0,9047619047...	1 - 10	4	7121569	RICE VOELKER	MANDEVILLE	Federal Express 2-...	3840 HIGHW...	"2ND & D S..."	SW HOUSE...
14	0,9166666666...	1 - 432	260	10111932	LA CURRENC...	LOS ANGELES	Federal Express 2-...	BRADLEY INT...	"330 CONG..."	6TH FL"
15	0,9166666666...	1 - 572	260	10112429	NAEGELE AN...	CHICAGO	Federal Express 2-...	401 S LASAL...	"330 CONG..."	6TH FL"
16	0,9166666666...	1 - 619	260	10112618	BANQUE FRA...	HOUSTON	United Parcel Serv...	333 CLAY ST...	"330 CONG..."	6TH FL"
17	0,9166666666...	1 - 666	260	10112777	UNIVERSAL S...	UNIVERSAL CI...	Federal Express 2-...	100 UNIVER...	"330 CONG..."	6TH FL"
18	0,9166666666...	1 - 807	1174	15951579	CREDIT LYON...	NEW YORK	US Postal 2-day	180 MAIDEN...	"330 CONG..."	6TH FL"
19	0,9583333333...	1 - 201	260	10110942	"WEST MAUI...	LLC"	LAHAINA	US Postal 2-d...	"5095 NAPI..."	SUITE 202 N..
20	1	1 - 3	16	1001604	KPMG PEAT...	WOODCLIFF L...	US Postal 2-day	530 CHESNU...	"50 BROAD..."	20TH FL"
21	0,9166666666...	1 - 456	260	10111994	"CREDIT AGRI...	INC."	NEW YORK	Federal Expre...	"520 MADI..."	3RD FLOOR"
22	0,9166666666...	1 - 238	260	10111126	"SMITH BARN...	INC."	CLAYTON	United Parcel...	"7701 FORS..."	12TH FL"
23	0,9166666666...	1 - 526	260	10112268	"BANK OF NE...	LLC"	NEW YORK	US Postal Ov...	"787 SEVEN..."	34TH FLR."

El registro contiene los siguientes campos:

Puntuación

Puntuación de coincidencia que indica el grado de similitud entre un registro y otro registro del clúster. Los registros con una puntuación de coincidencia de uno son registros duplicados que no necesitan revisarse. Un clúster en el que cualquier registro tiene una puntuación de coincidencia por debajo del umbral inferior no es un clúster duplicado.

Row_Identifier

Un número de fila que identifica de forma exclusiva cada fila en la tabla. Para este ejemplo, el identificador de fila contiene el ID de cliente.

Cluster_ID

Un identificador exclusivo para un clúster. Cada registro en un clúster recibe el mismo ID de clúster. Los primeros cuatro registros de los datos de salida de ejemplo son únicos. Cada registro tiene un ID de clúster único. Las filas entre la cinco y la nueve pertenecen al clúster cinco. Cada registro en este clúster es un duplicado registro debido a las similitudes en los campos de dirección.

Campos de datos de origen

El grupo de la tabla de salida estándar también recibe todos los campos de datos de origen.

Salida del clúster

La tabla Target_CustomerConsolidate recibe los registros del grupo de salida de clúster. El grupo de salida del clúster devuelve los registros que pueden ser de registros duplicados. Los registros de la tabla Target_CustomerConsolidate se deben revisar de forma manual en la herramienta del analista.

La siguiente imagen muestra algunos de los registros y campos de la tabla destino Target_CustomerConsolidate:

Salida												
Nombre: exc Consoli.Cluster Data												
	Row_Identifier	Sequential_Cluster_ID	Cluster_ID	Score	Is_Master	Workflow_ID	CUST_ID	COMPANY	CONTACT	TITLE	ADDR1	ADDR2
1	0	0	3	1	Y	DummyWorkflowID	1001622	INVESTEX	NEW YORK		"50 BROAD STREET	20TH FL"
2	1	0	3	1	N	DummyWorkflowID	1001622	INVESTEX	NEW YORK		"50 BROAD STREET	20TH FL"
3	2	1	4	1	Y	DummyWorkflowID	7121564		INC.*		Federal Express 2-day	
4	3	1	4	1	N	DummyWorkflowID	7121564		INC.*		Federal Express 2-day	
5	4	1	4		N	DummyWorkflowID	7121565		NEW YORK		767 FIFTH AVENUE	
6	5	1	4		N	DummyWorkflowID	7121566	K2 ADVISORS	NEW YORK	Courier		
7	6	1	4		N	DummyWorkflowID	7121567	MFP INVESTORS	SHORT HILLS		51 JFK PARKWAY	
8	7	1	4		N	DummyWorkflowID	7121568	EASTON	FORT LEE		ONE PARKER PLAZA	
9	8	1	4		N	DummyWorkflowID	7121569	RICE VOELKER	MANDEVILLE			
10	9	2	16	1	Y	DummyWorkflowID	1001590	E-AGENCY	OAKLAND		291 3RD STREET	
11	10	2	16	1	N	DummyWorkflowID	1001590	E-AGENCY	OAKLAND		291 3RD STREET	
12	11	2	16		N	DummyWorkflowID	1001599	BANK ONE	ELGIN			
13	12	2	16	1	N	DummyWorkflowID	1001604			US Postal 2-day		
14	13	2	16	1	N	DummyWorkflowID	1001658		WASHINGTON		"2ND & D STREET	
15	14	2	16		N	DummyWorkflowID	1001660		ENGLEWOOD		6803 S. TUCSON WAY	
16	15	2	16		N	DummyWorkflowID	1001659		FT. WORTH	Courier	301 COMMERCE ST.	
17	16	2	16	1	N	DummyWorkflowID	1001658		NEW YORK		875 3RD AVE.	
18	17	2	16		N	DummyWorkflowID	1001691	EPICENTRIC		US Postal 2-day	333 BRYANT STREET	
19	18	2	16		N	DummyWorkflowID	1001664		SAN ANTONIO		1100 N.E. LOOP 410	
20	19	2	16		N	DummyWorkflowID	1001694		BOSTON	US Postal 2-day	"125 BROAD STREET	
21	20	2	16	1	N	DummyWorkflowID	1001729					
22	21	2	16		N	DummyWorkflowID	1001724		NEW YORK			
23	22	2	16		N	DummyWorkflowID	1001732			Courier		

El registro contiene los siguientes campos:

Row_Identifier

Un número que identifica de forma exclusiva cada fila de la tabla.

Sequential_Cluster_ID

Un identificador secuencial de cada clúster para revisar en una tarea de humana. La transformación de excepción de registros duplicados añade el ID secuencial de clúster a los registros del grupo de salida de datos del clúster.

Cluster_ID

Un identificador exclusivo para un clúster. La transformación de coincidencia asigna un ID de clúster a todos los registros de salida. Los registros duplicados y los posibles registros duplicados comparten un ID de clúster. Un registro único recibe un ID del clúster, pero el registro no comparte el número de ID con ningún otro registro.

Puntuación

Puntuación de coincidencia que indica el grado de similitud entre un registro y otro registro del clúster. Los registros que requieren revisión manual tienen puntuaciones de coincidencia inferiores que .95 y mayores que .80.

Is_Master

Indica si el registro es el registro preferido del clúster.

WorkflowID

El identificador de flujo de trabajo es DummyWorkflowID porque la transformación no está en un flujo de trabajo.

Campos de registros

Los demás campos del registro contienen los datos de origen.

Creación de una transformación de excepción de registros duplicados

Cuando configure una transformación de excepción de registros duplicados, configure los puertos de entrada. Defina los umbrales superior e inferior para determinar las coincidencias. Configure dónde escribir los registros duplicados y los registros únicos.

1. Cree una transformación de excepción de registros duplicados reutilizable o no reutilizable.
 - Para crear una transformación reutilizable, haga clic en **Archivo > Nuevo > Transformación** y seleccione una transformación de excepción de registros duplicados.
 - Para crear una transformación no reutilizable, abra una asignación y añada la transformación en el lienzo de la asignación. Seleccione una transformación de excepción de registros duplicados en el asistente.
2. Haga clic en **Siguiente** o en **Finalizar**.

Si hace clic en **Finalizar**, puede actualizar los valores de umbral predeterminados y las opciones de distribución de datos antes de crear la transformación.
3. En la vista Configuración, configure los umbrales de puntuación de coincidencia superior e inferior.
4. En la sección **Opciones de distribución de datos**, configure las propiedades estándar de la salida y de la tabla de excepciones para establecer dónde la transformación escribirá cada tipo de registro.

También puede modificar dónde escribir los registros duplicados, registros duplicados para revisar y registros únicos.
5. También puede generar una tabla de registros únicos. Introduzca la conexión de base de datos y la información del nombre de tabla para la nueva tabla. Si genera una tabla de registros únicos, la transformación crea un objeto de base de datos en el repositorio de modelos.
6. Configure los puertos de entrada. Cuando se añade un puerto de entrada, Developer tool añade el mismo nombre de puerto a los grupos de salida.
 - Si crea una transformación reutilizable, seleccione la ficha **Puertos** y añada puertos para los datos que desea conectar a la transformación.
 - Si crea una transformación no reutilizable, añada otros objetos al lienzo de la asignación y arrastre los puertos de entrada para la transformación.
7. Conecte los puertos de salida de la transformación a uno o varios destinos de datos. Conecte los puertos de salida a los objetos de datos que correspondan a las opciones de salida que ha establecido en la vista **Configuración**.
 - Si crea una transformación reutilizable, añada la transformación a una asignación y conecte los puertos de salida.
 - Si crea una transformación no reutilizable, la transformación conecta los puertos a la tabla de datos del clúster. Debe conectar los puertos de salida al resto de los destinos de datos.

CAPÍTULO 16

Transformación de expresión

Este capítulo incluye los siguientes temas:

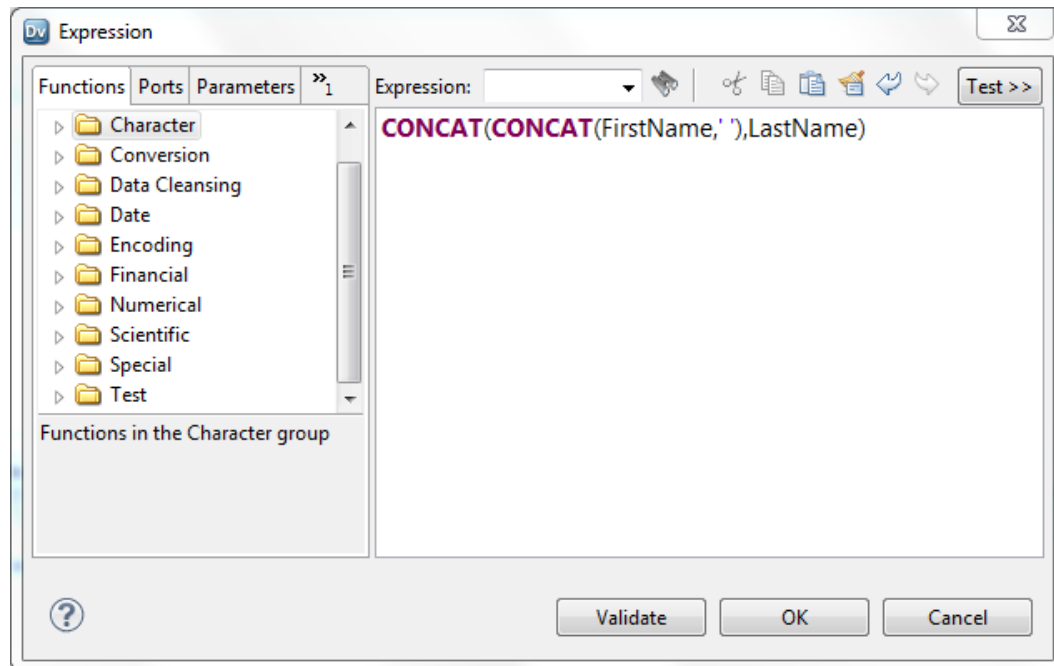
- [Resumen de la transformación de expresión, 287](#)
- [Puertos de transformación de expresión, 288](#)
- [Probar expresiones, 289](#)
- [Selectores de puerto, 291](#)
- [Administración de ventanas, 294](#)
- [Expresiones dinámicas, 299](#)
- [Aplanar una estructura dinámica, 303](#)
- [Propiedades avanzadas de la transformación de expresión, 304](#)
- [Transformación de expresión en un entorno no nativo, 305](#)

Resumen de la transformación de expresión

La transformación de expresión es una transformación pasiva que sirve para realizar cálculos o para probar instrucciones condicionales en una fila. En las transformaciones de expresión no reutilizables, puede definir una expresión de salida de asignación para agregarla cuando defina salidas de asignación.

En una sola fila, quizás necesite crear una expresión para corregir salarios de los empleados, concatenar nombres y apellidos o convertir cadenas en números.

La siguiente figura muestra una expresión en una transformación de expresión que concatena el nombre, un espacio y el apellido:



Si quiere especificar varias expresiones en una única transformación de expresión individual, puede crear una expresión para cada puerto de salida. Por ejemplo, quizá quiera calcular distintos tipos de impuestos en cada nómina de empleado, como los impuestos de renta locales y federales. Para realizar ambos cálculos impositivos, se necesitan el salario y la tasa de impuesto del empleado. Defina un puerto de salida aparte para cada cálculo. Defina una expresión diferente para cada puerto de salida. Puede definir puertos de transferencia para el salario y la tasa de impuesto, ya que los valores de puerto no cambian.

Puertos de transformación de expresión

Una transformación de expresión tiene distintos tipos de puerto a los que se puede hacer referencia al definir expresiones.

Una transformación de expresión tiene los siguientes tipos de puerto:

Entrada

Recibe los datos de las transformaciones de nivel superior. Si la transformación de expresión no cambia el valor de puerto, puede definir un puerto de transferencia en lugar de un puerto de entrada.

Salida

Contiene el valor de devolución de la expresión. La expresión se introduce como una opción de configuración del puerto de salida. También puede configurar un valor predeterminado para cada puerto.

Nota: Si una expresión genera errores numéricos, como una división entre cero o una raíz cuadrada de un número negativo, se devuelve un valor infinito o NaN.

Transferencia

Define un puerto de transferencia para pasar datos a través de la transformación sin modificar el valor. Puede hacer referencia a un puerto de transferencia en un cálculo, pero no cambiar el valor de los datos del puerto de transferencia.

Variable

Almacena temporalmente los datos que se van a utilizar en las expresiones. Puede almacenar datos en varias filas. Puede definir una expresión para devolver un valor a un puerto de variable.

Puerto dinámico

Recibe o devuelve puertos en una asignación dinámica. Un puerto dinámico puede recibir una o varias columnas de una transformación de nivel superior y crear un puerto generado para cada columna. Un puerto de salida dinámico puede devolver uno o varios puertos generados. Puede definir reglas de entrada para determinar qué columnas puede recibir un puerto dinámico. Un puerto de salida dinámico puede contener una expresión que genere varios puertos de salida.

Puerto generado

Puerto que representa una única columna dentro de un puerto dinámico. Los puertos generados en la transformación de expresión pueden cambiar en función de las columnas que la transformación de expresión reciba de una transformación de nivel superior.

Probar expresiones

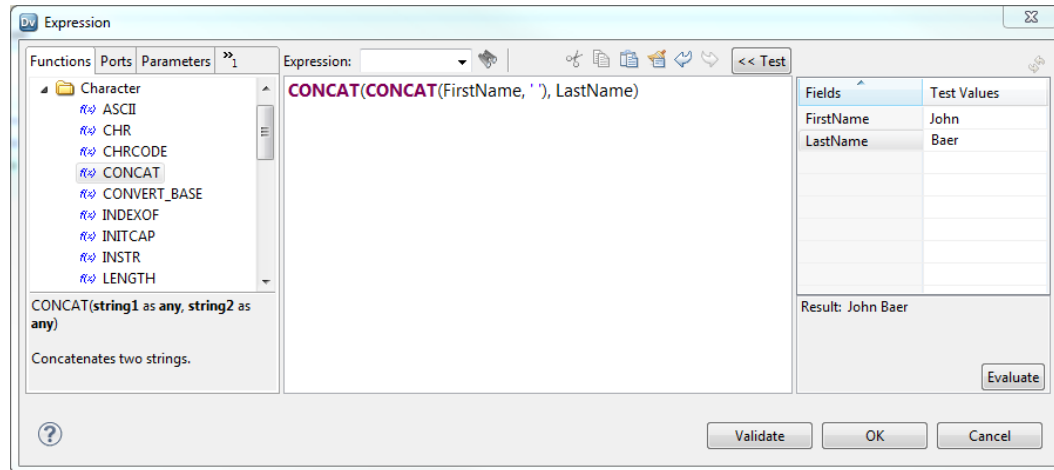
Puede probar las expresiones que tenga configuradas en el editor de expresiones. Para probar una expresión, hay que introducir datos de ejemplo y, a continuación, evaluar la expresión.

Al configurar expresiones, las puede probar de las siguientes formas:

- En un puerto de salida o de variable en la transformación de expresión
- En la vista Salidas de asignación de una transformación de expresión, después de añadir la transformación a una asignación

Por ejemplo, después de configurar una expresión que concatena el nombre, un espacio y el apellido, puede introducir datos de ejemplo de los puertos y, a continuación, evaluar la expresión para comprobar el resultado.

La siguiente imagen muestra los resultados de una expresión que concatena un nombre y un apellido de ejemplo:



Cadenas de formato de fecha para datos de ejemplo

Al probar una expresión que utiliza un puerto con el tipo de datos Fecha/Hora o Marca de tiempo con zona horaria, debe introducir datos de ejemplo para el puerto utilizando la cadena de formato de fecha correspondiente.

Para introducir datos de ejemplo para un puerto con el tipo de datos Fecha/Hora, utilice el formato MM/DD/YYYY HH24:MI:SS. Cuando la expresión se evalúe, el editor de expresiones mostrará el resultado con el formato que especifique en la expresión. Si omite la cadena de formato en la expresión, el editor de expresiones mostrará el resultado con el mismo formato, MM/DD/YYYY HH24:MI:SS.

Para introducir datos de ejemplo para un puerto con el tipo de datos Marca de tiempo con zona horaria, utilice el formato MM/DD/YYYY HH24:MI:SS TZR. Cuando la expresión se evalúe, el editor de expresiones mostrará el resultado con el formato YYYY-MM-DD HH24:MI:SS.NS TZR.

Probar una expresión

Pruebe una expresión en el editor de expresiones para evaluarla y comprobar los resultados.

1. Abra el editor de expresiones de una de las maneras siguientes:
 - En una transformación de expresión, haga clic en el botón **Abrir** (🔽) en la columna **Expresión** de un puerto de salida o de variable.
 - Seleccione una transformación de expresión incluida en una asignación. En la vista **Salidas de asignación**, haga clic en el botón **Abrir** (🔽) en la columna **Expresión** de una salida.
2. Configure la expresión.
3. Haga clic en **Probar >>** para abrir el panel de prueba.
4. Introduzca los datos de ejemplo de cada campo de la columna **Valores de prueba**.
Puede especificar valores de prueba para cada puerto o parámetro incluido en la expresión.
5. Haga clic en **Evaluar**.
El resultado de la expresión se muestra en la parte inferior del panel de prueba.

Selectores de puerto

Cuando una transformación tiene puertos generados, es necesario configurarla para que se ejecute correctamente cuando los puertos generados cambien. Puede utilizar un selector de puerto para saber qué puertos hay que usar en una expresión dinámica, en una condición de búsqueda o en una condición de combinación.

Un selector de puerto es una lista ordenada de puertos a los que puede hacer referencia en una expresión. Cuando los puertos generados cambian en una asignación dinámica, el selector de puerto puede contener diferentes puertos.

Por ejemplo, la siguiente expresión hace referencia a un puerto generado en una asignación dinámica:

```
Salary * 12
```

Configura la asignación para utilizar orígenes dinámicos, pero la columna que contiene la información sobre los salarios en cada archivo de origen tiene un nombre diferente. Los nombres de las columnas son `Salary`, `Monthly_Salary` o `Base_Salary`.

Realiza las siguientes tareas para incluir los diferentes nombres de columna:

1. Crea un selector de puerto llamado "Salary_PortSelector".
2. Crea una regla de selección para aceptar cualquier nombre de puerto con el sufijo "Salary".
3. Configura la expresión para incluir el nombre del selector de puerto en lugar del nombre de columna `Salary`. La expresión contiene la siguiente sintaxis:

```
Salary_PortSelector * 12
```

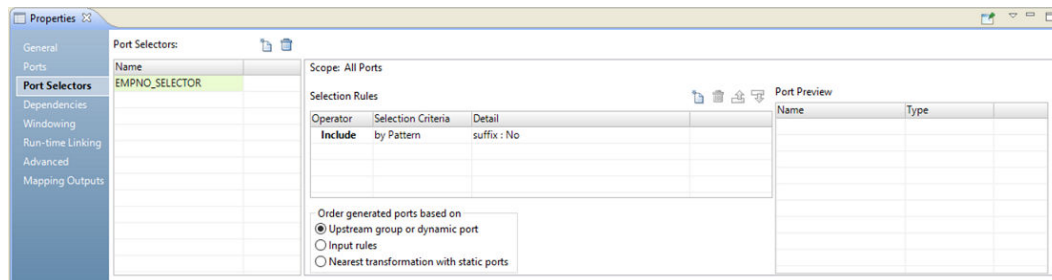
La expresión se ejecutará correctamente con cualquiera de los nombres de puerto `Salary`.

Configuración del selector de puerto

Al configurar un selector de puerto, puede definir reglas de selección para saber qué puertos generados incluir. Las reglas de selección son similares a las reglas de entrada que se configuran para los puertos dinámicos.

Un selector de puerto puede incluir puertos estáticos o puertos generados. Configure un selector de puerto en la ficha **Selector de puerto**.

La siguiente imagen muestra la ficha **Selector de puerto**:



Configure las siguientes propiedades de un selector de puerto:

Nombre

Identifica el selector de puerto. Puede crear varios selectores de puerto en una transformación y hacer referencia a ellos en las expresiones.

Ámbito

Identifica un grupo de puertos al que el selector de puerto se aplica. Al crear un selector de puerto para una transformación de búsqueda o de unión, debe elegir el ámbito. Estas transformaciones tienen varios grupos de entrada. La transformación de unión tiene un ámbito Principal o Detalle. La transformación de búsqueda, por su parte, tiene un ámbito Importar o Búsqueda. La transformación de expresión tiene un grupo de entrada. El ámbito es siempre Todos los puertos.

Reglas de selección

Determina los puertos que se van a incluir en el selector de puerto. Al crear la selección reglas, el panel **Vista previa del puerto** muestra los puertos que cumplen los requisitos de los puertos de entrada actuales. Estos puertos pueden cambiar. Configure las reglas de selección para dar cabida a puertos de diferentes orígenes.

Reglas de selección

Las reglas de selección asociadas con un selector de puerto determinan los puertos a incluir en el selector de puerto.

Al crear la selección reglas, el panel **Vista previa del puerto** muestra los puertos que cumplen los requisitos de los puertos de entrada actuales. Estos puertos pueden cambiar. Configure las reglas de selección para dar cabida a puertos de diferentes orígenes.

Cree reglas de selección basadas en los criterios siguientes:

Operador

Incluye o excluye los puertos que las reglas de selección devuelven. El valor predeterminado es Incluir. Para poder excluir puertos, antes debe incluirlos.

Criterios de selección

Tipo de regla de selección que desea crear Puede crear una regla basada en el nombre de columna, el tipo de puerto, el patrón o la definición del tipo de datos complejos. Para incluir puertos según el nombre de columna, busque nombres específicos o un patrón de caracteres en el nombre.

Detalle

Valores que se van a aplicar a los criterios de selección. Si los criterios de selección están definidos según el nombre de columna, configure la cadena o el nombre por el que realizar la búsqueda. Si los criterios de selección están definidos según el tipo de puerto, seleccione los tipos de puerto que quiera incluir.

En la siguiente tabla se describen los criterios de selección y cómo especificar los detalles de los criterios:

Criterios de selección	Descripción	Detalle
Todo	Incluye todos los puertos.	No se necesitan detalles.
Nombre	Filtra los puertos por el nombre del puerto.	Seleccione los nombres de puerto de una lista de valores o utilice un parámetro de tipo puerto o lista de puertos.
Tipo	Filtra los puertos por el tipo de datos de cada puerto.	Seleccione los tipos de datos de una lista.

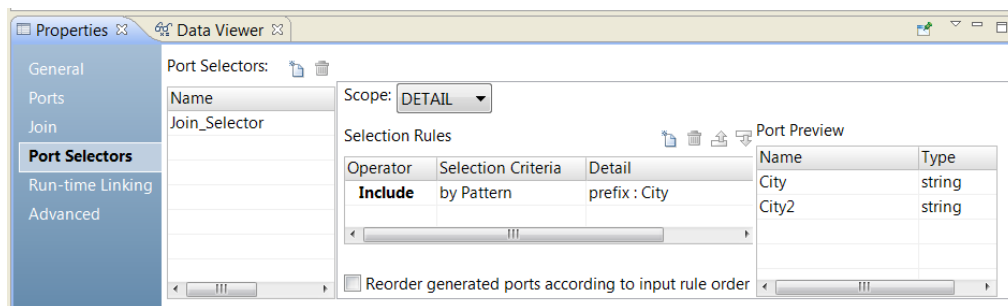
Crterios de seleccin	Descripcin	Detalle
Patr3n	Filtra los puertos por una cadena de caracteres en el nombre o por una expresi3n regular.	Elija Prefijo, Sufijo o Expresi3n regular como tipo de patr3n para el nombre del puerto. Luego, introduzca un valor para el patr3n o utilice un par3metro de tipo cadena.
Definici3n de tipo de datos complejo	Filtra los puertos a partir de una definici3n de tipo de datos complejo.	Elija un prefijo, un sufijo o una expresi3n regular como el tipo de patr3n de la definici3n de tipo de datos complejo. Luego, introduzca un valor para el patr3n o utilice un par3metro de tipo String.

Crear un selector de puerto

Cree un selector de puerto para saber qu3 puertos hay que usar en una expresi3n din3mica, en una condici3n de b3squeda o en una condici3n de combinaci3n.

- Haga clic en la ficha **Selectores de puerto**.
- En el 3rea **Selectores de puerto**, haga clic en **Nuevo**.
Developer tool crea un selector de puerto con una regla de selecci3n predeterminada que incluye todos los puertos.
- En el 3rea **Selectores de puerto**, cambie el nombre del selector de puerto por un nombre 3nico.
- Si est3 trabajando con una transformaci3n de combinaci3n o de b3squeda, seleccione el 3mbito.
Los puertos disponibles var3an seg3n el grupo de puertos que elija.
- En el 3rea **Reglas de selecci3n**, seleccione un **Operador**.
 - Incluir. Cree una regla que incluya puertos para el selector de puerto. Para poder excluir puertos, antes debe incluirlos.
 - Excluir. Cree una regla que excluya puertos espec3ficos del selector de puerto.
- Elija los **Crterios de selecci3n**.
 - Por nombre. Seleccione puertos concretos por su nombre. Puede seleccionar los nombres de puerto de una lista de puertos en el 3mbito.
 - Por tipo. Seleccione puertos por el tipo. Puede seleccionar uno o varios tipos de datos.
 - Por patr3n. Seleccione puertos siguiendo un patr3n de caracteres en el nombre del puerto. Puede realizar las b3squedas con caracteres espec3ficos o crear una expresi3n regular.

La siguiente imagen muestra la ficha Selector de puerto:



- Haga clic en la columna **Detalle**.
Se abre el cuadro de di3logo **Detalles de la regla de entrada**.

8. Seleccione los valores por los que quiera filtrar los puertos.
 - Por nombre. Decida si quiere crear una lista de puertos por valor o por un parámetro. Haga clic en **Elegir** para seleccionar los puertos de la lista.
 - Por tipo. Seleccione uno o varios tipos de datos de una lista. El área **Vista previa del puerto** muestra los puertos de los tipos que ha seleccionado.
 - Por patrón. Decida si quiere hallar un determinado patrón de caracteres en el prefijo o sufijo del nombre del puerto. También puede optar por crear una expresión regular con la que realizar la búsqueda. Configure un parámetro o el patrón de búsqueda.

El área **Vista previa del puerto** muestra los puertos en el selector de puerto a medida que configura las reglas.
9. Para reordenar los puertos en el selector de puerto, seleccione **Reordenar los puertos generados según el orden de regla de entrada**.

Administración de ventanas

Cuando una transformación contiene una función de ventana, es necesario configurar las propiedades de administración de ventanas. La administración de ventanas solo está disponible para las transformaciones en el motor Spark.

Las funciones de ventana operan en un grupo de filas y calculan un valor de devolución para cada fila de entrada.

Antes de definir una función de ventana en una transformación de expresión, debe describir la ventana configurando las propiedades de administración de ventanas. Las propiedades de administración de ventanas incluyen una especificación de marco, claves de partición y claves de ordenación. La especificación de marco indica qué filas se incluyen en el cálculo general para la fila actual. Las claves de partición determinan qué filas se incluyen en la misma partición. Las claves de ordenación determinan cómo se ordenan las filas de una partición.

Tras configurar las propiedades de administración de ventana, defina una función de ventana en la transformación de expresión. Informatica admite las funciones de ventana LEAD y LAG. También puede utilizar funciones de agregado como funciones de ventana en una transformación de expresión.

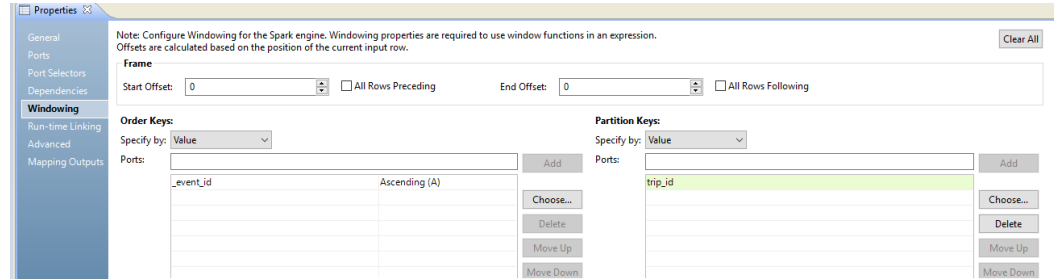
Configuración de la administración de ventanas

Cuando se incluye una función de ventana en una transformación de expresión, se configuran las propiedades de administración de ventanas asociadas a la función. Las propiedades de administración de

ventanas definen las particiones, los ordenamientos y los límites de marco asociados a un fila de entrada determinada.

Configure una transformación para la administración de ventanas en la ficha Administración de ventanas.

En la siguiente imagen se muestra la ficha Administración de ventanas:



Configure los siguientes grupos de propiedades en la ficha Administración de ventanas:

Marco

Define las filas que se incluyen en el marco de la fila de entrada actual, según los desplazamientos físicos de la posición de la fila de entrada actual.

El marco se configura cuando se utiliza una función de agregado como una función de ventana. Las funciones de ventana LEAD y LAG hacen referencia a filas individuales e ignoran la especificación de marco.

Claves de partición

Separe las filas de entrada en diferentes particiones. Si no se definen claves de partición, todas las filas pertenecerán a una sola partición.

Claves de ordenación

Defina cómo se ordenan las filas de una partición. Los puertos que elige determinan la posición de una fila dentro de una partición. La clave de ordenación puede ser ascendente o descendente. Si no se definen claves de ordenación, las filas no tendrán ningún orden en particular.

Marco

El marco determina qué filas se incluirán en el cálculo de la fila de entrada actual, basándose en su posición relativa con respecto a la fila actual.

Si utiliza una función de agregado en lugar de LEAD o LAG, debe especificar un marco de ventana. LEAD y LAG hacen referencia a filas individuales e ignoran la especificación del marco.

El inicio de desplazamiento y el fin de desplazamiento describen el número de filas que aparecen antes y después de la fila de entrada actual. Un desplazamiento de "0" representa la fila de entrada actual. Por ejemplo, un inicio de desplazamiento de -3 y un fin de desplazamiento de 0 describen un marco que incluye la fila de entrada actual y las tres filas anteriores a la fila actual.

En la siguiente imagen se muestra un marco con un inicio de desplazamiento de -1 y un fin de desplazamiento de 1:

Type	Category	Revenue
Action	Video game	1000
Arcade	Video game	1000
Sports	Video game	2000
Adventure	Video game	3000
Strategy	Video game	4000

Current input row →

← 1 PRECEDING

← 1 FOLLOWING

Para cada fila de entrada, la función realiza una operación de agregado en las filas dentro del marco. Si configura una expresión de agregado como SUM con el marco anterior, la expresión calcula la suma de los valores dentro del marco y devuelve un valor de 6000 para la fila de entrada.

También puede especificar un marco que no incluya la fila de entrada actual. Por ejemplo, un inicio de desplazamiento de 10 y un fin de desplazamiento de 15 describe un marco que incluye seis filas en total, desde la décima a la decimoquinta fila después de la fila actual.

Nota: El inicio de desplazamiento debe ser menor o igual que el fin de desplazamiento.

Los desplazamientos de **Todas las filas anteriores** y **Todas las filas siguientes** representan la primera fila de la partición y la última fila de la partición. Por ejemplo, si el inicio de desplazamiento es Todas las filas anteriores y el fin de desplazamiento es -1, el marco incluye una fila antes de la fila actual y todas las filas anteriores.

En la siguiente imagen se muestra un marco con un inicio de desplazamiento de 0 y un fin de desplazamiento de Todas las filas siguientes:

Genre	Recordings	Revenue
Jazz	233	5000
Gospel	214	1000
Country	145	2000
Ethnic	154	9000
Pop	317	4000
Rock	237	2100
Classical	221	3200
EDM	153	950
Hip Hop	839	2300
Punk	415	7650

Current input row →

All Rows Following

Claves de partición y ordenación

Configure las claves de partición y ordenación para que se formen grupos de filas y se defina el orden o la secuencia de las filas dentro de cada partición.

Utilice las siguientes claves para especificar cómo se agrupan y se ordenan las filas en una ventana:

Claves de partición

Configure las claves de partición para definir los límites de la partición, en lugar de realizar el cálculo en todas las entradas. La función de ventana opera en las filas que entran dentro de la misma partición que la fila actual.

Puede especificar las claves de partición por valor o parámetro. Seleccione **Valor** para usar los nombres de puerto. Seleccione **Parámetro** para usar un parámetro de lista de clave de ordenación. Un parámetro de lista de clave de ordenación contiene una lista de puertos por los que establecer la ordenación. Si no especifica las claves de partición, todos los datos se incluirán en la misma partición.

Claves de ordenación

Utilice las claves de ordenación para determinar cómo se ordenan las filas de una partición. Las claves de ordenación definen la posición de una fila determinada en una partición.

Puede especificar las claves de ordenación por valor o parámetro. Seleccione **Valor** para usar los nombres de puerto. Seleccione **Parámetro** para usar un parámetro de lista de clave de ordenación. Un parámetro de lista de clave de ordenación contiene una lista de puertos por los que establecer la ordenación. También debe elegir si desea organizar los datos en orden ascendente o descendente. Si no especifica las claves de ordenación, las filas de una partición no se organizarán en ningún orden concreto.

Ejemplo

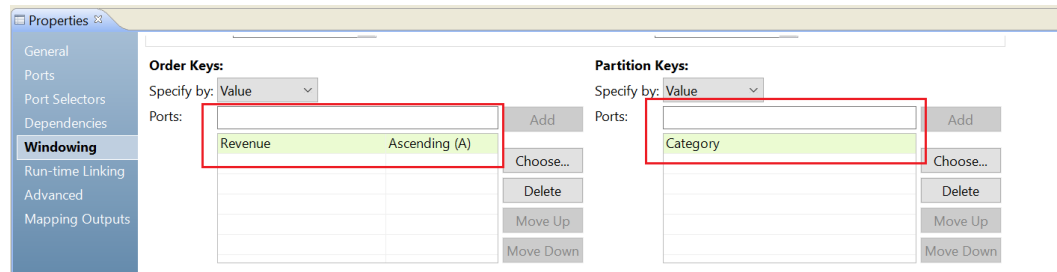
Es el dueño de una cafetería y una tetería. Quiere calcular cuáles son los dos cafés y tés que más se venden.

En la tabla siguiente se enumeran los productos, las categorías de producto correspondientes y los ingresos que provienen de cada producto:

Product	Category	Revenue
Espresso	Coffee	600
Black	Tea	550
Cappuccino	Coffee	500
Americano	Coffee	600
Oolong	Tea	250
Macchiato	Coffee	300
Green	Tea	450
White	Tea	650

Particione los datos por categoría y ordene los datos por ingresos descendentes.

En la siguiente imagen se muestran las propiedades que se configuran en la ficha Administración de ventanas:



En la tabla siguiente se muestran los datos agrupados en dos particiones según la categoría. Dentro de cada partición, los ingresos se organizan en orden descendente:

Product	Category	Revenue
Espresso	Coffee	600
Americano	Coffee	600
Cappuccino	Coffee	500
Macchiato	Coffee	300
White	Tea	650
Black	Tea	550
Green	Tea	450
Oolong	Tea	250

En función de las especificaciones de partición y ordenación, se determina que los dos cafés más vendidos son espresso y americano, y los dos té más vendidos son blanco y negro.

Reglas y directrices para la configuración de la administración de ventanas

Ciertas directrices se aplican cuando se configura una transformación para la administración de ventanas.

Tenga en cuenta las siguientes reglas y directrices cuando define las propiedades de administración de ventanas para una función de ventana:

- Al configurar un marco, el inicio de desplazamiento debe ser inferior o igual al fin de desplazamiento. De lo contrario, el marco no será válido.
- Configure una especificación de marco si utiliza una función de agregado como una función de ventana. LEAD y LAG funcionan basándose en el valor de desplazamiento e ignoran la especificación del marco.
- No es posible usar puertos complejos como claves de partición u ordenación.
- Asigne nombres de puerto exclusivos a las claves de partición y ordenación para evitar errores de tiempo de ejecución.
- Las claves de partición y ordenación no pueden utilizar a la vez un puerto dinámico y uno o varios puertos generados del mismo puerto dinámico. Debe seleccionar el puerto dinámico o los puertos generados.

Expresiones dinámicas

Cuando se configura una expresión en un puerto de salida dinámico, dicha expresión se convierte en una expresión dinámica. Una expresión dinámica puede generar varios puertos de salida.

Puede hacer referencia a un selector de puerto o a un puerto dinámico en una expresión dinámica. Cuando el selector de puerto o el puerto dinámico contiene varios puertos, la expresión dinámica se ejecutará en cada uno de ellos.

Al configurar una expresión dinámica, Developer tool no valida si los puertos generados son de un tipo válido en la expresión. Por ejemplo, si hace referencia a un selector de puerto que contiene puertos de tipo decimal en una expresión que requiere tipos de cadena, la expresión aparecerá como válida en el tiempo de diseño.

Ejemplo

Una transformación de expresión tiene los siguientes puertos de entrada generados:

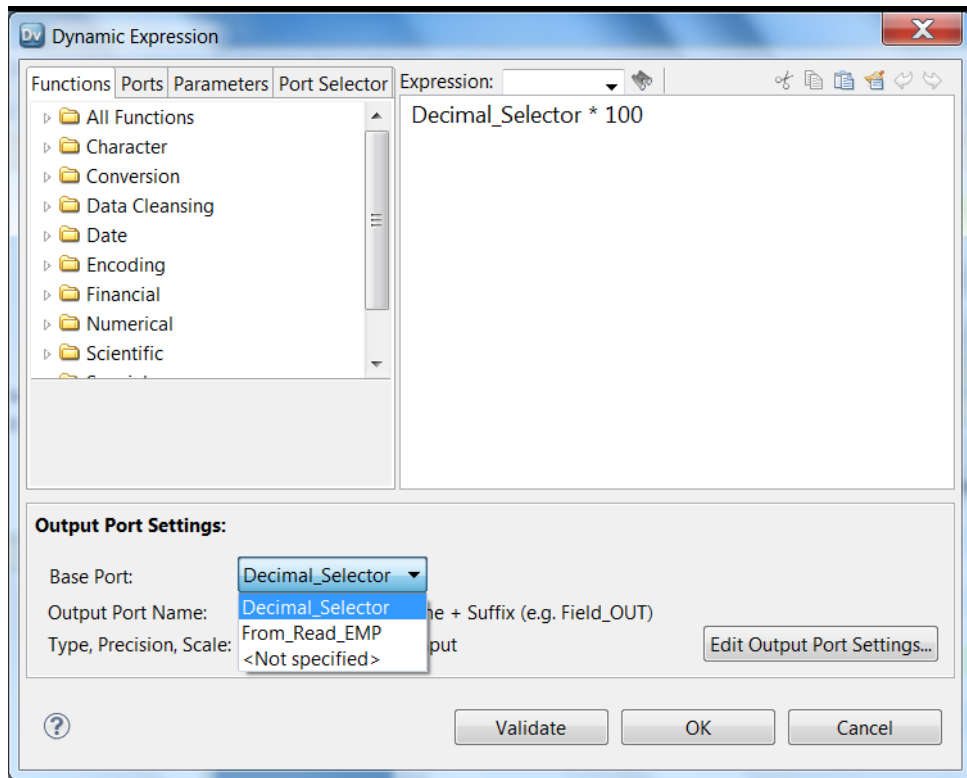
```
EMPNO    Decimal
NAME     String
SALARY   Decimal
DEPTNO   Decimal
```

La transformación contiene un puerto de salida dinámico denominado MyDynamicPort. El puerto de salida devuelve los resultados de una expresión dinámica. La expresión dinámica multiplica por 100 el valor de cada puerto en un selector de puerto. La expresión se ejecuta una vez por cada puerto del selector de puerto. Cada instancia puede devolver un resultado distinto. La transformación de expresión genera un puerto de salida independiente por cada resultado.

El selector de puerto Decimal_Selector tiene una regla de selección que incluye los puertos que sean de tipo de datos decimal:

```
EMPNO    Decimal
SALARY   Decimal
DEPTNO   Decimal
```

La siguiente imagen muestra una expresión dinámica en la que se hace referencia al selector de puerto Decimal_Selector:



Edite la configuración de puerto de salida para cambiar los nombres y las propiedades de los puertos de salida. También puede elegir el puerto base.

Configuración del puerto de salida

Puede indicar qué puertos quiere usar como entrada de una expresión dinámica. Seleccione los puertos en el área **Puerto base**.

Si selecciona el selector de puerto Decimal_Selector como puerto base, la expresión dinámica devuelve puertos de tipo decimal. La expresión dinámica no generará un puerto para el puerto NAME porque es una cadena.

La siguiente imagen muestra los puertos generados de la transformación:

	Name	Type	Precis...	Scale	Input	Output	Varia...	Expression
1	From_Read_EMP	dynamic		0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	From_Read_EMP
1	EMPNO	decimal	3	0				
2	NAME	string	10	0				
3	SALARY	decimal	4	0				
4	DEPTNO	decimal	4	0				
2	MyDynamicPort	dynamic		0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Decimal_Selector * 100
1	EMPNO_OUT	decimal	3	0				
2	SALARY_OUT	decimal	4	0				
3	DEPTNO_OUT	decimal	4	0				

Aunque el puerto dinámico From_Read_Emp es un puerto de entrada-salida, la transformación devolverá solamente los puertos del puerto de salida dinámico MyDynamicPort.

Puede configurar el modo en que quiere denominar a los puertos de salida. El nombre del puerto de salida predeterminado es una combinación del nombre del puerto de entrada y el sufijo `_OUT`.

Puede cambiar el puerto base por un selector de puerto.

La siguiente imagen muestra la configuración de puerto de salida en el editor de expresiones:

Output Port Settings:

Base Port: From_Read_EMP ▾

Output Port Name: Primary input port name + Suffix (e.g. Field_O)

Type, Precision, Scale: Inherit from primary input Edit Output Port Settings...

Si configura el puerto base como `From_Read_EMP`, selecciona el puerto dinámico que contiene todos los puertos de entrada generados. El Servicio de integración de datos ejecuta la expresión dinámica en todos los puertos en `From_Read_EMP`.

La siguiente imagen muestra los puertos de salida generados según la entrada `From_Read_EMP`:

General		Name	Type	Precis...	Scale	Input	Output	Varia...	Expression
Ports	1	From_Read_EMP	dynamic		0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	From_Read_EMP
Port Selectors	1	EMPNO	decimal	3	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Dependencies	2	NAME	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Parameters	3	SALARY	decimal	4	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Run-time Linking	4	DEPTNO	decimal	4	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Advanced	2	MyDynamicPort	dynamic		0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Decimal_Selector * 100
Mapping Outputs	1	EMPNO_OUT	decimal	3	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	2	NAME_OUT	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	3	SALARY_OUT	decimal	4	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	4	DEPTNO_OUT	decimal	4	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Los puertos de salida generados incluyen un puerto de salida denominado `NAME_OUT` que es de tipo cadena.

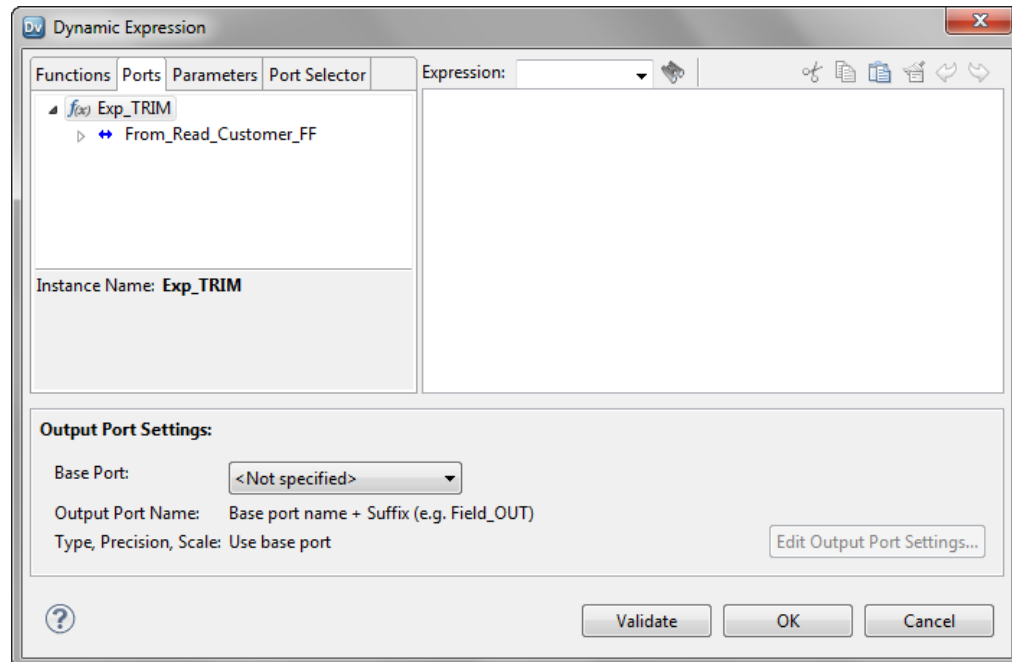
El Servicio de integración de datos genera puertos de salida por cada expresión dinámica. Así, si crea una expresión dinámica que genera 15 puertos y define otra que genera 5 puertos, el Servicio de integración de datos generará 20 puertos de salida en total. Cada puerto de salida dinámico genera un grupo de puertos distinto.

Crear una expresión dinámica

Cree una expresión dinámica en una transformación de expresión para ejecutar la expresión una vez por cada puerto de un puerto dinámico o de un selector de puerto. La expresión dinámica devuelve los resultados en otro puerto generado para cada instancia.

1. En la transformación de expresión, vaya a la vista **Propiedades** y haga clic en la ficha **Puertos**.
2. Haga clic en **Nuevo puerto dinámico**.
Developer tool creará un puerto dinámico con propiedades predeterminadas.
3. Cambie el nombre del puerto dinámico y deshabilite la opción de entrada.
El puerto dinámico debe ser un puerto de salida.
4. En la columna **Expresión** del puerto de salida dinámico, haga clic en el botón **Abrir** (📄).

Se abre el cuadro de diálogo **Expresión dinámica**:



5. En el editor de expresiones, introduzca una expresión. La expresión puede incluir un selector de puerto o un puerto dinámico.
Por ejemplo, `LTRIM(RTRIM(Dynamic_Customer))`, donde `Dynamic_Customer` es un puerto dinámico.
6. Haga clic en **Validar** para validar la expresión.
7. Haga clic en **Aceptar** para salir del cuadro de diálogo **Validar expresión**.
8. En el área **Configuración del puerto de salida**, seleccione el puerto de salida dinámico de la lista **Puerto base** o elija un selector de puerto al que haya hecho referencia en la expresión.
Developer tool genera puertos de salida según lo que seleccione.

9. Realice los siguientes pasos para cambiar el nombre de los puertos de salida:
 - a. Haga clic en **Editar configuración del puerto de salida**.

Se abre el cuadro de diálogo **Configuración del puerto de salida**.

- b. En la lista **Nombre**, seleccione una de las opciones e introduzca un valor de prefijo o sufijo. Si ha seleccionado **Cadena fija + Número automático**, introduzca el texto del nombre del puerto de salida. Por ejemplo, si introduce TRIM como nombre de puerto de salida, los nombres de los puertos de salida serán TRIM1, TRIM2, TRIM3, etc.
 - c. Si lo desea, elija **Especificar la configuración** en el área **Otras configuraciones** para cambiar el tipo, la precisión y la escala de los puertos de salida. Los puertos de salida usan la configuración de los puertos base de forma predeterminada.
 - d. Haga clic en **Aceptar**.
10. Haga clic en **Aceptar** para salir del **Expresión dinámica**.

Aplanar una estructura dinámica

Utilice la función `EXTRACT_STRUCT` para aplanar estructuras dinámicas en una transformación de expresión.

Utilice el operador de punto en expresiones para extraer elementos de una estructura.

Por ejemplo, puede utilizar la función `EXTRACT_STRUCT` para extraer todos los datos de una dirección sin especificar cada campo de dirección individualmente.

	Name	Type	Type Configuration	Precisi...	Scale	Input	Output	Variable	Expression	Default Value	Description	Input Rules
1	AddressInput_RegularStruct	struct	(Address)			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		AddressInput_RegularStruct			
1	Street	string	N/A	10	0							
2	City	string	N/A	10	0							
3	State	string	N/A	10	0							
4	Zip	string	N/A	10	0							
2	AddressInput_DynamicStruct	dynamic struct	N/A			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		AddressInput_DynamicStruct			Include All
1	Street	string	N/A	10	0							
2	City	string	N/A	10	0							
3	State	string	N/A	10	0							
4	Zip	string	N/A	10	0							
3	City_ExtractFromRegularStruct	string	N/A	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		AddressInput_RegularStruct.City		ERROR! transf...	
4	City_ExtractFromDynamicStruct	string	N/A	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>		AddressInput_DynamicStruct.City		ERROR! transf...	
5	ExtractAll_FromDynamicStruct	dynamic	N/A			<input type="checkbox"/>	<input checked="" type="checkbox"/>		EXTRACT_STRUCT(AddressInput_DynamicStruct)		ERROR! transf...	
1	Street_OUT	string	N/A	10	0							
2	City_OUT	string	N/A	10	0							
3	State_OUT	string	N/A	10	0							
4	Zip_OUT	string	N/A	10	0							

Propiedades avanzadas de la transformación de expresión

Configure las propiedades para ayudarle a determinar cómo el Servicio de integración de datos procesa los datos para la transformación de expresión.

Configure las siguientes propiedades avanzadas de una transformación de expresión:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Mantener orden de filas

Mantenga el orden de las filas de los datos de entrada en la transformación. Seleccione esta opción si el Servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas.

Quando el Servicio de integración de datos realiza optimizaciones, puede hacer que se pierda un orden establecido anteriormente en la asignación. Puede establecer el orden en una asignación con un origen de archivo sin formato ordenado, un origen relacional ordenado o una transformación de ordenación. Al configurar una transformación para mantener el orden de las filas, el Servicio de integración de datos tiene en cuenta esta configuración cuando realiza optimizaciones para la asignación. El Servicio de integración de datos realiza optimizaciones para la transformación si puede mantener el orden. El Servicio de integración de datos no realiza optimizaciones para la transformación si el optimización cambia el orden de las filas.

Transformación de expresión en un entorno no nativo

El procesamiento de transformaciones de expresión en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de expresión en el motor de Blaze

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Una transformación de expresión con una función definida por el usuario devuelve un valor nulo en las filas que tienen un error de excepción en la función.

Transformación de expresión en el motor de Spark

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Nota: Si una expresión genera errores numéricos, como una división entre cero o una raíz cuadrada de un número negativo, se devuelve un valor nulo y las filas no aparecen en la salida. En el entorno nativo, la expresión devuelve un valor infinito o NaN.

Transformación de expresión en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

No puede usar la función `EXTRACT_STRUCT` para asignaciones de transmisión.

Transformación de expresión en el motor de Databricks Spark

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene puertos variables con estado.
- La transformación contiene funciones no compatibles en una expresión.

Nota: Si una expresión genera errores numéricos, como una división entre cero o una raíz cuadrada de un número negativo, se devuelve un valor nulo y las filas no aparecen en la salida. En el entorno nativo, la expresión devuelve un valor infinito o NaN.

CAPÍTULO 17

Transformación de filtro

Este capítulo incluye los siguientes temas:

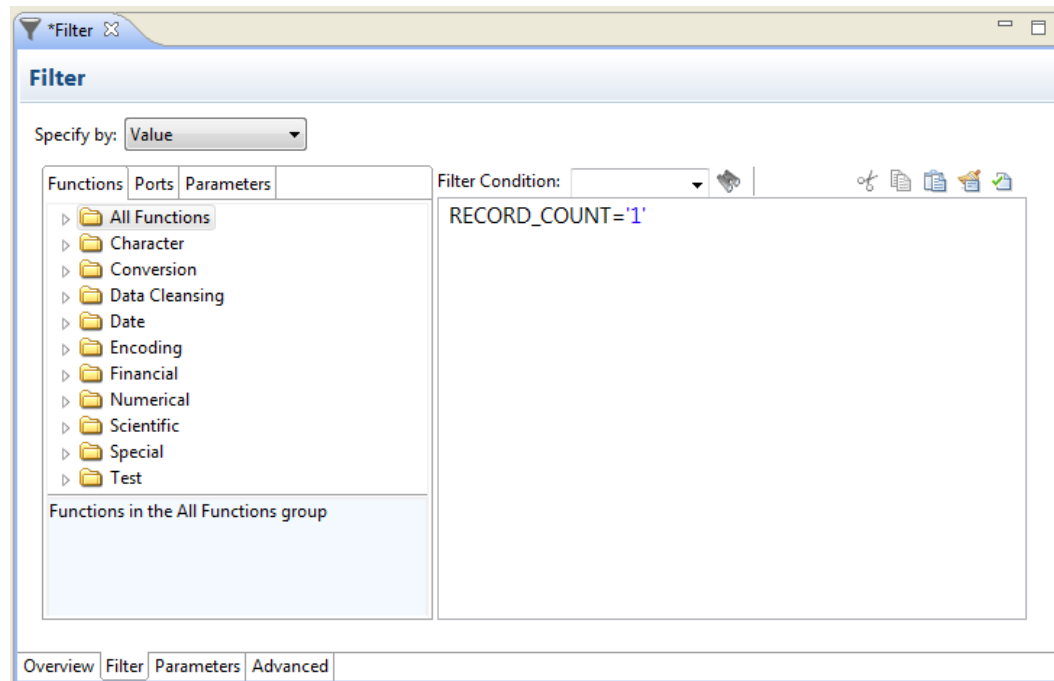
- [Resumen de la transformación de filtro, 306](#)
- [Transformaciones de filtro en asignaciones dinámicas, 307](#)
- [Condición de filtro, 308](#)
- [Propiedades avanzadas de la transformación de filtro, 310](#)
- [Consejos sobre el rendimiento de la transformación de filtro, 310](#)
- [Transformación de filtro en un entorno no nativo, 310](#)

Resumen de la transformación de filtro

Utilice la transformación de filtro para filtrar filas de una asignación. Como transformación activa, la transformación de filtro puede cambiar el número de filas que pasan por ella.

La transformación de filtro permite que pasen las filas que cumplen la condición de filtro especificada. Suelta las filas que no cumplen la condición. Se pueden filtrar los datos según una o más condiciones.

La siguiente imagen muestra una condición de filtro en una transformación de filtro:



Una condición de filtro devuelve TRUE o FALSE para cada fila que evalúa el Servicio de integración de datos si una fila cumple o no la condición especificada. Para cada fila que devuelve TRUE, el Servicio de integración de datos pasa por la transformación. Para cada fila que devuelve FALSE, el Servicio de integración de datos se suelta y escribe un mensaje en el registro.

No se pueden concatenar puertos de más de una transformación en la transformación de filtro. Los puertos de entrada para el filtro deben proceder de una transformación individual.

Transformaciones de filtro en asignaciones dinámicas

Puede utilizar una transformación de filtro en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados en la condición de filtro.

Puede parametrizar la condición de filtro completa. Configure un parámetro de expresión con un valor predeterminado que contenga la expresión completa. Developer tool no valida una condición de filtro en un valor predeterminado de parámetro.

Puede hacer referencia a un puerto dinámico en una condición de filtro. El puerto dinámico puede contener varios puertos generados. El Servicio de integración de datos expande la condición de filtro para contener cada puerto generado. Cada puerto generado debe ser de un tipo válido para incluirlo en la expresión.

Puede hacer referencia a un puerto generado en una condición de filtro. Sin embargo, si el puerto generado no existe en tiempo de ejecución, la asignación fallará.

Condición de filtro

La condición de filtro es una expresión que devuelve TRUE o FALSE.

Especifique condiciones en el editor de expresiones. La condición de filtro distingue entre mayúsculas y minúsculas.

Puede utilizar como filtro cualquier expresión que devuelve un valor individual. Por ejemplo, si desea filtrar las filas de los empleados cuyo salario es inferior o igual a 30 000 USD, especifique la siguiente condición:

```
SALARY > 30000
```

Puede especificar varios componentes para la condición utilizando los operadores lógicos AND y OR. Si desea filtrar los empleados que ganan menos de 30 000 USD y más de 100 000 USD, especifique la siguiente condición:

```
SALARY > 30000 AND SALARY < 100000
```

Puede utilizar puertos, parámetros, puertos dinámicos y puertos generados en la condición de filtro. Seleccione los puertos y los parámetros en el editor de expresiones.

Si utiliza un puerto dinámico en la condición de filtro, dicha condición se expande para incluir todos los puertos generados en el puerto dinámico. Por ejemplo, el puerto dinámico, MyDynamicPort, contiene tres puertos de decimales:

```
Salary  
Bonus  
Stock
```

Si configura la siguiente condición de filtro:

```
MyDynamicPort > 100
```

La condición de filtro se expande a la siguiente expresión:

```
Salary > 100 AND Bonus > 100 AND Stock > 100
```

Puede especificar una constante para la condición de filtro. El equivalente numérico de FALSE es cero (0). Todo valor distinto de cero equivale a TRUE. Por ejemplo, la transformación contiene un puerto llamado NUMBER_OF_UNITS con un tipo de datos numérico. Configure una condición de filtro para que devuelva FALSE si el valor de NUMBER_OF_UNITS es igual a cero. De lo contrario, la condición devuelve TRUE.

Nota: No se puede utilizar un único selector de puerto o puerto dinámico como valor booleano.

No necesita especificar TRUE o FALSE como valores en la expresión. TRUE y FALSE son valores de devolución implícitos de cualquier condición que se define. Si la condición de filtro tiene como resultado NULL, la fila será FALSE.

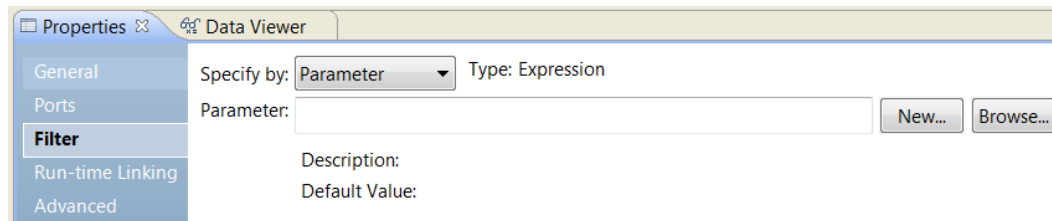
Parametrizar la condición de filtro

Puede configurar un parámetro de expresión para definir la condición de filtro. Un parámetro de expresión contiene la expresión completa.

Puede que sea necesario parametrizar la condición de filtro cuando la transformación de filtro está en una asignación dinámica. La condición de filtro podría cambiar según los puertos generados en la transformación en tiempo de ejecución.

Para utilizar un parámetro de expresión para la condición de filtro, elija **Especificar por parámetro** en la ficha **Filtro** de las propiedades de la transformación de filtro.

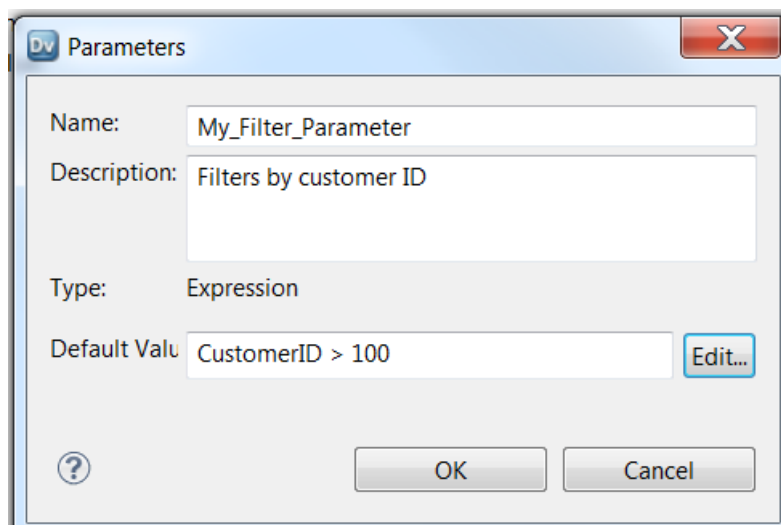
La siguiente imagen muestra la ficha **Filtro** cuando se especifica la condición de filtro con un parámetro:



Puede buscar y seleccionar un parámetro de expresión que ya haya creado. También puede crearlo.

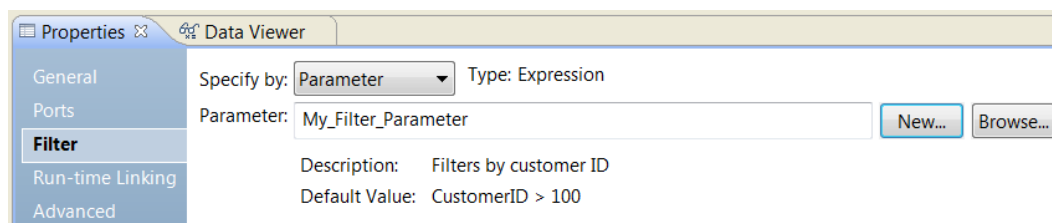
Para crear un parámetro de expresión, haga clic en **Nuevo**. Escriba un nombre para el parámetro, una descripción y el valor predeterminado de la expresión.

La siguiente imagen indica dónde se especifica el parámetro:



Puede introducir una expresión predeterminada en el cuadro de diálogo Parámetros. Si desea utilizar un editor de expresiones, haga clic en Editar. Si utiliza el editor de expresiones, puede seleccionar las funciones y puertos que se van a usar en la expresión. Puede validar la expresión.

La siguiente imagen muestra la ficha Filtro con un parámetro para la condición de filtro:



El parámetro de expresión es un parámetro de asignación. Puede reemplazar el parámetro en un conjunto de parámetros un archivo de parámetro en tiempo de ejecución.

Cómo filtrar filas con valores nulos

Para filtrar filas que contienen valores nulos o espacios, utilice las funciones ISNULL e IS_SPACES para probar el valor del puerto.

Por ejemplo, si desea filtrar las filas que contienen el valor NULL en el puerto FIRST_NAME, utilice la siguiente condición:

```
IIF (ISNULL (FIRST_NAME) , FALSE, TRUE)
```

Esta condición indica que, si el puerto FIRST_NAME es NULL, el valor de devolución es FALSE y que la fila debe descartarse. De lo contrario, la fila pasa a la siguiente transformación.

Propiedades avanzadas de la transformación de filtro

Configure las propiedades para ayudarle a determinar cómo el servicio de integración de datos procesa los datos para la transformación de filtro.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Consejos sobre el rendimiento de la transformación de filtro

Siga los consejos para aumentar el rendimiento de la transformación de filtro.

Utilice la transformación de filtro en los primeros lugares de la asignación.

Mantenga la transformación de filtro lo más cerca posible de los orígenes en la asignación. En lugar de pasar las filas que tiene previsto descartar mediante la asignación, puede filtrar los datos no deseados en los primeros lugares del flujo de datos desde los orígenes a los destinos.

Transformación de filtro en un entorno no nativo

El procesamiento de transformaciones de filtro en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.

- Motor de Databricks Spark. Compatible sin restricciones.

Transformación de filtro en el motor de Blaze

Cuando una asignación contiene una transformación de filtro en una columna particionada de un origen de Hive, el motor de Blaze solo puede leer las particiones que contengan datos que cumplan la condición del filtro. Para insertar el filtro en el origen de Hive, configure la transformación de filtro de forma que sea la siguiente transformación en la asignación después del origen.

CAPÍTULO 18

Transformación de jerárquica a relacional

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de jerárquica a relacional, 312](#)
- [Ejemplo: transformación de jerárquica a relacional, 313](#)
- [Puertos de salida relacionales y la vista Resumen, 314](#)
- [Puertos de la transformación de jerárquica a relacional, 315](#)
- [Referencias a esquemas, 316](#)
- [Procesamiento de instrucciones de consulta, 316](#)
- [Desarrollo de la transformación de jerárquica a relacional, 317](#)

Resumen de la transformación de jerárquica a relacional

La transformación de jerárquica a relacional procesa entradas jerárquicas XML o JSON y las transforma en salidas relacionales. Una transformación de jerárquica a relacional lee las entradas jerárquicas de los puertos de entrada y transforma los datos en salidas relacionales en los puertos de salida de la transformación. Para transformar la entrada jerárquica en una salida relacional, utilice un archivo de esquema para definir los datos jerárquicos.

Puede utilizar el asistente de transformación de jerárquica a relacional para asignar los datos automáticamente. Puede configurar la asignación a los puertos de salida relacionales en la vista **Resumen** de la transformación.

Después de que el asistente genera la transformación, puede pasar los datos de los puertos de salida relacionales a otra transformación de una asignación.

Nota: La transformación de jerárquica a relacional puede procesar hasta 10 000 elementos de esquema en un archivo .xsd. Para procesar más de 10 000 elementos, divida los datos en varios archivos.

Ejemplo: transformación de jerárquica a relacional

El departamento de logística de la empresa de transportes Harrinder debe procesar datos de envíos. Necesitan transformar el inventario y los datos de clientes del formato jerárquico a datos relacionales que pueden almacenar en tablas de bases de datos.

Necesitan crear una asignación que transforme los datos jerárquicos en datos relacionales. El sistema de inventario de la organización genera datos de inventario de envío en formato jerárquico. La asignación debe utilizar una transformación de jerárquica a relacional que permita la entrada de datos de envío y la salida de detalles en un formato relacional utilizable.

La entrada Envíos está en formato jerárquico. El elemento Envío contiene subelementos con datos de clientes y de inventario para cada envío:

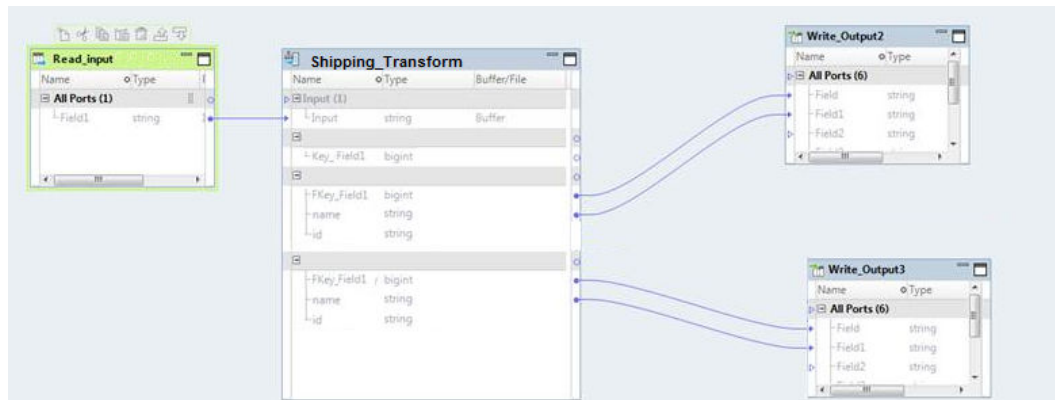
```
Shipments
Shipment
  Items
    Item_Name
    Inventory_ID
  Customer
    Customer_Name
    Customer_ID
    Customer_Address
```

En la salida relacional, el elemento Customer_ID es una clave principal en la tabla Cliente y es una clave externa en la tabla Envío.

Customer_ID	Customer_Name	Customer_Address
3543766	Tony Birch	6 Moby Drive
6342562	Sujita Man	22 Dan Street
6471862	Dwayne Horace	7 Jafendar Boulevard
7265204	Carmela Perez	23 Dan Street
4559672	Delilah Soraya	28 Jafendar Boulevard

Shipment_ID	Inventory_Item	Customer_ID
9173327437	908274	7265204
9174562342	553439	7265204
8484526471	546584	3543766
7023847265	908274	3543766
9174596725	553439	3543766

La siguiente imagen muestra la asignación en este ejemplo:



La asignación contiene los siguientes objetos:

Read_input

El origen que contiene la ruta de acceso al archivo con datos jerárquicos. Lee los datos de facturación desde un archivo XML.

Shipping_Transform

Una transformación de jerárquica a relacional que transforma la entrada XML en salida relacional.

Write_Output2

Un destino que almacena parte de los datos transformados, la tabla Cliente, en formato relacional.

Write_Output3

Un segundo destino que almacena otra parte de los datos transformados, la tabla Envío, en formato relacional.

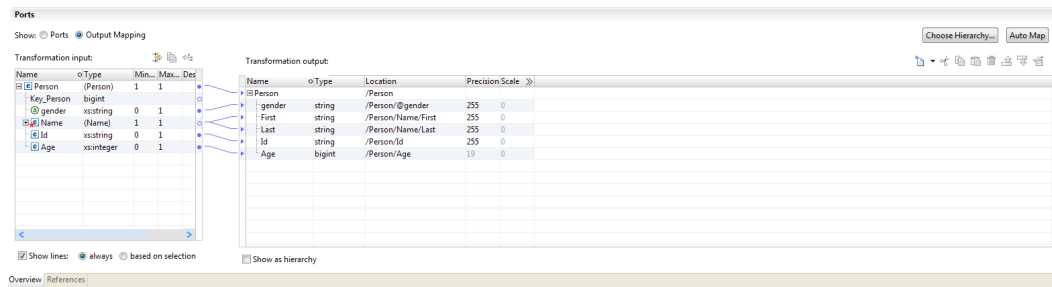
La asignación utiliza el archivo sin formato **Read_Input** para permitir la entrada de la ruta de acceso de destino para la entrada jerárquica. La asignación procesa y transforma los datos mediante la transformación **Shipping_Transform**. A continuación, la asignación almacena la salida en los dos destinos de salida.

Puertos de salida relacionales y la vista Resumen

Para convertir datos jerárquicos a salidas relacionales en la transformación de jerárquica a relacional, el asistente genera vínculos entre los nodos jerárquicos y los puertos relacionales. Utilice la vista **Resumen** para ver los vínculos entre los puertos relacionales y los puertos jerárquicos. También puede crear grupos de puertos de salida vinculando nodos de la salida jerárquica con grupos de puertos.

Para ver la asignación de grupos relacionales, utilice la vista **Resumen**. Seleccione **Asignación de salida**. El panel **Puertos** aparece en la vista **Resumen**.

La siguiente imagen muestra el panel **Puertos**:



El área **Entrada de transformación**, que muestra el esquema jerárquico, está a la izquierda. El área **Salida de transformación**, que muestra los puertos de salida relacionales, está a la derecha.

Puede definir los puertos de salida relacional en el área **Salida de transformación** y vincular nodos del esquema con los puertos. También puede arrastrar el puntero desde un nodo del esquema hacia un campo vacío del área **Salida de transformación** para crear un puerto. Cuando se arrastra un nodo desde el esquema de salida hacia un puerto, Developer tool muestra un vínculo entre ellos.

Puertos de la transformación de jerárquica a relacional

Los puertos de la transformación de jerárquica a relacional se definen en la vista **Resumen** de la transformación.

Una transformación de jerárquica a relacional puede leer entradas de un archivo o de un búfer. Los puertos de salida devuelven datos relacionales de la transformación.

Al crear una transformación de jerárquica a relacional, Developer tool crea un puerto de entrada predeterminado. El tipo de entrada determina el tipo de datos que el Servicio de integración de datos pasa a la transformación de jerárquica a relacional. El tipo de entrada determina si la entrada está formada por datos o si es la ruta de acceso de un archivo de origen.

Configure uno de los siguientes tipos de entrada:

Búfer

La transformación de jerárquica a relacional recibe filas de datos de origen en el puerto de entrada. Utilice el tipo de entrada de búfer cuando configure la transformación para recibir datos desde una transformación de Informatica.

Archivo

La transformación de jerárquica a relacional recibe la ruta de acceso al archivo de origen en el puerto de entrada. La transformación de jerárquica a relacional abre el archivo de origen. También puede utilizar el tipo de entrada de archivo para archivos de gran tamaño que puede requerir un lote de memoria del sistema para el proceso con un puerto de entrada de búfer.

Al crear la transformación con el asistente para nueva transformación, puede definir un archivo de entrada de ejemplo. Un archivo de entrada de ejemplo es un pequeño ejemplo del archivo de entrada. Haga referencia a un archivo de entrada de ejemplo cuando cree una transformación de jerárquica a relacional. También puede utilizar el archivo de entrada de ejemplo cuando pruebe la transformación en la vista **Visor de datos**.

La transformación contiene uno o varios grupos de puertos que devuelven datos relacionales.

Referencias a esquemas

Una transformación de jerárquica a relacional requiere un esquema jerárquico para definir la jerarquía de entrada en la transformación. Para utilizar el esquema en la transformación, puede definir una referencia de esquema.

Puede definir referencias de esquema de la transformación en la vista **Referencias** de la transformación.

La transformación de jerárquica a relacional hace referencia a los objetos de esquema del repositorio de modelos. Los objetos de esquema pueden existir en el repositorio antes de crear la transformación. También puede importar esquemas desde la vista **Referencias** de la transformación.

Un esquema puede hacer referencia a esquemas adicionales. La vista **Referencias** muestra el espacio de nombres y el prefijo para cada esquema al que la transformación de jerárquica a relacional hace referencia. Cuando se hace referencia a varios esquemas con espacios de nombres vacíos, la transformación no es válida.

Procesamiento de instrucciones de consulta

En el panel **Puertos**, la transformación muestra la asignación entre los nodos de esquema jerárquicos y los puertos relacionales. La transformación utiliza un esquema para definir la entrada jerárquica. Si el esquema tiene más de un elemento que puede ser un elemento raíz, elija un nodo para que sea el elemento raíz.

El asistente genera vínculos entre los nodos de esquema jerárquicos y los puertos relacionales. Si desea cambiar los vínculos generados, puede utilizar el panel **Puertos** para añadir, eliminar o editar los vínculos. Puede vincular nodos a puertos y crear un puerto.

Cuando vincula nodos al área **Salida de transformación**, Developer tool actualiza el campo de ubicación con la ubicación del nodo en la jerarquía. Si crea manualmente los puertos, debe asignar un nodo al puerto. Actualice la columna **Ubicación** y seleccione un nodo de la lista.

Cuando vincula un nodo de ocurrencia múltiple a un grupo que contiene el elemento primario, puede configurar el número de ocurrencias de elemento secundario que se van a incluir. También puede reemplazar el grupo principal con el grupo secundario de ocurrencia múltiple en la salida de transformación.

Para crear un grupo, vincule un nodo a una columna vacía en el área **Salida de transformación**. Si vincula un nodo secundario de ocurrencia múltiple a una columna de entrada o de salida vacía, Developer tool le pedirá que relacione el grupo con otros grupos de salida. Cuando se selecciona un grupo, Developer tool crea claves para relacionar los grupos.

Configure grupos relacionados de puertos de salida en el área **Salida de transformación**. Cuando Developer tool le insta a relacionar grupos de salida, añada las claves a los grupos. También puede añadir puertos manualmente para representar claves.

Desarrollo de la transformación de jerárquica a relacional

Utilice el asistente para nueva transformación para generar automáticamente una transformación de jerárquica a relacional. Elija un archivo de ejemplo de esquema o jerárquico para definir la jerarquía de entrada.

1. Cree la transformación en Developer tool.
2. Configure el puerto de entrada y la asignación.
3. Pruebe la transformación.

Crear la transformación de jerárquica a relacional

1. En Developer tool, haga clic en **Archivo > Nuevo > Transformación**.
2. Seleccione la transformación de jerárquica a relacional y haga clic en **Siguiente**.
3. Introduzca un nombre para la transformación, busque la ubicación en el repositorio de modelos donde colocarla y haga clic en **Siguiente**.
4. Para seleccionar un esquema, seleccione uno de los siguientes métodos:
 - Para utilizar un esquema del repositorio de modelos para definir la jerarquía de entrada, al lado del campo **Objeto de esquema**, examine para seleccionar el archivo de esquema en el repositorio.
 - Para importar un archivo de esquema, haga clic en **Crear un nuevo objeto de esquema**. En la ventana **Nuevo objeto de esquema**, puede examinar y seleccionar un archivo de esquema, o puede seleccionar crear un esquema desde un archivo jerárquico de muestra.
5. Elija la raíz de la jerarquía de salida. En el cuadro de diálogo **Raíz de jerarquía**, seleccione el elemento en el esquema que sea el elemento raíz del archivo jerárquico de salida. Para seleccionar el objeto raíz, puede añadir un archivo jerárquico de muestra. Para añadir un archivo de muestra, junto al campo **Archivo de ejemplo**, busque y seleccione el archivo del sistema de archivos.
6. Haga clic en **Finalizar**.
El asistente crea la transformación en el repositorio.

Configurar los puertos y la asignación

Configure los puertos de entrada y de salida en la vista **Resumen**.

1. Seleccione el tipo de datos, el tipo de puerto, la precisión y la escala del puerto de entrada.
2. Para ver la asignación, en el área **Puertos** de la vista **Resumen**, seleccione **Asignación de salida**.
3. Expanda los árboles en la cuadrícula **Puertos**. A la izquierda, el panel **Entrada de transformación** muestra la entrada jerárquica prevista, mientras que a la derecha, el panel **Salida de transformación** muestra la salida relacional.
4. Para definir un nodo como raíz, haga clic en **Seleccionar jerarquía**.
Developer tool solo muestra los nodos desde el nivel raíz y debajo del nivel raíz en el área **Entrada de transformación**.
5. Para ver las líneas que conectan los puertos con los nodos jerárquicos, haga clic en **Mostrar líneas**. Elija entre ver todas las líneas de conexión o solo las líneas de los puertos seleccionados.

6. Para añadir un grupo o un puerto de entrada al área **Salida de transformación**, utilice uno de los siguientes métodos:
 - Arrastre un elemento simple o complejo del área **Entrada de transformación** a una columna vacía del área **Salida de transformación**. Si el nodo es un nodo de grupo, Developer tool agrega un grupo relacional sin puertos.
 - Para añadir un grupo relacional, seleccione una fila y haga clic con el botón derecho para seleccionar **Nuevo > Grupo**.
 - Para añadir un puerto relacional, haga clic con el botón derecho para seleccionar **Nuevo > Campo**.
7. Para borrar la configuración de nodos jerárquica de las ubicaciones de puertos, utilice uno de los siguientes métodos:
 - Seleccione uno o varios nodos en el área **Entrada de transformación**, haga clic con el botón derecho y seleccione **Borrar**.
 - Seleccione una o más líneas que conecten los puertos relacionales con los nodos jerárquicos, haga clic con el botón derecho y seleccione **Eliminar**.
8. Para mostrar los puertos de salida en la jerarquía, haga clic en **Mostrar como jerarquía**. Cada grupo secundario aparece debajo del grupo primario.

Probar la transformación

Pruebe la transformación de jerárquica a relacional en la vista **Visor de datos**.

Antes de probar la transformación, compruebe que ha definido la ubicación de entrada del archivo. Puede definir la ubicación de entrada en el equipo del servicio de integración de datos en la columna Ubicación de entrada del panel **Puertos**, en la vista **Resumen**.

1. Abra la vista **Visor de datos**.
2. Haga clic en **Ejecutar**.

Developer tool valida la transformación. Si no hay errores, Developer tool muestra el contenido del archivo jerárquico en el panel **Salida**.

CAPÍTULO 19

Transformación de Java

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de Java, 319](#)
- [Cómo diseñar una transformación de Java, 323](#)
- [Puertos de transformación de Java, 324](#)
- [Propiedades avanzadas de la transformación de Java, 325](#)
- [Cómo desarrollar código Java, 328](#)
- [Propiedades de Java de una transformación de Java, 332](#)
- [Optimización de filtro con la transformación de Java, 334](#)
- [Cómo crear una transformación de Java, 337](#)
- [Cómo compilar una transformación de Java, 338](#)
- [Solución de problemas de una transformación de Java, 339](#)
- [Ejemplo de conversión a datos de tipo struct, 340](#)
- [Transformación Java en un entorno no nativo, 344](#)

Resumen de la transformación de Java

Utilice la transformación de Java para ampliar las funciones de Developer Tool.

La transformación de Java ofrece una simple interfaz de programación nativa para definir funciones de transformación con el lenguaje de programación Java. Puede utilizar la transformación de Java para definir funciones simples o moderadamente complejas sin tener conocimientos avanzados del lenguaje de programación Java o de un entorno de desarrollo externo de Java. La transformación de Java puede ser una transformación activa o pasiva.

Developer Tool utiliza el kit de desarrollo de Java (JDK) para compilar el código de Java y generar el código byte para la transformación. Developer Tool almacena el código byte en el repositorio de modelos.

El servicio de integración de datos utiliza el Java Runtime Environment (JRE) para ejecutar el código byte en tiempo de ejecución. Cuando el servicio de integración de datos ejecuta una asignación con una transformación de Java, utiliza el JRE para ejecutar el código byte y procesar filas de entrada y generar filas de salida.

Cree transformaciones de Java escribiendo snippets de código Java que definan una lógica de transformación. Defina el comportamiento de la transformación de Java de acuerdo con los siguientes eventos:

- La transformación recibe una fila de entrada.
- La transformación ha procesado todas las filas de entrada.

En las asignaciones que se ejecutan en el motor de Spark, puede usar tipos de datos complejos de transformaciones Java para procesar los datos jerárquicos. Con los tipos de datos complejos, el motor de Spark lee, procesa y escribe directamente los datos jerárquicos en archivos complejos Avro, Parquet y JSON.

Transformaciones de Java reutilizables y no reutilizables

Puede crear una transformación de Java reutilizable o no reutilizable.

Las transformaciones reutilizables pueden existir en varias asignaciones.
Las transformaciones no reutilizables existen en una única asignación.

Las vistas del editor donde se definen las propiedades y se crea el código Java difieren según se crea una transformación de Java reutilizable o no reutilizable.

Transformaciones de Java activas y pasivas

Una transformación Java genera filas de salida de modo diferente según si la transformación es activa o pasiva.

Tras crear la transformación, no se puede modificar si esta es activa o pasiva.

Transformación Java activa

Una transformación activa puede cambiar el número de filas que pasan por ella.

Para definir el número de filas en la salida, llame al método `generateRow()` en el código para generar cada fila de salida. Puede elegir entre generar varias filas de salida a partir de una única fila de entrada o generar una única fila de salida a partir de varias filas de entrada. Por ejemplo, si la transformación contiene dos puertos de entrada que representan una fecha de inicio y una fecha de finalización, puede llamar al método `generateRow()` para generar una fila de salida por cada fecha entre la fecha de inicio y la fecha de finalización.

Transformación Java pasiva

Una transformación pasiva no puede cambiar el número de filas que pasan por ella. La transformación llama al método `generateRow()` para generar una fila de salida tras procesar cada fila de entrada.

Conversión de tipo de datos

Una transformación Java convierte tipos de datos de de PowerCenter en tipos de datos Java según el tipo de puerto de la transformación Java.

Cuando una transformación Java lee filas de entrada, convierte los tipos de datos del puerto de entrada en tipos de datos Java.

Cuando una transformación Java escribe filas de salida, convierte los tipos de datos Java en tipos de datos del puerto de salida.

Por ejemplo, el siguiente procesamiento tiene lugar en un puerto de entrada con el tipo de datos Integer en una transformación Java:

1. La transformación Java convierte el tipo de datos Integer del puerto de entrada en el tipo de datos int primitivo de Java.
2. En la transformación, esta trata el valor del puerto de entrada como un tipo de datos int primitivo de Java.
3. Cuando la transformación genera la fila de salida, convierte el tipo de datos int primitivo de Java en el tipo de datos Integer.

En la siguiente tabla se muestra cómo la transformación Java asigna tipos de datos de de PowerCenter a tipos de datos Java primitivos y complejos:

Tipo de datos de Developer tool	Tipo de datos Java
array*	java.util.List
bigint	long
binary	byte[]
date/time	Con procesamiento de nanosegundos habilitado, BigDecimal con precisión al nanosegundo Con procesamiento de nanosegundos deshabilitado, long con precisión al milisegundo (número de milisegundos desde el 1 de enero de 1970 00:00:00.000 GMT).
decimal	Con procesamiento de alta precisión deshabilitado, double con precisión 15. Con procesamiento de alta precisión habilitado, BigDecimal
double	doble
entero	int
map*	java.util.Map
string	String
struct*	Clase Java Bean personalizada con establecedores y captadores para los elementos de campo de estructura
text	String
*Compatible solo en el motor de Spark.	

En Java, los tipos de datos java.util.List, java.util.Map, String, byte[] y BigDecimal son tipos de datos complejos. Los tipos de datos double, int y long son tipos de datos primitivos.

En Developer tool, los tipos de datos array, struct y map son tipos de datos complejos.

La transformación Java define los valores nulos en tipos de datos primitivos como cero. Puede usar los métodos API isNull y setNull en la ficha **En entrada** para definir valores nulos en el puerto de entrada como valores nulos en el puerto de salida. Para un ejemplo, véase [“setNull” en la página 355](#).

Nota: El tipo de datos decimal se asigna a BigDecimal cuando se habilita la alta precisión. BigDecimal no se puede usar con algunos operadores, como el operador +. Si el código Java contiene una expresión que usa

un puerto decimal o un puerto complejo con un elemento de un tipo de datos decimal y el puerto se usa con uno de los operadores, el código Java no se podrá compilar.

Conversión de tipos de datos complejos en el motor de Spark

Puede añadir puertos complejos a la transformación Java de las asignaciones que se ejecutan en el motor de Spark. Un puerto complejo es un puerto al que se asigna un tipo de datos complejo. La transformación Java convierte los tipos de datos complejos en tipos de datos complejos de Java comparables. También convierte los elementos de tipos de datos complejos en tipos de datos de Java comparables, que son una versión encuadrada de tipos de datos primitivos.

Tipo de datos array

Cuando una transformación Java lee filas de entrada, convierte el tipo de datos array en tipo de datos List de Java. La transformación convierte el tipo de datos de elementos de tipo array en una versión encuadrada de los tipos de datos de Java.

Por ejemplo, la transformación convierte el tipo de datos complejo de Developer tool `array<integer>` en un tipo de datos de Java `List<Integer>`.

Cuando una transformación Java escribe filas de salida, convierte los tipos de datos List de Java en tipos de datos array. La transformación convierte el tipo de datos de elementos de tipo List en una versión no encuadrada de tipos de datos de Developer tool.

Tipo de datos struct

Cuando una transformación Java lee filas de entrada del tipo de datos struct, genera una clase Java Bean equivalente. El nombre de la clase Java Bean es el mismo que el nombre del tipo de datos struct. La transformación convierte el tipo de datos de elementos de tipo struct en una versión encuadrada de tipos de datos de Java.

Si el tipo de datos struct usa caracteres especiales o palabras clave reservadas de Java, como `final` o `private`, la transformación Java reemplaza los caracteres especiales por un guion bajo (`_`) en el nombre de la clase. Puede abrir el código completo en la vista Java de la ficha de propiedades de la transformación para ver las clases generadas.

La transformación Java genera nombres de campo miembro de la clase que comienzan por un guion bajo (`_`). También genera captadores y establecedores para los campos miembro.

La clase Java Bean generada tiene el espacio de nombres del nombre de la biblioteca de definiciones de tipos como clase externa. El nombre de la clase externa es el mismo que el nombre de la biblioteca de definiciones de tipos. El nombre de tipo de datos de Java del puerto de estructura tiene el siguiente formato:

```
type_library_name.struct_type_name
```

Por ejemplo, el nombre de la biblioteca de tipos es `m_Type_Definition_Library`. La definición del tipo de datos complejo del puerto de estructura es:

```
Customer {
  name string
  age integer
}
```

La transformación Java genera las clases Java Bean con captadores y establecedores para los campos miembro. El siguiente fragmento de código muestra la clase externa y la clase interna:

```
public static final class m_Type_Definition_Library {
  public static final class Customer implements Serializable {
    private String _name;
    private Integer _age;

    public Customer() {}
```

El siguiente fragmento de código muestra el captador y el establecedor del nombre de elemento de estructura de tipo string:

```
public String get_name() {
    return _name;
}

public void set_name(String _name) {
    this._name = _name;
}
```

Cuando una transformación Java escribe filas de salida de tipo de datos struct, convierte la clase Java Bean en un tipo de datos struct. La transformación convierte los campos miembro de la clase en elementos de estructura. El tipo de datos de los campos miembro se convierte en una versión no encuadrada de los tipos de datos de Developer tool.

Tipo de datos map

Cuando una transformación Java lee filas de entrada, convierte el tipo de datos map en un tipo de datos map de Java. La transformación convierte el tipo de datos de elementos de asignación en una versión encuadrada de tipos de datos de Java.

Por ejemplo, el tipo de datos complejo de Developer tool `map<string, bigint>` se convierte en un tipo de datos de Java `Map<String, long>`.

Cuando una transformación Java escribe filas de salida, convierte los tipos de datos map de Java en tipos de datos map. La transformación convierte el tipo de datos de elementos de asignación en una versión no encuadrada de los tipos de datos de Developer tool.

Cómo diseñar una transformación de Java

Cuando se diseña una transformación de Java, es necesario tener en cuenta factores, como el tipo de transformación que se desea crear.

Cuando diseñe una transformación de Java, tenga en cuenta los siguientes aspectos:

- ¿Necesita crear una transformación de Java activa o pasiva?
Una transformación de Java pasiva genera una fila de salida para cada fila de entrada de la transformación.
Una transformación de Java activa genera varias filas de salida para cada fila de entrada de la transformación.
- ¿Necesita definir funciones en la transformación de Java? En caso afirmativo, ¿qué expresiones desea incluir en cada función?
Por ejemplo, puede definir una función que invoque una expresión para consultar los valores de los puertos de entrada y de salida o para buscar los valores de las variables de la transformación de Java.
- ¿Desea crear una transformación de Java reutilizable o no reutilizable?
Una transformación reutilizable puede existir en varias asignaciones.
Una transformación no reutilizable puede existir en una única asignación.

Puertos de transformación de Java

Una transformación de Java puede tener puertos de entrada y de salida.

Para crear y editar los puertos de una transformación de Java no reutilizable, utilice la ficha **Puertos** del editor. Para crear y editar los puertos de una transformación de Java reutilizable, utilice la vista **Resumen** en el editor.

Puede especificar valores predeterminados para los puertos. Después de añadir puertos a una transformación, puede usar los nombres de puerto como variables en snippets de código Java.

Cómo crear puertos

Cuando se crea una transformación de Java, ésta incluye un grupo de entrada y un grupo de salida.

Cuando se crea un puerto, la herramienta Tool lo añade debajo de la fila o el grupo seleccionado en ese momento.

Cómo configurar valores de puerto predeterminados

Se pueden definir valores predeterminados para los puertos en una transformación de Java.

La transformación de Java inicializa las variables del puerto con el valor de puerto predeterminado basado en el tipo de datos del puerto.

Puertos de entrada y de salida

La transformación de Java inicializa el valor de puertos de entrada no conectados o de puertos de salida que no tienen un valor asignado en los snippets de código Java.

La transformación Java inicializa puertos según los siguientes tipos de datos Java:

Tipo de datos primitivo

Si se define un valor predeterminado para el puerto que no es igual a nulo, la transformación inicializa el valor de la variable del puerto con el valor predeterminado. De lo contrario, inicializa el valor de la variable del puerto con 0.

Compleja tipo de datos

Si se define un valor predeterminado para el puerto, la transformación crea un objeto e inicializa el objeto con el valor predeterminado. De lo contrario, la transformación inicializa la variable de puerto con un valor nulo. Por ejemplo, si se define un valor predeterminado para un puerto de tipo string, la transformación crea un objeto de tipo string e inicializa ese objeto de tipo string con el valor predeterminado.

Nota: Si se accede a una variable de puerto de entrada con un valor nulo en el código Java, se produce una `NullPointerException`.

Puede habilitar un puerto de entrada como una clave de partición y una clave de ordenación, y puede asignar una dirección de ordenación. El servicio de integración de datos crea particiones de los datos y los ordena en cada partición según la clave de ordenación y la dirección de ordenación. La clave de partición y la clave de ordenación son válidas cuando el ámbito de la transformación se establece en Todas las entradas.

Utilice las siguientes propiedades para particionar y ordenar los datos:

Clave de partición

Puerto de entrada que determina las filas de datos para agruparlas en la misma partición.

Puede habilitar una o más filas de entrada como claves de partición. El servicio de integración de datos utiliza las claves de partición para la repartición de datos antes de que el código se ejecute. Si no selecciona una fila de entrada como clave de partición, los datos se procesan utilizando el esquema de partición predeterminado.

Clave de ordenación

Puerto de entrada que determina los criterios de ordenación dentro de cada partición.

Dirección

Orden ascendente o descendente. El orden ascendente es el predeterminado.

Propiedades avanzadas de la transformación de Java

La transformación de Java incluye propiedades avanzadas, tanto para el código de la transformación como para la propia transformación.

Cuando se utiliza la transformación en una asignación, se pueden reemplazar las propiedades de la transformación.

Puede definir las siguientes propiedades avanzadas para la transformación de Java en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Admite partición

La transformación se puede procesar con varios subprocesos. Desactive esta opción si quiere que el Servicio de integración de datos utilice un subproceso para procesar la transformación. El Servicio de integración de datos puede usar varios subprocesos para procesar el resto de etapas del canal de la asignación.

Desactive las particiones para una transformación de Java cuando el código Java requiera que la transformación se procese con un subproceso.

Habilitar alta precisión

Procesa un puerto de tipo de datos decimal con una precisión menor o igual a 38 como puerto de tipo de datos BigDecimal de Java.

Deshabilite la alta precisión para procesar un puerto de tipo de datos decimal como puerto de tipo de datos double de Java.

En la siguiente tabla se muestra cómo una transformación de Java trata un valor en un puerto de entrada con tipo de datos decimal dependiendo si se ha habilitado o no la opción de alta precisión:

Ejemplo	Procesamiento de alta precisión habilitado	Procesamiento de alta precisión deshabilitado
Un puerto de entrada con tipo decimal recibe el valor 40012030304957666903.	La transformación de Java deja el valor como está.	La transformación de Java lo convierte en el siguiente valor: 4.00120303049577 x 10 ¹⁹

Si la transformación Java contiene un puerto decimal o un puerto complejo con un elemento de un tipo de datos decimal, la transformación deberá usar el mismo modo de precisión que la asignación. Por ejemplo, si la transformación Java tiene habilitada la alta precisión, deberá habilitar la alta precisión en la asignación.

Usar nanosegundos en Fecha/Hora

Convierte los puertos con tipo de datos date/time en puertos con tipo de datos de Java BigDecimal con precisión al nanosegundo.

Deshabilite el procesamiento al nanosegundo de modo que el código Java generado convierta los puertos con tipo de datos date/time en puerto con tipo de datos Java Long y una precisión al milisegundo.

Classpath

Establece la classpath para directorios de archivos jar o class que están asociados a paquetes Java jo estándar que se importan en la ficha **Importaciones**.

Los directorios de los archivos jar o class deben ser accesibles en el equipo cliente de Developer tool para poder compilar el código Java.

Según el sistema operativo, separe cada una de las entradas de classpath del siguiente modo:

- En UNIX, utilice el signo de dos puntos para separar cada una de las entradas de classpath.
- En Windows, utilice un punto y coma para separar cada una de las entradas de classpath.

Por ejemplo, si importa el paquete de conversión de Java en la ficha **Importaciones** y define el paquete en converter.jar, debe añadir la ubicación del archivo converter.jar a la classpath antes de compilar el código Java de la transformación de Java.

Nota: No es necesario definir la classpath para paquetes Java incorporados. Por ejemplo, puesto que java.io es un paquete Java incorporado, no es necesario definir classpath para java.io.

Está activo

Indica si la transformación es una transformación activa. Una transformación activa puede cambiar el número de filas que pasan por ella.

Esta propiedad no se puede modificar después de crear la transformación. Si necesita cambiarla, cree una nueva transformación.

Ámbito de transformación

Define el método que utiliza el servicio de integración de datos para aplicar la lógica de transformación a los datos entrantes. Puede seleccionar uno de los siguientes valores:

- Fila. Aplica la lógica de transformación a una fila de datos cada vez. Elija Fila cuando los resultados del procedimiento dependan de una única fila de datos.
- Transacción. Aplica la lógica de transformación a todas las filas de una transacción. Elija Transacción cuando los resultados del procedimiento dependan de todas las filas en la misma transacción, pero no de las filas en otras transacciones. Cuando elija Transacción, debe conectar todos los grupos de entrada con el mismo punto de control de transacción.
- Todas las entradas. Aplica la lógica de transformación a todos los datos entrantes. Cuando elija Todas las entradas, el servicio de integración de datos quita los límites de transacción. Elija Todas las entradas cuando los resultados del procedimiento dependen de todas las filas de datos en el origen.

Sin estado

Mantenga el orden de las filas de los datos de entrada en la transformación. Seleccione esta opción si el Servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas.

Cuando el Servicio de integración de datos realiza optimizaciones, puede hacer que se pierda un orden establecido anteriormente en la asignación. Puede establecer el orden en una asignación con un origen de archivo sin formato ordenado, un origen relacional ordenado o una transformación de ordenación. Al configurar una transformación para mantener el orden de las filas, el Servicio de integración de datos tiene en cuenta esta configuración cuando realiza optimizaciones para la asignación. El Servicio de integración de datos realiza optimizaciones para la transformación si puede mantener el orden. El Servicio de integración de datos no realiza optimizaciones para la transformación si el optimización cambia el orden de las filas.

Cómo configurar la classpath para el cliente de Developer Tool

Puede añadir archivos jar o directorios de archivos de clase a la classpath del Cliente de Developer Tool.

Para definir la classpath para el equipo en el que se ejecuta el Cliente de Developer Tool, lleve a cabo una de las siguientes tareas:

- Configure la variable de entorno CLASSPATH. Defina la variable de entorno CLASSPATH en el equipo donde reside el Cliente de Developer Tool. Esto se aplica a todos los procesos Java que se ejecutan en la máquina.
- Para una transformación de Java no reutilizable, configure la classpath en las propiedades avanzadas de la transformación de Java. Esto se aplica a las asignaciones que incluyen esta transformación de Java. El cliente de Developer tool incluye archivos dentro de la ruta de clases cuando compila el código Java.

Para añadir directorios de archivos jar o de clase a la classpath en una transformación de Java, lleve a cabo los siguientes pasos:

1. En la ficha **Avanzadas**, haga clic en el icono de flecha descendiente en la columna **Valor**, junto a **Classpath**.
Se abrirá el cuadro de diálogo **Editar ruta de clase**.
2. Para añadir una ruta de clase, complete los siguientes pasos:
 - a. Haga clic en **Añadir**.
Aparecerá la ventana **Guardar como**.
 - b. En la ventana **Guardar como**, localice el directorio donde se encuentra el archivo jar.
 - c. Haga clic en **Aceptar**.
La classpath aparece en el cuadro de diálogo **Editar classpath**.
3. Para eliminar un archivo jar o un directorio de archivos de clase, seleccione el directorio de archivos jar o de clase y haga clic en **Quitar**.
El directorio desaparece de la lista de directorios.

Configurar la classpath para el Servicio de integración de datos

Se pueden añadir directorios de archivos jar o class que necesita la classpath en tiempo de ejecución en el nodo del Servicio de integración de datos.

Ponga los archivos jar que se necesitan en tiempo de ejecución en el siguiente directorio del nodo del Servicio de integración de datos:

```
$INFA_HOME/services/shared/jars
```

Los archivos jar que se encuentran en esta ubicación se cargan dinámicamente. Cualquier archivo de clase que necesita una asignación individual en tiempo de ejecución se encuentra y se carga desde este directorio.

Nota: La transformación de Java añade los archivos jar de este directorio a la classpath a nivel de asignación.

Cómo desarrollar código Java

Utilice las fichas de introducción de código de la vista **Java** para escribir y compilar código Java que defina el comportamiento de una transformación ante sucesos de transformación específicos.

Puede desarrollar snippets de código en las fichas de introducción de código en el orden que desee. Puede ver, aunque no editar, el código Java completo en la ficha **Código completo**.

Después de desarrollar snippets de código, puede compilar los snippets de código o el código Java completo y ver los resultados de la compilación en la ventana **Resultados** de las propiedades **Compilación**, en la vista **Java**.

Cada ficha de introducción de código contiene componentes que utiliza para escribir, ver y compilar código Java:

Propiedades de código

Ofrece controles que permiten ver y escribir código Java, incluyendo métodos API de una transformación de Java. La siguiente tabla describe los controles que están disponibles en las propiedades **Código**:

Control	Descripción
Navegador	<p>Muestra puertos de entrada, puertos de salida y métodos API de transformación de Java invocables.</p> <p>Haga clic en un elemento dentro del navegador para mostrar una descripción del elemento.</p> <p>Haga doble clic en un elemento para añadirlo a la ventana Código Java. También puede arrastrar un elemento desde el navegador hasta la ventana Código Java.</p> <p>El navegador está disponible en las siguientes fichas de introducción de código:</p> <ul style="list-style-type: none">- Aplicaciones auxiliares- En entrada- Al final
Ventana Código Java	<p>Permite ver o escribir código Java para la transformación. La ventana Código Java muestra código Java mediante el resaltado de sintaxis básica de Java.</p> <p>Nota: En la ficha Código completo puede ver, aunque no editar, la clase completa del código para la transformación de Java.</p> <p>La ventana Código Java está disponible en las siguientes fichas de introducción de código:</p> <ul style="list-style-type: none">- Importaciones- Aplicaciones auxiliares- En entrada- Al final- Funciones- Interfaces del optimizador- Código completo

Control	Descripción
Comando Nueva función	Abre el cuadro de diálogo Definir función que se usa para definir funciones que invocan expresiones de Java. El comando Función está disponible en la ficha Funciones .
Barra de herramientas de edición	Permite hacer clic en iconos de herramientas como cortar y pegar para editar código Java. La barra de herramientas de edición está disponible en las siguientes fichas de introducción de código: <ul style="list-style-type: none"> - Importaciones - Aplicaciones auxiliares - En entrada - Al final - Funciones - Interfaces del optimizador

Propiedades de compilación

Proporciona controles para compilar y depurar código Java. La siguiente tabla describe los controles en las propiedades **Compilación**:

Control	Descripción
Comando Compilar	Compila el código Java de la transformación.
Ventana Resultados	Muestra los resultados de compilación de la clase de la transformación de Java y permite buscar el origen de errores en el código. Para buscar un error en el código, haga clic con el botón derecho en un mensaje de error de la ventana Resultados y selecciónelo para ver el error en el código del snippet o en el código completo. También puede hacer doble clic en un mensaje de error en la ventana Resultados para encontrar el origen del mismo.

Cómo crear snippets de código Java

Para crear snippets de código que definan el comportamiento de una transformación, utilice la ventana **Código Java** de las fichas de introducción de código .

1. Haga clic en la ficha de introducción de código apropiada.

La siguiente tabla describe las tareas que se pueden completar en las fichas de introducción de código en la vista **Java**:

Ficha	Descripción
Importaciones	Importa paquetes de Java de terceros, incorporados y personalizados para una transformación de Java activa o pasiva. Después de importar paquetes, puede utilizarlos en otras fichas de introducción de código.
Aplicaciones auxiliares	Declara variables y métodos definidos por el usuario para la clase de transformación de Java en una transformación de Java activa o pasiva. Después de declarar las variables y los métodos, puede usarlos en cualquier otra ficha de introducción de código, excepto en la ficha Importaciones .
En entrada	Indica el comportamiento de una transformación de Java activa o pasiva cuando recibe una fila de entrada. El código Java que se define en esta ficha se ejecuta una vez para cada fila de entrada. En esta ficha también puede acceder y utilizar datos de puertos de entrada y de salida, variables y métodos API de transformación de Java.
Al final	Indica el comportamiento de una transformación de Java activa o pasiva cuando procesa todos los datos de entrada. En esta ficha también puede definir los datos para transformaciones activas e invocar métodos API de transformación de Java.
Funciones	Define las funciones que invocan expresiones en una transformación de Java mediante el lenguaje de programación Java. Por ejemplo, puede definir una función que invoque una expresión que busque los valores de puertos de entrada o de salida o que busque los valores de variables de una transformación de Java. En la ficha Funciones puede definir funciones manualmente o hacer clic en Nueva función para invocar el cuadro de diálogo Definir función , que permite definir una función fácilmente.
Interfaces del optimizador	Define la optimización de primera selección o de filtro de inserción. Seleccione el método de optimización en el navegador. Actualice los fragmentos de código para habilitar la optimización. Defina los puertos de entrada y los puertos de salida asociados para insertar la lógica de filtro.
Código completo	Solo lectura. En esta ficha puede ver, aunque no editar, la clase completa del código de la transformación de Java.

- Para acceder a variables de columna de entrada o de salida en el snippet, expanda la lista **Entrada o Salida** en el navegador y haga doble clic en el nombre del puerto.
- Para invocar un método API de transformación de Java en el snippet, expanda la lista **API invocables** en el navegador y haga doble clic en el nombre del método. En caso necesario, configure los valores de entrada apropiados para el método.
- Escriba el código Java apropiado, basado en el tipo de ficha de introducción de código.

Visualice el código de clase completo de la transformación de Java, en las ventanas **Código Java**, en la ficha **Código completo**.

Cómo importar paquetes de Java

En la ficha **Importaciones** se pueden importar paquetes de Java para transformaciones de Java activas o pasivas.

Puede importar paquetes de Java desarrollados por terceros, incorporados o personalizados. Después de importar paquetes de Java, puede utilizarlos en otras fichas de introducción de código.

Nota: En la ficha **Importaciones** no se pueden declarar o usar variables estáticas, variables de instancia o métodos de usuario.

En el Developer Tool, cuando se exportan o importan datos que contienen una transformación de Java, los archivos jar o class que contienen los paquetes de terceros o paquetes personalizados requeridos por la transformación de Java no se incluyen en la exportación ni en la importación.

Si se importan metadatos que contienen una transformación de Java, debe copiar los archivos jar o class que contienen los paquetes de terceros o paquetes personalizados al cliente de Developer Tool y al nodo del servicio de integración de datos.

Por ejemplo, para importar el paquete Java I/O, especifique el siguiente código en la ficha **Importaciones**:

```
import java.io.*;
```

Cuando se importan paquetes de Java que no son estándar, añada el paquete o la clase a la classpath en la de la transformación de Java.

Cómo definir el código de una aplicación auxiliar

En la ficha **Aplicaciones auxiliares**, puede declarar variables y métodos definidos por el usuario para la clase de una transformación de Java en transformaciones de Java activas o pasivas.

Después de declarar variables y métodos en la ficha **Aplicaciones auxiliares**, puede utilizar las variables y los métodos en cualquier ficha de introducción de código, excepto en la ficha **Importaciones**.

En la ficha **Aplicaciones auxiliares**, puede declarar los siguientes tipos de código, variables y métodos:

- Código estático y variables estáticas.

Dentro de un bloque estático puede declarar variables estáticas y código estático. Todas las instancias de una transformación de Java reutilizable en una asignación comparten código estático y variables estáticas. En una transformación de Java, el código estático se ejecuta antes de cualquier otro código.

Por ejemplo, el siguiente código declara una variable estática para almacenar el umbral de error para todas las instancias de una transformación de Java en una asignación:

```
static int errorThreshold;
```

Utilice esta variable para almacenar el umbral de error para la transformación y acceder al mismo desde todas las instancias de la transformación de Java de una asignación.

Nota: Debe sincronizar las variables estáticas en una transformación de Java reutilizable.

- Variables de instancia.

Varias instancias de una transformación de Java reutilizable en una asignación no comparten variables estáticas. Declare variables de instancia con un prefijo para evitar conflictos e inicializar variables de instancia no primitivas.

Por ejemplo, el siguiente código utiliza una variable booleana para decidir si se genera o no una fila de salida:

```
// boolean to decide whether to generate an output row  
// based on validity of input  
private boolean generateRow;
```

- Métodos estáticos o instancias definidas por el usuario.

Amplía las funciones de la transformación de Java. Los métodos de Java declarados en la ficha **Aplicaciones auxiliares** pueden usar o modificar variables de salida o variables de instancia declaradas localmente. No se puede acceder a variables de entrada desde métodos de Java en la ficha **Aplicaciones auxiliares**.

Por ejemplo, utilice el siguiente código en la ficha **Aplicaciones auxiliares** para declarar una función que añada dos enteros:

```
private int myTXAdd (int num1,int num2)
{
    return num1+num2;
}
```

Propiedades de Java de una transformación de Java

Utilice las fichas de introducción de código de la vista **Java** para escribir y compilar código Java que defina el comportamiento de una transformación ante sucesos de transformación específicos.

Las siguientes fichas son las fichas para la introducción de código:

- **Importaciones**
- **Aplicaciones auxiliares**
- **En entrada**
- **Al final**
- **Funciones**
- **Interfaces del optimizador**

Visualice el código de clase completo para la transformación de Java en la ficha **Código completo**.

Ficha Importaciones

En la ficha **Importaciones** se pueden importar paquetes de Java de terceros, incorporados o personalizados para transformaciones de Java activas o pasivas.

Para importar un paquete de Java, introduzca el código para importar el paquete en la ventana **Código Java**, en las propiedades **Código** de la ficha **Importaciones**.

Por ejemplo, puede introducir el siguiente código para importar el paquete java.io:

```
import java.io.*;
```

Para compilar el código que importa paquetes de Java, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **Importaciones**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **Importaciones**.

Después de importar paquetes de Java, puede utilizarlos en otras fichas de introducción de código.

Ficha Aplicaciones auxiliares

En la ficha **Aplicaciones auxiliares** puede declarar variables y métodos definidos por el usuario para la clase de transformación de Java en un archivo o en una transformación de Java pasiva.

Para declarar variables y métodos definidos por el usuario, introduzca el código en la ventana **Código Java**, en las propiedades **Código** de la ficha **Aplicaciones auxiliares**.

Para compilar el código de la aplicación auxiliar para la transformación de Java, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **Aplicaciones auxiliares**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **Aplicaciones auxiliares**.

Después de declarar las variables y los métodos, puede usarlos en cualquier otra ficha de introducción de código, excepto en la ficha **Importaciones**.

Ficha En entrada

En la ficha **En entrada** se indica el comportamiento de una transformación de Java activa o pasiva cuando recibe una fila de entrada. En esta ficha también puede acceder y utilizar datos de puertos de entrada y de salida, variables y métodos API de transformación de Java.

El código Java que se define en esta ficha se ejecuta una vez para cada fila de entrada.

Para definir cómo se comporta una transformación de Java cuando recibe una fila de entrada, especifique el código en la ventana **Código Java**, en las propiedades **Código** de la ficha **En entrada**.

Desde el explorador de la ficha **En entrada**, puede acceder a y definir las siguientes variables y métodos API:

- Variables de puerto de entrada y de puerto de salida. Acceda a los datos de los puertos de entrada y de salida utilizando el nombre del puerto como nombre de la variable. Por ejemplo, si "in_int" es un puerto de entrada con el tipo de datos Integer, puede acceder a los datos de este puerto haciendo referencia al mismo como la variable "in_int" con el tipo de datos primitivo de Java int. No es necesario que declare puertos de entrada y de salida como variables.

No asigne un valor a una variable de puerto de entrada. Si asigna un valor a una variable de entrada en la ficha **En entrada**, no se pueden obtener los datos de entrada para el puerto correspondiente en la fila actual.

- Variables de instancia y métodos definidos por el usuario. Utilice cualquier variable de instancia o estática o método definido por el usuario que haya declarado en la ficha **Aplicaciones auxiliares**.

Por ejemplo, una transformación de Java activa tiene dos puertos de entrada BASE_SALARY y BONUSSES, con un tipo de datos Integer y un puerto de salida individual TOTAL_COMP, con un tipo de datos Integer. Cree un método definido por el usuario en la ficha **Aplicaciones auxiliares**, myTXAdd, que añada dos enteros y devuelva el resultado. Utilice el siguiente código Java en la ficha **En entrada** para asignar los valores totales de los puertos de entrada al puerto de salida y generar una fila de salida:

```
TOTAL_COMP = myTXAdd (BASE_SALARY, BONUSSES);
generateRow();
```

Cuando la transformación de Java recibe una fila de entrada, añade los valores de los puertos de entrada BASE_SALARY y BONUSSES, asigna el valor al puerto de salida TOTAL_COMP y genera una fila de salida.

- Métodos API de transformación de Java. Puede invocar los métodos API proporcionados por la transformación de Java.

Para compilar el código de la transformación de Java, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **En entrada**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **En entrada**.

Ficha Al final

En la ficha **Al final** se define el comportamiento de una transformación de Java activa o pasiva después de procesar todos los datos de entrada. En esta ficha también puede definir los datos de salida para las transformaciones activas y llamar métodos API de una transformación de Java.

Para definir el comportamiento de una transformación de Java después de que haya procesado todos los datos de entrada, especifique el código en la ventana **Código Java** en las propiedades **Código** de la ficha **Al final**.

Puede acceder a las siguientes variables y definir las para métodos API en la ficha **Al final**:

- Variables de puerto de salida. Puede usar los nombres de cualquier puerto de salida que haya definido como variables en la ficha **Puertos** o definir datos de salida para las transformaciones de Java activas.
- Variables de instancia y métodos definidos por el usuario. Utilice cualquier variable de instancia o método definido por el usuario que haya declarado en la ficha **Aplicaciones auxiliares**.
- Métodos API de transformación de Java. Invoque los métodos API proporcionados por la transformación de Java.

Por ejemplo, utilice el siguiente código Java para escribir información en el registro cuando se alcance el final de los datos:

```
logInfo("Number of null rows for partition is: " + partCountNullRows);
```

Para compilar el código de la transformación de Java, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **Al final**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **Al final**.

Ficha Funciones

En la ficha **Funciones** se definen funciones que invocan expresiones en una transformación de Java con el lenguaje de programación Java.

Por ejemplo, puede definir una función que invoque una expresión para consultar los valores de los puertos de entrada y de salida o para buscar los valores de las variables de la transformación de Java.

Para definir una función, puede definir funciones manualmente en la ventana **Código Java**, en las propiedades **Código** de la ficha **Funciones**, o puede hacer clic en **Nueva función** para abrir el cuadro de diálogo **Definir función**, que permite definir fácilmente una función.

Para compilar el código, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **Funciones**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **Funciones**.

Ficha Código completo

En la ficha **Código completo** puede ver, aunque no editar, el código completo de la clase para la transformación de Java, y compilar el código.

Puede ver el código completo de la clase en la ventana **Código Java**, en las propiedades **Código**.

Para compilar el código completo de la transformación de Java, haga clic en **Compilar**, en las propiedades **Compilación** de la ficha **Código completo**. Los resultados de la compilación aparecen en la ventana **Resultados** de la ficha **Código completo**.

Optimización de filtro con la transformación de Java

El servicio de integración de datos puede aplicar la optimización del filtro para las transformaciones de Java activas. Cuando defina la transformación de Java, puede añadir código para la optimización del filtro en la ficha **Interfaces del optimizador** de la transformación de Java.

Optimización de primera selección con la transformación de Java

Puede habilitar una transformación de Java activa o pasiva para la optimización de primera selección si la transformación Java no tiene ningún efecto secundario. El optimizador pasa la lógica de filtro a través de la transformación de Java y modifica la condición de filtro según sea necesario.

Para ver los fragmentos de código para la optimización de primera selección, seleccione PredicatePushOptimization en el navegador de la ficha **Interfaces del optimizador**.

allowPredicatePush

Booleano. Habilita la primera selección. Cambie la función para devolver un resultado TRUE y el mensaje para poder habilitar la primera selección. El valor predeterminado es FALSE y la función devuelve un mensaje que no es compatible con la optimización.

```
public ResultAndMessage allowPredicatePush(boolean ignoreOrderOfOp) {
    // To Enable PredicatePushOptimization, this function should return true
    //return new ResultAndMessage(true, "");
    return new ResultAndMessage(false, "Predicate Push Optimization Is Not
Supported");
}
```

canGenerateOutputFieldEvalError

Booleano. Indica si la transformación de Java puede devolver o no un error de campo de salida, tal como un error de dividir por cero. Cambie la función devolver un resultado FALSE si la transformación de Java no genera errores del campo de salida. Cuando la transformación de Java puede generar errores de campos, el servicio de integración de datos no puede utilizar la optimización de primera selección.

```
public boolean canGenerateOutputFieldEvalError() {
    // If this Java transformation can never generate an output field evaluation error,
    // return false.
    return true;
}
```

getInputExpr

Devuelve una expresión de Informativa que describe qué valores de entrada de los campos de entrada tienen un campo de salida. El optimizador necesita saber qué campos de entrada tienen un campo de salida para insertar la lógica de filtro a la transformación.

```
public InfaExpression getInputExpr(TransformationField field,
TransformationDataInterface group) {
    // This should return an Informativa expression for output fields in terms of input
fields
    // We will only push predicate that use fields for which input expressions are
defined.
    // For example, if you have two input fields in0 and in1 and three output fields
out0, out1, out2
    // out0 is the pass-through of in1, out2 is sum of in1 and in2, and out3 is unknown,
the code should be:
    //if (field.getName().equals("out0"))
    //    return new InfaExpression("in0", instance);
    //else if (field.getName().equals("out1"))
    //    return new InfaExpression("in0 + in1", instance);
    //else if (field.getName().equals("out2"))
    //    return null;
    return null;
}
```

Por ejemplo, una asignación contiene una expresión de filtro, "out0 > 8". Out0 es el valor del puerto de salida de out0 en la transformación de Java. Puede definir el valor de out0 como el valor del puerto de entrada in0 + 5. El optimizador puede insertar la siguiente expresión "(in0 + 5) > 8" más allá de la transformación de Java con la optimización de primera selección. Puede devolver NULL si un campo de salida no tiene una expresión de campo de entrada. El optimizador no inserta expresiones de filtro después de los campos de salida sin expresiones de entrada.

Puede incluir el siguiente código:

```
if (field.getName().equals("out0"))
    return new InfaExpression("in0 + 5", instance);
else if (field.getName().equals("out2"))
    return null;
```

inputGroupsPushPredicateTo

Devuelve una lista de grupos que pueden recibir la lógica de filtro. La transformación de Java tiene un grupo de entrada. No modifique esta función para la transformación de Java.

```
public List<TransformationDataInterface> inputGroupsPushPredicateTo(
    List<TransformationField> fields) {
    // This functions returns a list of input data interfaces to push predicates to.
    // Since JavaTx only has one input data interface, you should not have to modify
    this function
    AbstractTransformation tx = instance.getTransformation();
    List<DataInterface> dis = tx.getDataInterfaces();
    List<TransformationDataInterface> inputDIs = new
    ArrayList<TransformationDataInterface>();
    for (DataInterface di : dis){
        TransformationDataInterface tdi = (TransformationDataInterface) di;
        if (tdi.isInput())
            inputDIs.add(tdi);
    }
    if(inputDIs.size() == 1)
        return inputDIs;
    else
        return null;
}
```

Optimización de inserción con la transformación de Java

Puede habilitar una transformación de Java activa para la optimización de inserción si no tiene ningún efecto secundario y la optimización no afecta a los resultados de la asignación.

Cuando configure la optimización de inserción para la transformación de Java, puede definir una forma para que la transformación de Java almacene la condición de filtro que recibe desde el optimizador. Añada el código que examina la condición de filtro. Si la transformación de Java puede absorber la lógica de filtro, entonces la transformación de Java devuelve una condición TRUE al optimizador. El optimizador elimina la transformación de filtro de la asignación optimizada.

Cuando configure la transformación de Java escriba el código que almacena la condición de filtro como metadatos de transformación durante la optimización. También puede escribir el código para recuperar la condición de filtro en tiempo de ejecución y para quitar las filas según la lógica de filtro.

Cuando defina la transformación de Java, puede añadir código para la optimización de inserción en la ficha **Interfaces del optimizador** de la transformación de Java. Para acceder a los fragmentos de código para la optimización de inserción, seleccione FilterPushdownOptimization en el navegador de la ficha **Interfaces del optimizador** de la transformación.

Developer Tool muestra los fragmentos de código para habilitar la optimización de inserción y para recibir la condición de filtro del optimizador. Actualice los fragmentos de código para habilitar la optimización y para guardar la lógica de filtro como metadatos de transformación.

isFilterSupported

Devuelve TRUE para habilitar la optimización de inserción. Devuelve FALSE para deshabilitar la optimización de inserción.

Cambie la función para que devuelva TRUE para poder habilitar la optimización de inserción.

```
public ResultAndMessage isFilterSupported() {
    // To enable filter push-into optimization this function should return true
    // return new ResultAndMessage(true, "");
}
```



```
        return new ResultAndMessage(false, "Filter push-into optimization is not supported");
    }
}
```

pushFilter

Recibe la condición de filtro del optimizador.

Añada el código para examinar el filtro y determinar si la lógica de filtro puede utilizarse en la transformación. Si la transformación puede absorber el filtro, entonces use el siguiente método para almacenar la condición de filtro como metadatos de transformación:

```
storeMetadata(String key, String data)
```

La clave es un identificador de los metadatos. Puede definir cualquier cadena como clave. Los datos son los datos que desea almacenar en orden para determinar qué filas soltar en tiempo de ejecución. Por ejemplo, los datos pueden ser la condición de filtro que la transformación de Java recibe desde el optimizador.

```
public ResultAndMessage pushFilter(InfraExpression condition) {
    // Add code to absorb the filter
    // If filter is successfully absorbed return new ResultAndMessage(true, ""); and the
optimizer
// will remove the filter from the mapping
// If the filter is not absorbed, return new ResultAndMessage(false, msg);
    return new ResultAndMessage(false, "Filter push-into optimization is not supported");
}
```

Cómo crear una transformación de Java

En Developer Tool, puede crear una transformación de Java reutilizable o no reutilizable.

Cómo crear una transformación de Java reutilizable

Puede haber transformaciones reutilizables en varias asignaciones.

Cree una transformación de Java reutilizable en Developer Tool.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de Java.
4. Haga clic en **Siguiente**.
5. Especifique un nombre para la transformación.
6. Para crear una transformación activa, seleccione la opción **Crear como activa**.
7. Haga clic en **Finalizar**.

La transformación aparece en el editor.

8. En la vista **Puertos**, haga clic en el botón **Nuevo** para añadir un puerto a la transformación.
9. Edite el puerto para definir el nombre, el tipo de datos y la precisión.
Utilice nombres de puerto como variables en los snippets de código Java.
10. En la vista **Java**, utilice las fichas de introducción de código para escribir y compilar el código Java de la transformación.

11. En la vista **Java**, utilice la ficha **Funciones** para definir funciones que invoquen expresiones.
12. En cualquier ficha de introducción de código, haga doble clic en los mensajes de error que aparecen en la ventana **Resultados**, en las propiedades **Compilación** para localizar y solucionar errores de compilación del código Java de la transformación.
13. En la vista **Avanzadas**, edite las propiedades de la transformación.

Cómo crear una transformación de Java no reutilizable

Las transformaciones no reutilizables existen en una única asignación.

Cree una transformación de Java no reutilizable en Developer Tool.

1. En una asignación o en un mapplet, arrastre una transformación de Java desde la paleta Transformación hasta el editor.
2. En el cuadro de diálogo **Nueva transformación de Java**, especifique un nombre para la transformación.
3. Para crear una transformación activa, seleccione la opción **Crear como activo**.
4. Haga clic en **Finalizar**.
La transformación aparece en el editor.
5. En la ficha **General**, edite el nombre y la descripción de la transformación.
6. En la ficha **Puertos**, haga clic en el botón **Nuevo** para añadir puertos a la transformación.
7. Edite el puerto para definir el nombre, el tipo de datos y la precisión.
Utilice nombres de puerto como variables en los snippets de código Java.
8. En la vista **Java**, utilice las fichas de introducción de código para escribir y compilar el código Java de la transformación.
9. En la vista **Java**, utilice la ficha **Funciones** para definir funciones que invoquen expresiones.
10. En cualquier ficha de introducción de código, haga doble clic en los mensajes de error que aparecen en la ventana **Resultados**, en las propiedades **Compilación** para localizar y solucionar errores de compilación del código Java de la transformación.
11. En la vista **Avanzadas**, edite las propiedades de la transformación.

Cómo compilar una transformación de Java

cliente de PowerCenter utiliza el compilador de Java para compilar el código Java y generar el código byte para la transformación.

El compilador de Java compila el código Java y muestra los resultados de la compilación en la ventana **Resultados** en las propiedades **Compilación** en las fichas de introducción de código. El compilador de Java se instala junto con el Developer Tool en el directorio `java/bin`.

Para compilar el código completo de la transformación de Java, haga clic en **Compilar**, en las propiedades **Compilación**, en la ficha **Código completo**.

Cuando se crea una transformación de Java, ésta contiene una clase de Java que define las funciones básicas de la transformación de Java. El código completo de la clase de Java contiene la clase de plantilla para la transformación, además del código Java que se define en las fichas de introducción de código.

Cuando se compila una transformación de Java, el Developer Tool añade el código desde las fichas de introducción de código para la transformación para generar el código de clase completo para la

transformación. Luego, el Developer Tool invoca el compilador de Java para compilar el código de clase completo. El compilador de Java compila la transformación y genera el código byte para la transformación.

Los resultados de la compilación se muestran en la ventana **Resultados**. Utilice los resultados de la compilación para identificar y localizar errores del código Java.

Solución de problemas de una transformación de Java

En la ventana **Resultados** de las propiedades **Compilación** de cualquier ficha de introducción de código, puede encontrar y solucionar los errores del código Java.

Se pueden producir errores en una transformación de Java debido a un error del código en una ficha de introducción de código o en el código completo de la clase de la transformación de Java.

Para solucionar los problemas de una transformación de Java, lleve a cabo los siguientes pasos de alto nivel:

1. Busque la causa del error en el código del snippet de Java en el código completo de la clase de la transformación.
2. Identifique el tipo de error. Utilice los resultados de la compilación de la ventana **Resultados** y la ubicación del error para identificar el tipo de error.
3. Repare el código Java en la ficha de introducción de código.
4. Vuelva a compilar la transformación.

Cómo buscar el origen de errores de compilación

Para buscar el origen de errores de compilación, utilice los resultados de la compilación que se muestran en la ventana **Resultados** de las propiedades **Compilación** en una ficha de introducción de código o en la ficha **Código completo**.

Cuando se hace doble clic en un mensaje de error, en la ventana **Resultados**, el código de origen que ha causado el error aparece resaltado en la ventana **Código Java**, en la ficha de introducción de código o en la ficha **Código completo**.

Se pueden buscar errores en la ficha **Código completo**, aunque no puede editar el código Java en la ficha **Código completo**. Para corregir los errores que aparecen en la ficha de introducción de código **Código completo**, modifique el código que aparece en la ficha de introducción de código apropiada. Quizás necesite utilizar la ficha **Código completo** para ver los errores causados al añadir código de usuario al código de clase completo de la transformación.

Cómo encontrar un error en una ficha de introducción de código o en la ficha Código completo

Puede encontrar errores de compilación en una ficha de introducción de código o en la ficha **Código completo**.

1. En la ventana **Resultados** de las propiedades **Compilación** de cualquier ficha de introducción de código o en la ficha **Código completo**, haga doble clic en un mensaje de error.

2. Haga clic en **Mostrar en > Snippet** o en **Mostrar en > Ficha Código completo**.

Developer Tool resalta el origen del error en la ficha seleccionada.

Nota: Puede ver, aunque no corregir, errores en la ficha **Código completo**. Para corregir los errores debe ir hasta la ficha de introducción de código apropiada.

Cómo identificar el origen de errores de compilación

Los errores de compilación pueden aparecer como resultado de errores en el código escrito por el usuario.

Los errores en el código escrito por el usuario también pueden generar un error en el código no escrito por el usuario en la clase. Los errores de compilación se producen en el código, escrito o no por el usuario, para la transformación de Java.

Errores de código de usuario

Puede haber errores en el código de usuario en las fichas de introducción de código. Los errores de código de usuario incluyen errores estándar de sintaxis Java y de lenguaje.

También puede haber errores de código de usuario cuando el Developer Tool añade el código del usuario desde las fichas de introducción de código al código de clase completo.

Por ejemplo, una transformación de Java tiene un puerto de entrada con el nombre `int1` y un tipo de datos `Integer`. El código completo de la clase declara la variable de puerto de entrada con el siguiente código:

```
int int1;
```

Sin embargo, si se utiliza el mismo nombre de variable en la ficha **En entrada**, el compilador de Java emite un error para una redeclaración de una variable. Para corregir el error, cambie el nombre de la variable en la ficha **En entrada**.

Errores ajenos al código de usuario

El código de usuario en las fichas de introducción de código puede causar errores en el código no creado por el usuario.

Por ejemplo, una transformación de Java tiene un puerto de entrada y un puerto de salida con los nombres `int1` y `out1` con tipos de datos `Integer`. Escriba el siguiente código en la ficha de introducción de código **En entrada** para calcular el interés para el puerto de entrada `int1` y asignarlo al puerto de salida `out1`:

```
int interest;
interest = CallInterest(int1); // calculate interest
out1 = int1 + interest;
}
```

Cuando se compila la transformación, el Developer Tool añade el código desde la ficha de introducción de código **En entrada** al código de clase completo para la transformación. Cuando el compilador de Java compila el código Java, el corchete no cerrado hace que un método del código de clase completo finalice prematuramente y el compilador de Java emite un error.

Ejemplo de conversión a datos de tipo struct

La organización necesita convertir un volumen extenso de datos de clientes de un archivo sin formato en datos de tipo struct, así como escribirlos en un archivo de Avro. El archivo de entrada contiene detalles de

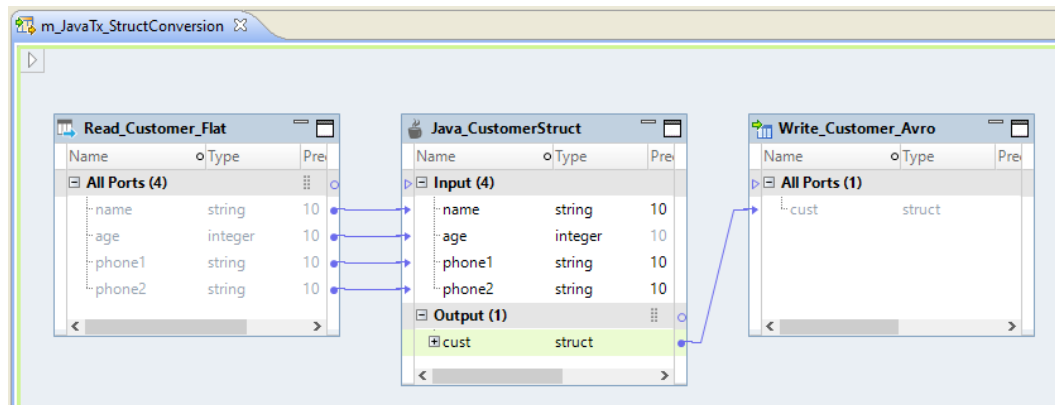
los clientes, como el nombre, la edad y los números de teléfono. Si el nombre del cliente es nulo en el archivo de entrada, no quiere que los detalles de los clientes se añadan al archivo de salida.

Puede desarrollar una asignación con una transformación Java para definir la funcionalidad de la transformación. En el entorno Hadoop, ejecute la asignación en el motor de Spark para transformar los datos y escribir los datos de tipo struct en un archivo de Avro.

Cree una asignación y configure las siguientes transformaciones:

- Una transformación de lectura que lea la información de los clientes desde un origen de archivo sin formato
- Una transformación Java que actúa como transformación activa que convierte los datos sin formato en datos de tipo struct y elimina los datos incoherentes
- Una transformación de escritura que escribe los datos de tipo struct en un archivo de Avro

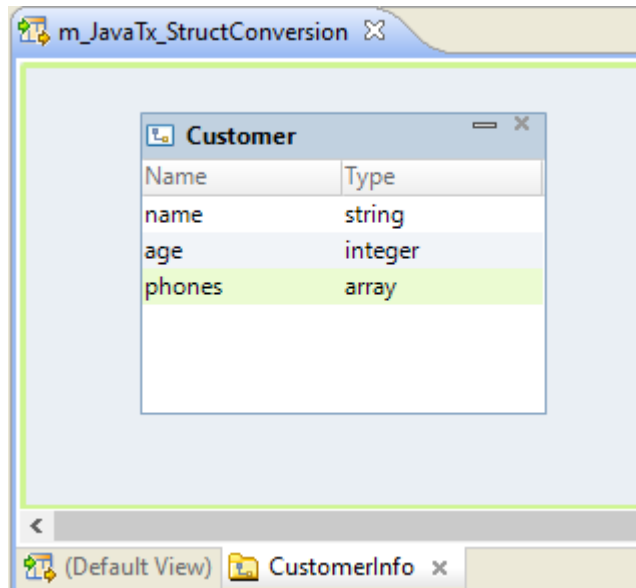
En la siguiente imagen se muestra la asignación con una transformación de lectura, una transformación Java y una transformación de escritura.



En la ficha de la biblioteca de definiciones de tipos del editor de la asignación, cree una definición de tipo de datos complejo Customer. La definición de tipo de datos complejo representa el esquema de los datos de tipo struct. Cambie el nombre de la biblioteca de definiciones de tipos a CustomerInfo. Añada los siguientes elementos a la definición de tipo de datos complejo:

- Elemento name de tipo string
- Elemento age de tipo integer
- Elemento phones de tipo array con elementos de tipo string

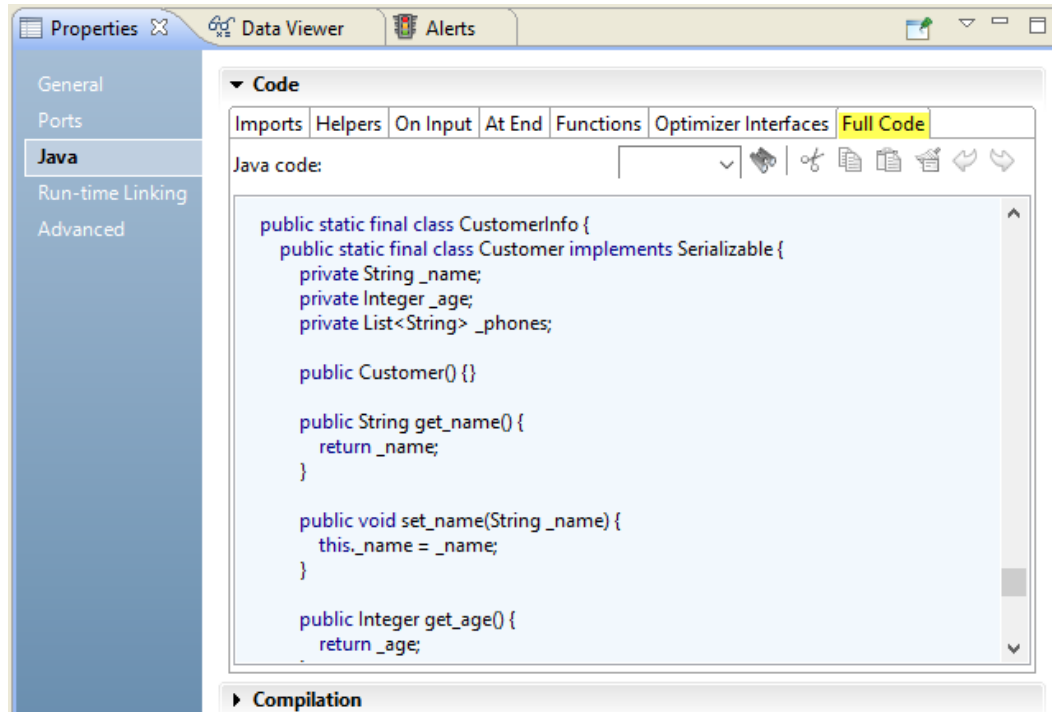
En la siguiente imagen se muestra la definición de tipo de datos complejo en la biblioteca de definiciones de tipos:



En la transformación Java, añada un puerto de salida de estructura y especifique la configuración de tipo del puerto para hacer referencia a la definición de tipo de datos complejo que ha creado. La transformación Java genera una clase Customer con establecedores y captadores para leer y establecer los campos miembro. La clase contiene los siguientes campos miembro:

- _name
- _age
- _phones

En la siguiente imagen se muestra la clase creada para el puerto de estructura en la ficha **Código completo** de la vista **Java**:



```
public static final class CustomerInfo {
    public static final class Customer implements Serializable {
        private String _name;
        private Integer _age;
        private List<String> _phones;

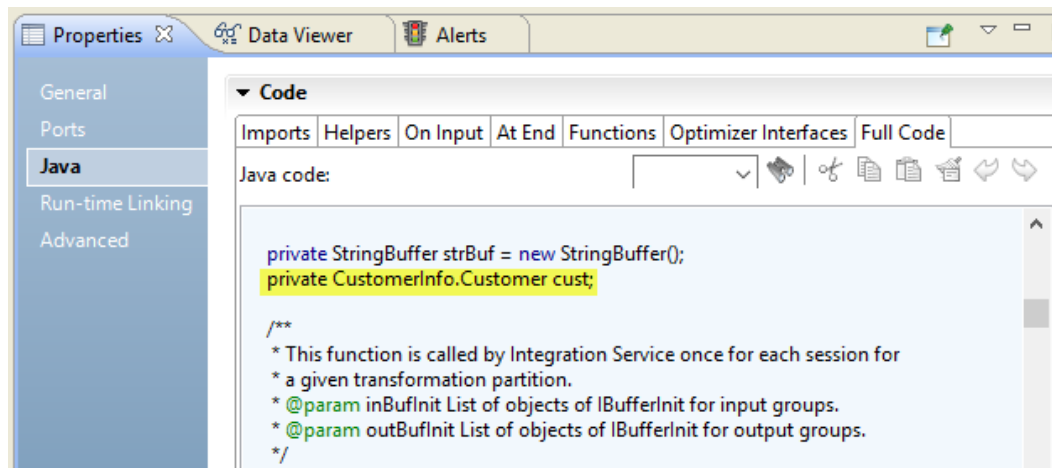
        public Customer() {}

        public String get_name() {
            return _name;
        }

        public void set_name(String _name) {
            this._name = _name;
        }

        public Integer get_age() {
            return _age;
        }
    }
}
```

El tipo de datos de Java del puerto de estructura usa el nombre de la biblioteca de definiciones de tipos y la definición de tipos de datos complejos. En la siguiente imagen se muestra el nombre del tipo de datos de Java CustomerInfo.Customer del campo cust en el código generado:

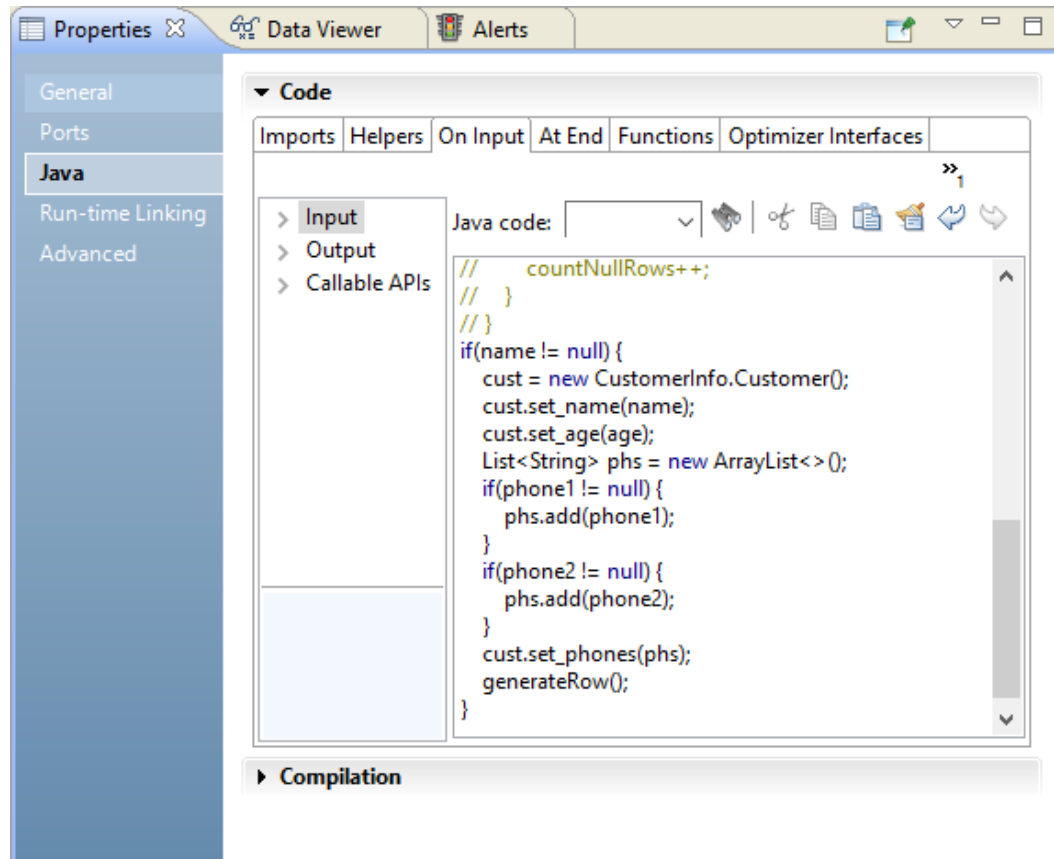


```
private StringBuffer strBuf = new StringBuffer();
private CustomerInfo.Customer cust;
```

/**
 * This function is called by Integration Service once for each session for
 * a given transformation partition.
 * @param inBufInnit List of objects of IBufferInnit for input groups.
 * @param outBufInnit List of objects of IBufferInnit for output groups.
 */

En la vista **Java** de la transformación Java, importe los paquetes Java de terceros, integrados o personalizados que la transformación requiera. Escriba y compile el código Java para convertir los datos sin formato en datos de tipo struct y para quitar la fila de cliente si el nombre de cliente es nulo.

En la siguiente imagen se muestra el código en la ficha **En entrada**:



Valide la asignación y ejecútela en el motor de Spark para escribir los datos transformados en la salida del archivo de Avro.

Transformación Java en un entorno no nativo

El procesamiento de transformaciones Java en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. No compatible.

Transformación Java en el motor de Blaze

Haga lo siguiente para usar archivos .jar externos en una transformación Java:

1. Copie los archivos .jar externos en el directorio de instalación de Informática en el equipo del servicio de integración de datos, en la siguiente ubicación: <directorio de instalación de Informática>/services/shared/jars. A continuación, recicle el servicio de integración de datos.

2. En el equipo que aloja Developer tool donde desarrolla y ejecuta la asignación que contiene la transformación Java:
 - a. Copie los archivos .jar externos en un directorio en el equipo local.
 - b. Edite la transformación Java para que incluya una instrucción de importación que apunte a los archivos .jar locales.
 - c. Actualice la ruta de clase de la transformación Java.
 - d. Compile la transformación.

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

Partición

- Las siguientes restricciones son aplicables a la propiedad **Ámbito** de la transformación:
 - El valor **Transacción** del ámbito de transformación no es válido.
 - Si habilita un puerto de entrada para la clave de partición, el ámbito de la transformación debe establecerse en **Todas las entradas**.
 - La propiedad **Sin estado** debe estar habilitada si el ámbito de transformación son filas.

Transformación Java en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Restricciones generales

La transformación Java es compatible con las siguientes restricciones en el motor de Spark:

- El código Java de la transformación no puede escribir la salida en una salida estándar cuando se inserta lógica de transformación en Hadoop. El código Java puede escribir la salida en un error estándar que aparece en los archivos de registro.
- En relación con los valores de fecha/hora, el motor de Spark admite una precisión de microsegundos como máximo. Si un valor de fecha/hora contiene nanosegundos, los dígitos finales se truncarán.

Particiones

La transformación Java tiene las siguientes restricciones cuando se usa con particiones:

- La propiedad **Admite particiones** debe estar habilitada en la transformación Java. La transformación no se puede ejecutar en una partición.
- Las siguientes restricciones son aplicables a la propiedad **Ámbito** de la transformación:
 - El valor **Transacción** del ámbito de transformación no es válido.
 - Si habilita un puerto de entrada para la clave de partición, el ámbito de la transformación debe establecerse en **Todas las entradas**.
 - La propiedad **Sin estado** debe estar habilitada si el ámbito de transformación son filas.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- Hace referencia a una transformación de búsqueda no conectada desde una expresión de una transformación Java.
- Selecciona un puerto con un tipo de datos complejo como la clave de partición u ordenación.

- Habilita el procesamiento al nanosegundo de tipos date/time y la transformación Java contiene un puerto de tipo de datos complejo con un elemento de tipo date/time. Por ejemplo, un puerto de tipo `array<data/time>` no es válido si se habilita el procesamiento al nanosegundo de tipos date/time.

La asignación genera un error en las siguientes situaciones:

- La transformación Java y la asignación usan modos de precisión diferentes cuando la transformación Java contiene un puerto decimal o un puerto complejo con un elemento de un tipo de datos decimal. Incluso si la alta precisión está habilitada en la asignación, la asignación procesa los datos en el modo de baja precisión en algunas situaciones, como cuando la asignación contiene un puerto complejo con un elemento de un tipo de datos decimal o la asignación es una asignación de transmisión. Si la alta precisión está habilitada tanto en la transformación Java como en la asignación, pero la asignación procesa los datos en modo de baja precisión, se produce un error en la asignación.
- Los caracteres binarios nulos se pasan a un puerto de salida. Para evitar un error de asignación, puede añadir código a la transformación Java que reemplace los caracteres nulos binarios por un carácter alternativo antes de escribir los datos en los puertos de salida.

Usar archivos .jar externos

Haga lo siguiente para usar archivos .jar externos en una transformación Java:

1. Copie los archivos .jar externos en el directorio de instalación de Informática en el equipo del servicio de integración de datos, en la siguiente ubicación:
`<directorio de instalación de Informática>/services/shared/jars`
2. Recicle el servicio de integración de datos.
3. En el equipo que aloja Developer tool donde desarrolla y ejecuta la asignación que contiene la transformación Java:
 - a. Copie los archivos .jar externos en un directorio en el equipo local.
 - b. Edite la transformación Java para que incluya una instrucción de importación que apunte a los archivos .jar locales.
 - c. Actualice la ruta de clase de la transformación Java.
 - d. Compile la transformación.

Transformación de Java en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

El servicio de integración de datos ignora las siguientes propiedades:

- Admite partición. El servicio de integración de datos ignora la propiedad y puede procesar la transformación con varios subprocesos.
- Sin estado. El servicio de integración de datos ignora la propiedad y mantiene el orden de filas de los datos de entrada.

CAPÍTULO 20

Referencia API de transformación de Java

Este capítulo incluye los siguientes temas:

- [Resumen de los métodos API de transformación de Java, 347](#)
- [defineJExpression, 348](#)
- [failSession, 349](#)
- [generateRow, 350](#)
- [getInRowType, 350](#)
- [getMetadata, 351](#)
- [incrementErrorCount, 351](#)
- [invokeJExpression, 352](#)
- [isNull, 353](#)
- [logError, 354](#)
- [logInfo, 354](#)
- [resetNotification, 355](#)
- [setNull, 355](#)
- [storeMetadata, 356](#)

Resumen de los métodos API de transformación de Java

En las fichas de introducción de código de la vista **Java** del editor, puede añadir métodos API al código Java para definir el comportamiento de la transformación.

Para añadir un método API al código, expanda la lista **API invocables** en el navegador de la ficha de introducción de código, y luego haga doble clic en el nombre del método que desee añadir al código.

También puede arrastrar el método desde el navegador hasta el snippet de código Java o especificar manualmente el método API en el snippet de código Java.

Puede añadir los siguientes métodos API al código Java en una transformación de Java:

defineJExpression

Define la expresión de Java.

failSession

Devuelve una excepción con un mensaje de error y la de sesión genera un error.

generateRow

Genera una fila de salida para las transformaciones de Java activas.

getInRowType

Devuelve el tipo de entrada de la fila actual en la transformación.

incrementErrorCount

Incrementa el recuento de errores para la de sesión.

invokeJExpression

Invoca una expresión de Java definida mediante el método defineJExpression.

isNull

Comprueba si hay un valor nulo en una columna de entrada.

logError

Escribe un mensaje de error en el de sesión.

logInfo

Escribe un mensaje informativo en registro de .

resetNotification

Si el equipo del servicio de integración de datos se ejecuta en modo de reinicio, restablece variables que utiliza en el código Java después de la ejecución de una asignación.

setNull

Establece como nulo el valor de una columna de salida en una transformación de Java activa o pasiva.

defineJExpression

Define una expresión, incluyendo la cadena de la expresión y los parámetros de entrada. Los argumentos para el método defineJExpression incluyen una matriz de objetos JExprParamMetadata que contiene los parámetros de entrada y un valor de cadena que define la sintaxis de la expresión.

Utilice la siguiente sintaxis:

```
defineJExpression(  
    String expression,  
    Object[] paramMetadataArray  
);
```

La tabla siguiente describe los parámetros:

Parámetro	Tipo	Tipo de datos	Descripción
expresión	Entrada	String	Cadena que representa la expresión.
paramMetadataArray	Entrada	Object[]	Matriz de objetos JExprParamMetadata que contiene los parámetros de entrada para la expresión.

Se puede añadir el método `defineExpression` al código Java en cualquier ficha de entrada de código, excepto en las fichas **Importaciones** y **Funciones**.

Para utilizar el método `defineJExpression`, debe crear una instancia de una matriz de objetos `JExprParamMetadata` que representan los parámetros de entrada de la expresión. Defina los valores de metadatos para los parámetros y pase la matriz en forma de parámetro al método `defineJExpression`.

Por ejemplo, el siguiente código Java crea una expresión para buscar el valor de dos cadenas:

```
JExprParamMetadata params[] = new JExprParamMetadata[2];
params[0] = new JExprParamMetadata(EDataType.STRING, 20, 0);
params[1] = new JExprParamMetadata(EDataType.STRING, 20, 0);
defineJExpression(":lkp.mylookup(x1,x2)",params);
```

Nota: Debe numerar los parámetros que se pasan a la expresión de forma consecutiva, y comenzar los parámetros con la letra x. Por ejemplo, para pasar tres parámetros a una expresión, nombre los parámetros x1, x2 y x3.

failSession

Devuelve una excepción con un mensaje de error y genera un fallo en la de la sesión.

Utilice la siguiente sintaxis:

```
failSession(String errorMessage);
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
errorMessage	Entrada	String	Cadena del mensaje de error

Utilice el método `failSession` para finalizar la de la sesión. No utilice el método `failSession` en un bloque `try/catch` en una ficha de introducción de código.

Se puede añadir el método `failSession` al código Java en cualquier ficha de entrada de código, excepto en las fichas **Importaciones** y **Funciones**.

El siguiente código Java muestra cómo probar el puerto `input1` para un valor nulo y hacer fallar la de la sesión si es nulo:

```
if(isNull("input1")) {
    failSession("Cannot process a null value for port input1.");
}
```

generateRow

Genera una fila de salida para transformaciones de Java activas.

Utilice la siguiente sintaxis:

```
generateRow();
```

Cuando se llama al método generateRow, la transformación de Java genera una fila de salida utilizando el valor actual de las variables del puerto de salida. Si se desea generar varias filas correspondientes a una fila de entrada, puede llamar al método generateRow más de una vez para cada fila de entrada. Si no se utiliza el método generateRow en una transformación de Java activa, la transformación no generará filas de salida.

Se puede añadir el método generateRow al código Java en cualquier ficha de introducción de código, excepto en las fichas **Importaciones y Funciones**.

Solamente puede llamar al método generateRow en transformaciones activas. Si se llama al método generateRow en una transformación pasiva, el servicio de integración de datos genera un error.

Utilice el siguiente código Java para generar una fila de salida, modificar los valores de los puertos de salida y generar otra fila de salida:

```
// Generate multiple rows.
if(!isNull("input1") && !isNull("input2"))
{
    output1 = input1 + input2;
    output2 = input1 - input2;
}
generateRow();
// Generate another row with modified values.
output1 = output1 * 2;
output2 = output2 * 2;
generateRow();
```

getInRowType

Devuelve el tipo de entrada de la fila actual en la transformación. El método devuelve un valor de inserción, actualización, eliminación o rechazo.

Utilice la siguiente sintaxis:

```
rowType getInRowType();
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
rowType	Salida	String	Devuelve el tipo de estrategia de actualización, que es uno de los siguientes valores: <ul style="list-style-type: none">- DELETE- INSERT- REJECT- UPDATE

Puede añadir el método getInRowType al código Java en la ficha de introducción de código **En entrada**.

Puede utilizar el método `getInRowType` en transformaciones activas configuradas para definir la estrategia de actualización. Si se invoca este método en una transformación activa que no está configurada para definir la estrategia de actualización, el de la sesión genera un error.

getMetadata

Recupera los metadatos de la transformación de Java en tiempo de ejecución. El método `getMetadata` recupera los metadatos que haya guardado con el método `storeMetadata`, como una condición de filtro que el optimizador pasa a la transformación de Java en la función `pushFilter`.

Utilice la siguiente sintaxis:

```
getMetadata (String key);
```

La tabla siguiente describe los parámetros:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
clave	Entrada	String	Identifica los metadatos. El método <code>getMetadata</code> emplea la clave para determinar qué metadatos recuperar.

Puede añadir el método `getMetadata` al código Java en las siguientes fichas de introducción de código:

- Aplicación auxiliar
- En entrada
- Al final
- Interfaces del optimizador
- Funciones

Puede configurar el método `getMetadata` para recuperar las condiciones de filtro para la optimización de inserción. El método `getMetadata` puede recuperar cada condición de filtro que haya almacenado desde el optimizador.

```
// Retrieve a filter condition  
String mydata = getMetadata ("FilterKey");
```

incrementErrorCount

Incrementa el recuento de errores. Si el recuento de errores alcanza el umbral de error, la de la sesión genera un error.

Utilice la siguiente sintaxis:

```
incrementErrorCount (int nErrors);
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
Número de errores	Entrada	Integer	Número por el cual incrementar el recuento de errores.

Se puede añadir el método `incrementErrorCount` al código Java en cualquier ficha de introducción de código, excepto en las fichas **Importaciones** y **Funciones**.

El siguiente código Java muestra cómo incrementar el recuento de errores si el puerto de entrada de una transformación tiene un valor nulo:

```
// Check if input employee id and name is null.
if (isNull ("EMP_ID_INP") || isNull ("EMP_NAME_INP"))
{
    incrementErrorCount(1);
    // if input employee id and/or name is null, don't generate a output row for this
    input row
    generateRow = false;
}
```

invokeJExpression

Invoca una expresión y devuelve el valor para la expresión.

Utilice la siguiente sintaxis:

```
(datatype) invokeJExpression(
    String expression,
    Object[] paramMetadataArray);
```

Los parámetros de entrada para el método `invokeJExpression` son un valor de cadena que representa la expresión y una matriz de objetos que contienen parámetros de entrada de expresión.

La tabla siguiente describe los parámetros:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
Expresión	Entrada	String	Cadena que representa la expresión.
<code>paramMetadataArray</code>	Entrada	Object[]	Matriz de objetos que contiene los parámetros de entrada para la expresión.

Se puede añadir el método `invokeJExpression` al código Java en cualquier ficha de introducción de código, excepto las fichas **Importaciones** y **Funciones**.

Tenga en cuenta las siguientes reglas y directrices cuando utilice el método `invokeJExpression`:

- Tipo de datos de devolución. El tipo de datos de devolución del método `invokeJExpression` es un objeto. Debe lanzar el valor de devolución de la función con el tipo de datos apropiado. Puede devolver valores con los tipos de datos Integer, Double, String y byte[].
- Tipo de fila. El tipo de fila para valores de devolución del método `invokeJExpression` es INSERT.

Utilice la interfaz avanzada para utilizar un tipo de fila diferente para el valor de devolución.

- Valores nulos. Si se pasa un valor nulo como parámetro o si el valor de devolución del método `invokeJExpression` es `NULL`, el valor se trata como un indicador nulo.

Por ejemplo, si el valor de devolución de una expresión es `NULL` y el tipo de datos de devolución es `String`, se devuelve una cadena con un valor nulo.

- Tipo de datos `Date`. Debe convertir los parámetros de entrada con un tipo de datos `Date` al tipo de datos `String`.

Para utilizar la cadena en una expresión como tipo de datos `Date`, utilice la función `to_date()` para convertir la cadena a un tipo de datos `Date`.

Asimismo, debe lanzar el tipo de devolución de cualquier expresión que devuelva un tipo de datos `Date` como tipo de datos `String`.

El siguiente ejemplo concatena las cadenas "John" y "Smith" y devuelve la cadena "John Smith":

```
(String)invokeJExpression("concat(x1,x2)", new Object [] { "John ", "Smith" });
```

Nota: Debe numerar los parámetros que pasa a la expresión de forma consecutiva e iniciar el parámetro con la letra x. Por ejemplo, para pasar tres parámetros a una expresión, nombre los parámetros x1, x2 y x3.

isNull

Comprueba el valor de una columna de entrada para un valor nulo.

Utilice la siguiente sintaxis:

```
Boolean isNull(String strColName);
```

La tabla siguiente describe el parámetro:

Parámetros	Tipo de parámetro	Tipo de datos	Descripción
strColName	Entrada	String	Nombre de una columna de entrada.

Puede añadir el método `isNull` al código Java en la ficha de introducción de código **En entrada**.

El siguiente código Java muestra cómo comprobar si el valor de la columna de entrada `SALARY` es nulo antes de añadirlo a la variable de instancia `totalSalaries`:

```
// if value of SALARY is not null
if (!isNull("SALARY")) {
    // add to totalSalaries
    TOTAL_SALARIES += SALARY;
}
```

También puede utilizar el siguiente código Java para conseguir los mismos resultados:

```
// if value of SALARY is not null
String strColName = "SALARY";
if (!isNull(strColName)) {
    // add to totalSalaries
    TOTAL_SALARIES += SALARY;
}
```

logError

Escribe un mensaje de error en el registro de la .

Utilice la siguiente sintaxis:

```
logError(String msg);
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
msg	Entrada	String	Cadena del mensaje de error

Se puede añadir el método `logError` al código Java en cualquier ficha de introducción de código, excepto en las fichas **Importaciones** y **Funciones**.

El siguiente código Java muestra cómo registrar un error cuando el puerto de entrada es nulo:

```
// check BASE_SALARY
if (isNull("BASE_SALARY")) {
    logError("Cannot process a null salary field.");
}
```

Cuando se ejecuta el código aparece el siguiente mensaje en el registro de la :

```
[JTX_1013] [ERROR] Cannot process a null salary field.
```

logInfo

Escribe un mensaje informativo en el registro de la .

Utilice la siguiente sintaxis:

```
logInfo(String msg);
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
msg	Entrada	String	Cadena de mensaje informativo.

Se puede añadir el método `logInfo` al código Java en cualquier ficha de introducción de código, excepto en las fichas **Importaciones** y **Funciones**.

El siguiente código Java muestra cómo escribir un mensaje en el registro de la después de que una transformación de Java procesa un umbral de mensajes de 1000 filas:

```
if (numRowsProcessed == messageThreshold) {
    logInfo("Processed " + messageThreshold + " rows.");
}
```

resetNotification

Si el equipo del servicio de integración de datos se ejecuta en modo de reinicio, restablece variables que utiliza en el código Java después de la ejecución de una asignación.

En el modo de reinicio, el servicio de integración de datos no se desinicializa, sino que se restablece tras una solicitud de modo que el servicio de integración de datos pueda procesar la siguiente solicitud.

Para una transformación de Java, utilice el método `resetNotification` para restablecer variables en el código Java después de ejecutar una sesión.

Utilice la siguiente sintaxis:

```
public int resetNotification(IGroup group) {  
    return EStatus.value;  
}
```

La tabla siguiente describe los parámetros:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
int	Salida	EStatus.value	Valor de devolución, donde <i>value</i> tiene uno de los siguientes valores: <ul style="list-style-type: none">- SUCCESS. Correcto.- FAILURE. Anomalía.- NOIMPL. No implementado.
group	Entrada	IGroup	El grupo de entrada.

Puede añadir el método `resetNotification` al código Java en la ficha de introducción de código, en la ficha **Aplicaciones auxiliares**.

El método `resetNotification` no aparece en la lista de API invocables.

Por ejemplo, en el supuesto de que el código Java declare una variable estática llamada `out5_static` la inicialice con el valor 1, el siguiente código Java restablece la variable `out5_static` a 1 después de la siguiente ejecución de asignación:

```
public int resetNotification(IGroup group) {  
    out5_static=1;  
    return EStatus.SUCCESS;  
}
```

Este método no es necesario. No obstante, si el servicio de integración de datos se ejecuta en modo de reinicio y la asignación contiene una transformación de Java que no implementa el método `resetNotification`, el mensaje de advertencia `JSDK_42075` aparecerá en el registro.

setNull

Establece el valor de una columna de salida como nulo en una transformación de Java activa o pasiva.

Utilice la siguiente sintaxis:

```
setNull(String strColName);
```

La tabla siguiente describe el parámetro:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
strColName	Entrada	String	Nombre de una columna de salida.

El método `setNull` establece el valor de una columna de salida como nulo en una transformación de Java activa o pasiva. Después de establecer una columna de salida como nula, no se puede modificar el valor hasta que no se haya generado una fila de salida.

Se puede añadir el método `setNull` al código Java en cualquier ficha de introducción de código, excepto en las fichas **Importaciones** y **Funciones**.

El siguiente código Java muestra cómo comprobar el valor de una columna de entrada y definir como nulo el valor correspondiente de una columna de salida:

```
// check value of Q3RESULTS input column
if(isNull("Q3RESULTS")) {
    // set the value of output column to null
    setNull("RESULTS");
}
```

También puede utilizar el siguiente código Java para conseguir los mismos resultados:

```
// check value of Q3RESULTS input column
String strColName = "Q3RESULTS";
if(isNull(strColName)) {
    // set the value of output column to null
    setNull(strColName);
}
```

storeMetadata

Almacena los metadatos de la transformación de Java que puede recuperar en tiempo de ejecución con el método `getMetadata`.

Utilice la siguiente sintaxis:

```
storeMetadata (String key String data);
```

La tabla siguiente describe los parámetros:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
clave	Entrada	String	Identifica los metadatos. El método <code>storeMetadata</code> requiere una clave para identificar los metadatos. Defina la clave como cualquier cadena.
datos	Entrada	String	Los datos que desee almacenar como metadatos de transformación de Java.

Puede añadir el método `storeMetadata` al código Java a las siguientes fichas de introducción de código:

- Aplicación auxiliar
- En entrada

- Al final
- Interfaces del optimizador
- Funciones

Puede configurar el método `storeMetadata` en una transformación activa para aceptar las condiciones de filtro para la optimización de inserción. El método `storeMetadata` almacena una condición de filtro que el optimizador inserta desde la asignación a la transformación de Java.

```
// Store a filter condition
storeMetadata ("FilterKey", condition);
```

CAPÍTULO 21

Expresiones de Java

Este capítulo incluye los siguientes temas:

- [Resumen de las expresiones de Java, 358](#)
- [Cómo utilizar el cuadro de diálogo Definir función para definir una expresión, 359](#)
- [Cómo trabajar con la interfaz simple, 361](#)
- [Cómo trabajar con la interfaz avanzada, 363](#)
- [Referencia API de la clase JExpression, 367](#)

Resumen de las expresiones de Java

Puede invocar expresiones de en una transformación de Java mediante el lenguaje de programación Java.

Utilice expresiones para ampliar las funciones de una transformación de Java. Por ejemplo, puede invocar una expresión en una transformación de Java para buscar los valores de puertos de entrada y de salida, o buscar los valores de variables de una transformación de Java.

Para invocar expresiones en una transformación de Java, genere el código Java o utilice métodos API de transformación de Java para invocar la expresión. Invoque la expresión y utilice el resultado de la expresión en la ficha de introducción de código apropiada. Puede generar el código Java que invoque una expresión o utilizar métodos API para escribir el código Java que invoque la expresión.

La siguiente tabla describe los métodos que se pueden usar para crear e invocar expresiones en una transformación de Java:

Método	Descripción
Definir función	Permite crear una función que invoque una expresión y genere el código para una expresión.
Interfaz simple	Permite invocar un método API individual para invocar una expresión y obtener el resultado de la expresión.
Interfaz avanzada	Permite definir la expresión, invocar la expresión y utilizar el resultado de la expresión. Si está familiarizado con la programación orientada a objetos y desea ejercer un mayor control en la invocación de la expresión, utilice la interfaz avanzada.

Tipos de función de expresión

Puede crear expresiones para una transformación de Java usando el cuadro de diálogo **Definir función**, o utilizando la interfaz simple o la avanzada.

Puede especificar como parámetros de entrada expresiones que utilizan variables de puerto de entrada o de salida, o variables en el código Java.

Si utiliza el cuadro de diálogo **Definir función**, puede validar la expresión antes de utilizarla en una transformación de Java.

Puede invocar los siguientes tipos de funciones de expresión en una transformación de Java:

Tipo de función de expresión	Descripción
Funciones de lenguaje de transformación	Funciones similares a SQL, diseñadas para manejar expresiones comunes.
Funciones definidas por el usuario	Funciones que se crean en Developer Tool basadas en funciones del lenguaje de transformación.
Funciones personalizadas	Funciones que se crean con la API Custom Function.

También puede usar transformaciones no conectadas y variables incorporadas. Por ejemplo, puede utilizar una transformación de búsqueda no conectada en una expresión.

Cómo utilizar el cuadro de diálogo Definir función para definir una expresión

Cuando se define una expresión de Java se configura la función, se crea la expresión, y se genera el código que invoca la expresión.

Puede definir la función y crear la expresión en el cuadro de diálogo **Definir función**.

Para crear una función de expresión y utilizar la expresión en una transformación de Java, realice las siguientes tareas de alto nivel:

1. Configure la función que invoca la expresión, incluyendo el nombre, descripción y parámetros de la función. Utilice parámetros de función cuando cree la expresión.
2. Cree la sintaxis de la expresión y válidela.
3. Genere el código Java que invoca la expresión.

Developer pone el código en la ficha de introducción de código **Funciones**.

Después de generar el código Java, invoque la función generada en la ficha de introducción de código apropiada para invocar una expresión u obtener un objeto JExpression, en función de si utiliza la interfaz simple o la avanzada.

Nota: Para validar una expresión cuando se crea la expresión debe usar el cuadro de diálogo **Definir función**.

Paso 1. Configurar la función

Configure el nombre y descripción de la función y los parámetros de entrada de la función de Java que invoca la expresión.

Tenga en cuenta las siguientes reglas y directrices cuando configure la función:

- Utilice un nombre de función único que no esté en conflicto con una función Java existente en la transformación o con palabras clave de Java reservadas.
- Debe configurar el nombre del parámetro, el tipo de datos de Java, la precisión y la escala. Los parámetros de entrada son los valores que se pasan cuando se llama la función del código Java para la transformación.
- Para pasar el tipo de datos Date a una expresión, utilice el tipo de datos String para el parámetro de entrada.

Si una expresión devuelve el tipo de datos Date, puede utilizar el valor de devolución como tipo de datos String en la interfaz simple y el tipo de datos String o Long en la interfaz avanzada.

Paso 2. Crear y validar la expresión

Cuando cree la expresión, utilice los parámetros configurados para la función.

También puede utilizar funciones de lenguaje de transformación, funciones personalizadas u otras funciones definidas por el usuario en la expresión. Puede crear y validar la expresión en los cuadros de diálogo **Definir función** .

Paso 3. Generar código Java para la expresión

Después de configurar la función, sus parámetros, y de definir y validar la expresión, puede generar el código Java que invoca la expresión.

Developer coloca el código Java generado en la ficha de introducción de código **Funciones**. Utilice el código Java generado para llamar las funciones que invocan la expresión en las fichas de introducción de código. Se puede generar código Java simple o avanzado.

Después de generar el código Java que invoca una expresión, no se puede editar la expresión y volver a validarla. Para modificar una expresión después de generar el código se debe volver a crear la expresión.

Cómo crear una expresión y generar código Java mediante el cuadro de diálogo Definir función

Puede crear una función que invoque una expresión en el cuadro de diálogo **Definir función**.

Realice los siguientes pasos para crear una función que invoque una expresión:

1. En Developer, abra una transformación de Java o cree una nueva transformación de Java.
2. En la ficha **Código Java**, haga clic en **Nueva función**.
Aparece el cuadro de diálogo **Definir función**.
3. Escriba un nombre de función.
4. Si lo desea, puede escribir una descripción para la expresión.
Puede escribir hasta 2.000 caracteres
5. Cree los argumentos para la función.
Cuando cree los argumentos, configure el nombre, tipo de datos, precisión y escala del argumento.

6. En la ficha **Expresión**, cree una expresión con los argumentos creados.
7. Para validar la expresión, haga clic en **Validar**.
8. Si lo desea, escriba la expresión en el cuadro **Expresión**. Luego, haga clic en **Validar** para validar la expresión.
9. Para generar el código Java mediante la interfaz avanzada, seleccione la opción **Generar código avanzado**. Luego, haga clic en **Generar**.
Developer genera la función para invocar la expresión en la ficha de introducción de código **Funciones**.

Plantillas de expresiones de Java

Se puede generar código Java para una expresión utilizando el código Java simple o avanzado para expresiones.

El código Java para expresiones se genera conforme la plantilla de la expresión.

El siguiente ejemplo muestra la plantilla para una expresión de Java generada para código Java simple:

```
Object function_name (Java datatype x1[,
                        Java datatype x2 ...] )
                        throws SDK Exception
{
return (Object)invokeJExpression( String expression,
                                new Object [] { x1[, x2, ... ]} );
}
```

El siguiente ejemplo muestra la plantilla para una expresión de Java generada utilizando la interfaz avanzada:

```
JExpression function_name () throws SDKException
{
    JExprParamMetadata params[] = new JExprParamMetadata[number of parameters];
    params[0] = new JExprParamMetadata (
        EDataType.STRING, // data type
        20, // precision
        0 // scale
    );
    ...
    params[number of parameters - 1] = new JExprParamMetadata (
        EDataType.STRING, // data type
        20, // precision
        0 // scale
    );
    ...
    return defineJExpression(String expression,params);
}
```

Cómo trabajar con la interfaz simple

Utilice el método API de Java `invokeJExpression` para invocar una expresión en la interfaz simple.

invokeJExpression

Invoca una expresión y devuelve el valor para la expresión.

Utilice la siguiente sintaxis:

```
(datatype) invokeJExpression(  
    String expression,  
    Object[] paramMetadataArray);
```

Los parámetros de entrada para el método `invokeJExpression` son un valor de cadena que representa la expresión y una matriz de objetos que contienen parámetros de entrada de expresión.

La tabla siguiente describe los parámetros:

Parámetro	Tipo de parámetro	Tipo de datos	Descripción
expresión	Entrada	String	Cadena que representa la expresión.
paramMetadataArray	Entrada	Object[]	Matriz de objetos que contiene los parámetros de entrada para la expresión.

Se puede añadir el método `invokeJExpression` al código Java en cualquier ficha de introducción de código, excepto las fichas **Importaciones** y **Funciones**.

Tenga en cuenta las siguientes reglas y directrices cuando utilice el método `invokeJExpression`:

- Tipo de datos de devolución. El tipo de datos de devolución del método `invokeJExpression` es un objeto. Debe lanzar el valor de devolución de la función con el tipo de datos apropiado. Puede devolver valores con los tipos de datos `Integer`, `Double`, `String` y `byte[]`.
- Tipo de fila. El tipo de fila para valores de devolución del método `invokeJExpression` es `INSERT`. Utilice la interfaz avanzada para utilizar un tipo de fila diferente para el valor de devolución.
- Valores nulos. Si se pasa un valor nulo como parámetro o si el valor de devolución del método `invokeJExpression` es `NULL`, el valor se trata como un indicador nulo. Por ejemplo, si el valor de devolución de una expresión es `NULL` y el tipo de datos de devolución es `String`, se devuelve una cadena con un valor nulo.
- Tipo de datos `Date`. Debe convertir los parámetros de entrada con un tipo de datos `Date` al tipo de datos `String`. Para utilizar la cadena en una expresión como tipo de datos `Date`, utilice la función `to_date()` para convertir la cadena a un tipo de datos `Date`. Asimismo, debe lanzar el tipo de devolución de cualquier expresión que devuelva un tipo de datos `Date` como tipo de datos `String`.

Nota: Debe numerar los parámetros que pasa a la expresión de forma consecutiva e iniciar el parámetro con la letra `x`. Por ejemplo, para pasar tres parámetros a una expresión, nombre los parámetros `x1`, `x2` y `x3`.

Ejemplo de interfaz simple

Puede definir y llamar expresiones que usan el método API `invokeJExpression` en las fichas de introducción de código **Auxiliares** y **En entrada**.

El siguiente ejemplo muestra cómo realizar una búsqueda en los puertos de entrada `NAME` y `ADDRESS` de una transformación de Java y asignar el valor de devolución al puerto de salida `COMPANY_NAME`.

Escriba el siguiente código en la ficha de introducción de código **En entrada**:

```
COMPANY_NAME = (String) invokeJExpression(":lkp.my_lookup(X1,X2)", new Object []
{str1 ,str2} );
generateRow();
```

Cómo trabajar con la interfaz avanzada

En la interfaz avanzada se pueden usar métodos API orientados al objeto para definir, invocar y obtener el resultado de una expresión.

La siguiente tabla describe las clases y métodos API que están disponibles en la interfaz avanzada:

Clase o método API	Descripción
Clase EDataType	Enumera los tipos de datos para una expresión.
Clase JExprParamMetadata	Contiene los metadatos para cada uno de los parámetros en una expresión. Los metadatos de un parámetro incluyen el tipo de datos, la precisión y la escala.
Método API defineJExpression	Define la expresión. Incluye la cadena y los parámetros de la expresión .
Método API invokeJExpression	Invoca una expresión.
clase JExpression	Contiene los métodos para crear, invocar y obtener los metadatos y obtener el resultado de una expresión, y también para comprobar el tipo de datos de devolución.

Cómo invocar una expresión con la interfaz avanzada

Utilice la interfaz avanzada para definir, invocar y obtener el resultado de una expresión.

1. En las fichas de introducción de código **Aplicaciones auxiliares** o **En entrada**, cree una instancia de la clase JExprParamMetadata para cada de parámetro para la expresión y defina el valor de los metadatos. Si lo desea, puede crear una instancia del objeto JExprParamMetadata en el método defineJExpression.
2. Utilice el método defineJExpression para obtener el objeto JExpression para la expresión.
3. En la ficha de introducción de código apropiada, invoque la expresión con el método invokeJExpression.
4. Compruebe el resultado del valor de devolución o con el método isResultNull.
5. Puede obtener el tipo de datos del valor de devolución o los metadatos del valor de devolución con los métodos getResultDataType y getResultMetadata.
6. Obtenga el resultado de la expresión utilizando el método API apropiado. Puede utilizar los métodos getInt, getDouble, getStringBuffer y getBytes.

Reglas y directrices para trabajar con la interfaz avanzada

Cuando se trabaja con las interfaces avanzadas, debe tener en cuenta ciertas reglas y directrices.

Tenga en cuenta las siguientes reglas y directrices:

- Si se pasa un valor nulo como parámetro, o si el resultado de una expresión es nulo, el valor se tratará como un indicador nulo. Por ejemplo, si el resultado de una expresión es nulo y el tipo de datos de devolución es String, se devuelve una cadena con valor nulo. Puede comprobar el resultado de una expresión mediante el método `isResultNull`.
- Debe convertir parámetros de entrada con el tipo de datos Date al tipo de datos String antes de poder usarlos en una expresión. Para utilizar la cadena en una expresión como tipos de datos Date, utilice la función `to_date()` para convertir la cadena a un tipo de datos Date.

Puede obtener el resultado de una expresión que devuelve un tipo de datos Date como tipo de datos String o Long.

Para obtener el resultado de una expresión que devuelve un tipo de datos Date como tipo de datos String, utilice el método `getStringBuffer`. Para obtener el resultado de una expresión que devuelve un tipo de datos Date como tipo de datos Long, utilice el método `getLong`.

Clase EDataType

Enumera los tipos de datos de Java utilizados en las expresiones. Obtiene el tipo de datos de devolución de una expresión o asigna el tipo de datos para un parámetro en el objeto `JExprParamMetadata`. No es necesario crear una instancia de la clase `EDataType`.

Las siguiente tabla muestra los valores enumerados para los tipos de datos de Java en expresiones:

Tipo de datos	Valor enumerado
INT	1
DOUBLE	2
STRING	3
BYTE_ARRAY	4
DATE_AS_LONG	5

El siguiente ejemplo de código Java muestra cómo utilizar la clase `EDataType` para asignar un tipo de datos String a un objeto `JExprParamMetadata`:

```
JExprParamMetadata params[] = new JExprParamMetadata[2];
params[0] = new JExprParamMetadata (
    EDataType.STRING, // data type
    20, // precision
    0 // scale
);
...
```

Clase JExprParamMetadata

Crea una instancia de un objeto que representa los parámetros para una expresión y define los metadatos para los parámetros.

Utilice una matriz de objetos JExprParamMetadata como entrada del método defineJExpression para definir los metadatos para los parámetros de entrada. Puede crear una instancia del objeto JExprParamMetadata en la ficha de introducción de código **Funciones** o en defineJExpression.

Utilice la siguiente sintaxis:

```
JExprParamMetadata paramMetadataArray[] = new JExprParamMetadata[numberOfParameters];
paramMetadataArray[0] = new JExprParamMetadata(datatype, precision, scale);
...
paramMetadataArray[numberOfParameters - 1] = new JExprParamMetadata(datatype, precision,
scale);;
```

La tabla siguiente describe los argumentos:

Argumento	Tipo de argumento	Tipo de datos de argumento	Descripción
datatype	Entrada	EDataType	Tipo de datos del parámetro.
precision	Entrada	Entero	Precisión del parámetro.
scale	Entrada	Entero	Escala del parámetro.

Por ejemplo, utilice el siguiente código Java para crear una instancia de dos objetos JExprParamMetadata con tipo de datos String, precisión 20 y escala 0:

```
JExprParamMetadata params[] = new JExprParamMetadata[2];
params[0] = new JExprParamMetadata(EDataType.STRING, 20, 0);
params[1] = new JExprParamMetadata(EDataType.STRING, 20, 0);
return defineJExpression(":LKP.LKP_addresslookup(X1,X2)", params);
```

defineJExpression

Define una expresión, incluyendo la cadena de la expresión y los parámetros de entrada. Los argumentos para el método defineJExpression incluyen una matriz de objetos JExprParamMetadata que contiene los parámetros de entrada y un valor de cadena que define la sintaxis de la expresión.

Utilice la siguiente sintaxis:

```
defineJExpression(
    String expression,
    Object[] paramMetadataArray
);
```

La tabla siguiente describe los parámetros:

Parámetro	Tipo	Tipo de datos	Descripción
Expresión	Entrada	String	Cadena que representa la expresión.
paramMetadataArray	Entrada	Object[]	Matriz de objetos JExprParamMetadata que contiene los parámetros de entrada para la expresión.

Se puede añadir el método defineExpression al código Java en cualquier ficha de entrada de código, excepto en las fichas **Importaciones** y **Funciones**.

Para utilizar el método `defineJExpression`, debe crear una instancia de una matriz de objetos `JExprParamMetadata` que representan los parámetros de entrada de la expresión. Defina los valores de metadatos para los parámetros, y pase la matriz en forma de parámetro al método `defineJExpression`.

Por ejemplo, el siguiente código Java crea una expresión para buscar el valor de dos cadenas:

```
JExprParamMetadata params[] = new JExprParamMetadata[2];
params[0] = new JExprParamMetadata(EDatatype.STRING, 20, 0);
params[1] = new JExprParamMetadata(EDatatype.STRING, 20, 0);
defineJExpression(":lkp.mylookup(x1,x2)",params);
```

Nota: Debe numerar los parámetros que se pasan a la expresión de forma consecutiva y comenzar los parámetros con la letra x. Por ejemplo, para pasar tres parámetros a una expresión, nombre los parámetros x1, x2 y x3.

Clase JExpression

Contiene métodos que permiten crear e invocar una expresión, devolver el valor de una expresión y comprobar el tipo de datos de devolución.

La siguiente tabla enumera los métodos de la clase `JExpression`:

Nombre de método	Descripción
<code>invoke</code>	Invoca una expresión.
<code>getResultDataType</code>	Devuelve el tipo de datos del resultado de la expresión.
<code>getResultMetadata</code>	Devuelve los metadatos del resultado de la expresión.
<code>isResultNull</code>	Comprueba el valor del resultado de un resultado de expresión.
<code>getInt</code>	Devuelve el valor del resultado de una expresión como un tipo de datos Integer.
<code>getDouble</code>	Devuelve el valor del resultado de una expresión como un tipo de datos Double.
<code>getStringBuffer</code>	Devuelve el valor del resultado de una expresión como un tipo de datos String.
<code>getBytes</code>	Devuelve el valor del resultado de una expresión como un tipo de datos byte[].

Ejemplo de interfaz avanzada

Puede utilizar la interfaz avanzada para crear e invocar una expresión de búsqueda en una transformación de Java.

El siguiente código Java de ejemplo muestra cómo crear una función que llama una expresión y cómo invocar la expresión para obtener el valor de devolución. Este ejemplo pasa los valores de dos puertos de entrada con un tipo de datos String, NAME y COMPANY, a la función `myLookup`. La función `myLookup` utiliza una expresión de búsqueda para buscar el valor del puerto de salida ADDRESS.

Nota: Este ejemplo asume que tiene una transformación de búsqueda no conectada llamada `LKP_addresslookup`.

Utilice el siguiente código Java en la ficha **Aplicaciones auxiliares**:

```
JExpression addressLookup() throws SDKException
{
    JExprParamMetadata params[] = new JExprParamMetadata[2];
    params[0] = new JExprParamMetadata (
        EDataType.STRING,          // data type
        50,                        // precision
        0                          // scale
    );
    params[1] = new JExprParamMetadata (
        EDataType.STRING,          // data type
        50,                        // precision
        0                          // scale
    );
    return defineJExpression(":LKP.LKP_addresslookup(X1,X2)",params);
}
JExpression lookup = null;
boolean isJExprObjCreated = false;
```

Utilice el siguiente código Java en la ficha **En entrada** para invocar la expresión y devolver el valor del puerto ADDRESS:

```
...
if(!isJExprObjCreated)
{
    lookup = addressLookup();
    isJExprObjCreated = true;
}
lookup = addressLookup();
lookup.invoke(new Object [] {NAME,COMPANY}, ERowType.INSERT);
EDataType addressDataType = lookup.getResultDataType();
if(addressDataType == EDataType.STRING)
{
    ADDRESS = (lookup.getStringBuffer()).toString();
} else {
    logError("Expression result datatype is incorrect.");
}
...
```

Referencia API de la clase JExpression

La clase JExpression contiene métodos API que permiten crear e invocar una expresión, devolver el valor de una expresión, y comprobar el tipo de datos de devolución.

La clase JExpression contiene los siguientes métodos API:

- getBytes
- getDouble
- getInt
- getLong
- getResultDataType
- getResultMetadata
- getStringBuffer
- invoke
- isResultNull

getBytes

Devuelve el valor del resultado de una expresión como un tipo de datos `byte[]`. Obtiene el resultado de una expresión que cifra datos con la función `AES_ENCRYPT`.

Utilice la siguiente sintaxis:

```
objectName.getBytes();
```

Utilice el siguiente código Java para obtener el resultado de una expresión que cifra los datos binarios con la función `AES_ENCRYPT`, donde `JExprEncryptData` es un objeto `JExpression`:

```
byte[] newBytes = JExprEncryptData.getBytes();
```

getDouble

Devuelve el valor del resultado de una expresión como un tipo de datos `Double`.

Utilice la siguiente sintaxis:

```
objectName.getDouble();
```

Por ejemplo, utilice el siguiente código Java para obtener el resultado de una expresión que devuelve un valor de salario en forma de número doble, donde `JExprSalary` es un objeto `JExpression`:

```
double salary = JExprSalary.getDouble();
```

getInt

Devuelve el valor del resultado de una expresión como un tipo de datos `Integer`.

Utilice la siguiente sintaxis:

```
objectName.getInt();
```

Por ejemplo, utilice el siguiente código Java para obtener el resultado de una expresión que devuelve el número de ID de un empleado en forma de número entero, donde `findEmpID` es un objeto `JExpression`:

```
int empID = findEmpID.getInt();
```

getLong

Devuelve el valor del resultado de una expresión como un tipo de datos `Long`. Obtiene el resultado de una expresión como un tipo de datos `Date`.

Utilice la siguiente sintaxis:

```
objectName.getLong();
```

Utilice el siguiente código Java para obtener el resultado de una expresión que devuelve un valor `Date` como tipo de datos `Long`, donde `JExprCurrentDate` es un objeto `JExpression`:

```
long currDate = JExprCurrentDate.getLong();
```

getResultDataType

Devuelve el tipo de datos del resultado de una expresión. Devuelve un valor de `EDataType`.

Utilice la siguiente sintaxis:

```
objectName.getResultDataType();
```


Utilice el siguiente código Java de ejemplo para invocar una expresión y asignar el tipo de datos del resultado a la variable `dataType`:

```
myObject.invoke(new Object[] { NAME,COMPANY }, ERowType INSERT);
EDataType dataType = myObject.getResultDataType();
```

getResultMetadata

Devuelve los metadatos para el resultado de una expresión. Puede usar `getResultMetadata` para obtener la precisión, la escala y el tipo de datos del resultado de una expresión. Se pueden asignar los metadatos del valor de devolución de una expresión a un objeto `JExprParamMetadata`. Utilice los métodos de objeto `getScale`, `getPrecision` y `getDataType` para recuperar los metadatos del resultado.

Utilice la siguiente sintaxis:

```
objectName.getResultMetadata();
```

Utilice el siguiente código Java de ejemplo para asignar la escala, la precisión y el tipo de datos del valor de devolución de `myObject` a variables:

```
JExprParamMetadata myMetadata = myObject.getResultMetadata();
int scale = myMetadata.getScale();
int prec = myMetadata.getPrecision();
int datatype = myMetadata.getDataType();
```

Nota: El método de objeto `getDataType` devuelve el valor entero del tipo de dato, tal y como se enumera en `EDataType`.

getStringBuffer

Devuelve el valor del resultado de una expresión como un tipo de datos `String`.

Utilice la siguiente sintaxis:

```
objectName.getStringBuffer();
```

Utilice el siguiente código Java para obtener el resultado de una expresión que devuelve dos cadenas concatenadas, donde `JExprConcat` es un objeto `JExpression`:

```
String result = JExprConcat.getStringBuffer();
```

invoke

Invoca una expresión. Los argumentos de `invoke` incluyen un objeto que define los parámetros de entrada, así como el tipo de fila. Debe crear una instancia de un objeto `JExpression` antes de poder usar el método `invoke`. Para el tipo de fila, utilice `ERowType.INSERT`, `ERowType.DELETE` y `ERowType.UPDATE`.

Utilice la siguiente sintaxis:

```
objectName.invoke(
    new Object[] { param1[, ... paramN ]},
    rowType
);
```

En la tabla siguiente se describen los argumentos:

Argumento	Tipo de datos	Entrada/ Salida	Descripción
objectName	JExpression	Entrada	Nombre del objeto JExpression.
parámetros	-	Entrada	Matriz de objetos que contiene los valores de entrada para la expresión.

Por ejemplo, cree una función en la ficha de introducción de código **Funciones** llamada `address_lookup()` que devuelva un objeto `JExpression` que represente la expresión. Utilice el siguiente código para invocar la expresión que utiliza los puertos de entrada `NAME` y `COMPANY`:

```
JExpression myObject = address_lookup();
myObject.invoke(new Object[] { NAME,COMPANY }, ERowType INSERT);
```

isResultNull

Comprueba el valor del resultado de una expresión.

Utilice la siguiente sintaxis:

```
objectName.isResultNull();
```

Utilice el siguiente código Java de ejemplo para invocar una expresión y asignar el valor de devolución de la misma a la dirección variable si el valor de devolución no es nulo:

```
JExpression myObject = address_lookup();
myObject.invoke(new Object[] { NAME,COMPANY }, ERowType INSERT);
if(!myObject.isResultNull()) {
    String address = myObject.getStringBuffer();
}
```

CAPÍTULO 22

Transformación de incorporación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de combinación, 371](#)
- [Propiedades avanzadas de la transformación de combinación, 372](#)
- [Memorias caché de combinación, 373](#)
- [Puertos de transformación de combinación, 374](#)
- [Transformaciones de combinación en asignaciones dinámicas, 375](#)
- [Selectores de puerto en una transformación de combinación, 375](#)
- [Definir una condición de combinación, 378](#)
- [Tipos de unión, 382](#)
- [Entrada ordenada para una transformación de incorporación, 385](#)
- [Cómo unir datos de un mismo origen, 389](#)
- [Cómo bloquear los canales de origen, 391](#)
- [Consejos sobre el rendimiento de la transformación de combinación, 392](#)
- [Reglas y directrices para una transformación de combinación, 393](#)
- [Transformación de combinación en un entorno no nativo, 393](#)

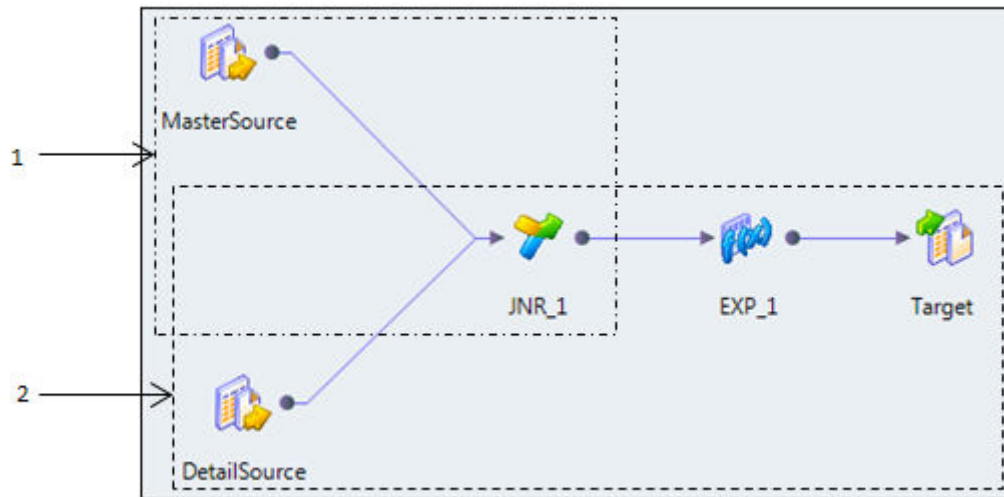
Resumen de la transformación de combinación

La transformación de combinación une datos de origen de dos orígenes heterogéneos relacionados de ubicaciones distintas o de sistemas de archivos distintos. También se pueden unir datos de un mismo origen. La transformación de combinación es una transformación activa de varios grupos.

La transformación de combinación une orígenes con, por lo menos, una columna coincidente. La transformación de combinación utiliza una condición que coincide con uno o más pares de columnas que comparten los dos orígenes.

Los dos canales de entrada incluyen un canal principal y un canal de detalle, o una ramificación principal y una ramificación de detalle. El canal maestro finaliza en la transformación de combinación, mientras que el canal de detalle continúa hacia el destino.

La siguiente imagen muestra los canales principal y de detalle en una asignación con una transformación de combinación:



1. Canal principal
2. Canal de detalle

Para unir más de dos orígenes en una asignación, una la salida de la transformación de combinación con otro canal de origen. Añada transformaciones de combinación a la asignación hasta que haya unido todos los canales de origen.

Propiedades avanzadas de la transformación de combinación

Configure las propiedades para ayudarle a determinar cómo el Servicio de integración de datos procesa los datos para la transformación de incorporación.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Tamaño de la memoria caché de datos de combinación

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de datos para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Tamaño de la memoria caché de índice de incorporación

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de índice para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Directorio de la memoria caché

Directorio en el que el Servicio de integración de datos crea los archivos de memoria caché de índice y de datos. Compruebe que el directorio existe y que tiene suficiente espacio en disco para los archivos de memoria caché.

Introduzca varios directorios separados por punto y coma para aumentar el rendimiento durante la creación de particiones en la memoria caché. La creación de particiones en la memoria caché crea una memoria caché independiente para cada partición que procesa la transformación.

El valor predeterminado es el parámetro del sistema CacheDir. Para esta propiedad, puede configurar otro parámetro del sistema o definido por el usuario.

Entrada ordenada

Indica que los datos de entrada se ordenan previamente por grupos. Elija Entrada ordenada para unir datos ordenados. Utilizar la entrada ordenada puede aumentar el rendimiento.

Orden de clasificación principal

Especifica el orden de clasificación para los datos de origen principales. Elija Ascendente si los datos de origen principal están en orden ascendente. Si elige Ascendente, habilite también la entrada ordenada. El valor predeterminado es Auto.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

TEMAS RELACIONADOS

- [“Tamaño de la memoria caché” en la página 74](#)

Memorias caché de combinación

Cuando se ejecuta una asignación que utiliza una transformación de combinación, el Servicio de integración de datos crea una memoria caché de índice y una memoria caché de datos en la memoria para ejecutar la transformación. Si el Servicio de integración de datos necesita más espacio del que hay disponible en la memoria caché, almacena los datos desbordados en archivos de memoria caché.

Al ejecutar una asignación que utiliza una transformación de combinación, el Servicio de integración de datos lee las filas de los orígenes principal y detallado de forma simultánea y crea memorias caché de índice y de datos basándose en las filas principales. El Servicio de integración de datos realiza la combinación basándose en los datos del origen detallado y los datos principales almacenados en la memoria caché.

El tipo de transformación de combinación determina el número de filas que el Servicio de integración de datos almacena en la memoria caché.

En la siguiente tabla, se describe la información que el Servicio de integración de datos almacena en las memorias caché para los distintos tipos de transformaciones de combinación:

Tipo de transformación de combinación	Memoria caché de índice	Memoria caché de datos
Entrada no ordenada	Almacena todas las filas principales en la condición de combinación con claves de índice únicas.	Almacena todas las filas principales.
Entrada ordenada con diferentes orígenes	Almacena 100 filas principales en la condición de combinación con claves de índice únicas.	Almacena filas principales que corresponden a las filas almacenadas en la memoria caché de índice. Si los datos principales contienen varias filas con la misma clave, el Servicio de integración de datos almacena más de 100 filas en la memoria caché de datos.
Entrada ordenada con el mismo origen	Almacena todas las filas principales y detalladas en la condición de combinación con claves únicas. Almacena las filas detalladas si el Servicio de integración de datos procesa el canal detallado más rápidamente que el canal principal. De lo contrario, almacena las filas principales. El número de filas que almacena depende de la velocidad de procesamiento de los canales principal y detallado. Si un canal procesa sus filas más rápidamente que el otro, el Servicio de integración de datos almacena en la memoria caché todas las filas que ya se han procesado. El servicio conserva las filas en la memoria caché hasta que el otro canal termina de procesar sus filas.	Almacena datos de las filas almacenadas en la memoria caché de índice. Si la memoria caché de índice almacena claves para el canal principal, la memoria caché de datos almacena los datos para el canal principal. Si la memoria caché de índice almacena claves para el canal detallado, la memoria caché de datos almacena los datos para el canal detallado.

Puertos de transformación de combinación

Una transformación de combinación tiene distintos tipos de puerto que determinan cómo el Servicio de integración de datos llevará a cabo la unión.

Una transformación de combinación tiene los siguientes tipos de puerto:

Principal

Puertos que enlazan con el origen principal en la asignación.

Detalle

Puertos que enlazan con el origen de detalle en la asignación.

Puerto dinámico

Recibe o devuelve puertos en una asignación dinámica. Un puerto dinámico puede recibir una o varias columnas de una transformación de nivel superior y crear un puerto generado para cada columna. Un puerto de salida dinámico puede devolver uno o varios puertos generados. Puede definir reglas de entrada para determinar qué columnas puede recibir un puerto dinámico.

Puede cambiar un puerto principal a puerto de detalle. También puede cambiar un puerto de detalle a puerto principal. Cuando se cambia el tipo de un puerto, se cambia el tipo de puerto de todos los puertos. Por tanto,

cuando se cambia un puerto principal a puerto de detalle, se cambian todos los puertos principales a puertos de detalle y viceversa.

Transformaciones de combinación en asignaciones dinámicas

Puede utilizar una transformación de combinación en una asignación dinámica. Puede hacer referencia a los puertos dinámicos y puertos generados de las condiciones de combinación.

Una asignación dinámica es una asignación en la que se pueden cambiar orígenes, destinos y lógicas de transformación en tiempo de ejecución. Puede establecer los parámetros y reglas con los que cambiar la estructura de los datos. Al utilizar una transformación de combinación en una asignación dinámica, la estructura del origen puede cambiar. Los puertos de entrada y los puertos en la condición de combinación también cambian.

Puede realizar las siguientes tareas para configurar una transformación de combinación en una asignación dinámica:

Definir puertos dinámicos.

Defina puertos dinámicos y puertos generados para dar cabida a las distintas columnas de entrada de un origen dinámico. Puede incluir los puertos dinámicos y puertos generados en la condición de combinación.

Definir selectores de puerto.

Defina un selector de puerto que contenga los puertos que se van a utilizar en la condición de combinación. Configure reglas de selección que determinen los puertos que se van a incluir en el selector de puerto. Puede parametrizar el selector de puerto para que incluya puertos específicos en tiempo de ejecución.

Parametrizar la condición de combinación.

Puede parametrizar toda la condición de combinación. Configure un parámetro de expresión por cada condición de combinación que pueda necesitar.

Para obtener más información sobre las asignaciones dinámicas, consulte la *Guía de asignación de Informática Developer*.

Selectores de puerto en una transformación de combinación

Cuando la transformación de combinación tiene puertos generados, es necesario configurar una condición de combinación que sea válida si la transformación tiene puertos generados distintos en tiempo de ejecución.

Por ejemplo, una asignación dinámica incluye una transformación de combinación con la siguiente condición de combinación:

```
CustomerID = CustomerNo
```

CustomerID es un puerto generado en la transformación de combinación. Dado que la asignación tiene un origen dinámico, podrá ejecutarse con diferentes formatos de archivo de origen. La columna que contiene el número de cliente tiene un nombre diferente en cada archivo de origen: CustomerID, CustomerNum o CustNO.

Puede crear un selector de puerto en la transformación de combinación para dar cabida a los distintos nombres de columna de cliente del origen dinámico. Configure un selector de puerto con una regla de selección que incluya cualquier nombre del puerto con el prefijo "Cust".

Luego, configure la condición de combinación de modo que incluya el nombre del selector de puerto en lugar del nombre de columna CustomerID:

```
Customer_PortSelector = CustomerNo
```

La condición de combinación es válida con cualquier nombre de puerto que empiece por "Cust".

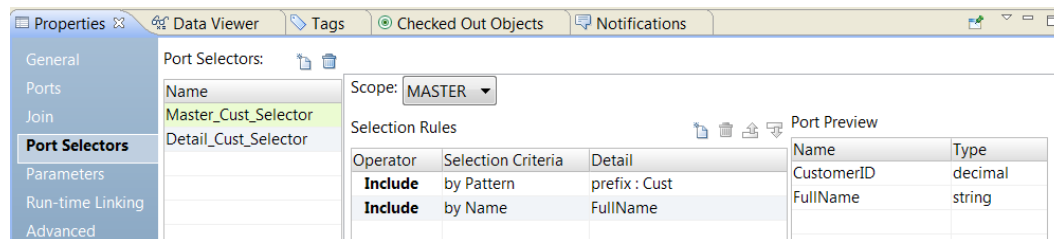
Un selector de puerto puede contener uno o más puertos. La condición de combinación puede incluir varios puertos si el grupo principal y el grupo de detalles de la condición de combinación contienen el mismo número de puertos.

Reglas de selección

Al configurar un selector de puerto, puede definir reglas de selección para saber qué puertos generados incluir. Las reglas de selección son similares a las reglas de entrada que se configuran para los puertos dinámicos.

Un selector de puerto puede incluir puertos o puertos generados. Configure un selector de puerto en la ficha **Selector de puerto**.

La siguiente imagen muestra la ficha **Selector de puerto**:



Configure las siguientes propiedades de un selector de puerto:

Nombre

Identifica el selector de puerto. Puede crear varios selectores de puerto en una transformación y hacer referencia a ellos en las expresiones.

Ámbito

Identifica un grupo de puertos al que el selector de puerto se aplica. Elija el ámbito Principal o Detalle.

Reglas de selección

Determina los puertos que se van a incluir en el selector de puerto. Al crear la selección reglas, el panel **Vista previa del puerto** muestra los puertos que cumplen los requisitos de los puertos de entrada actuales. Estos puertos pueden cambiar. Configure las reglas de selección para dar cabida a puertos de diferentes orígenes.

Puede crear reglas de selección según los siguientes criterios:

Operador

Incluye o excluye los puertos que las reglas de selección devuelven. El valor predeterminado es Incluir. Para poder excluir puertos, antes debe incluirlos.

Criterios de selección

Tipo de regla de selección que desea crear. Puede crear una regla basada en el tipo de puerto o en el nombre de columna. Para incluir puertos según el nombre de columna, busque nombres específicos o un patrón de caracteres en el nombre.

Detalle

Valores que se van a aplicar a los criterios de selección. Si los criterios de selección están definidos según el nombre de columna, configure la cadena o el nombre por el que realizar la búsqueda. Si los criterios de selección están definidos según el tipo de puerto, seleccione los tipos de puerto que quiera incluir.

En la siguiente tabla se describen los criterios de selección y cómo especificar los detalles de los criterios:

Criterios de selección	Descripción	Detalle
Todo	Incluye todos los puertos.	No se necesitan detalles.
Nombre	Filtra los puertos por el nombre del puerto.	Seleccione los nombres de puerto de una lista de valores o utilice un parámetro de tipo puerto o lista de puertos.
Tipo	Filtra los puertos por el tipo de datos de cada puerto.	Seleccione los tipos de datos de una lista.
Patrón	Filtra los puertos por una cadena de caracteres en el nombre o por una expresión regular.	Elija Prefijo, Sufijo o Expresión regular como tipo de patrón para el nombre del puerto. Luego, introduzca un valor para el patrón o utilice un parámetro de tipo cadena.

Crear un selector de puerto

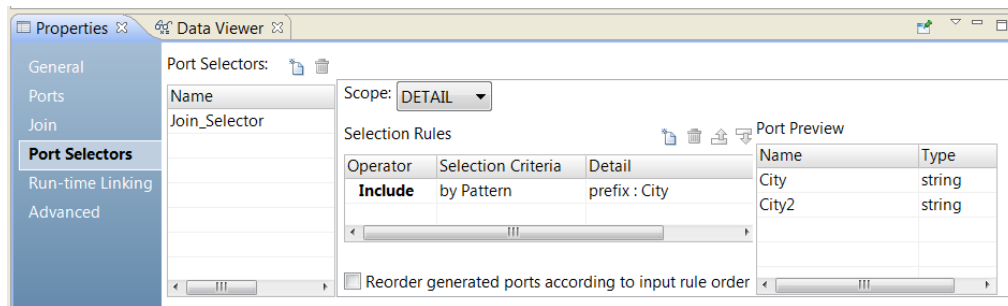
Cree un selector de puerto para saber qué puertos hay que usar en una expresión dinámica, en una condición de búsqueda o en una condición de combinación.

1. Haga clic en la ficha **Selectores de puerto**.
2. En el área **Selectores de puerto**, haga clic en **Nuevo**.
Developer tool crea un selector de puerto con una regla de selección predeterminada que incluye todos los puertos.
3. En el área **Selectores de puerto**, cambie el nombre del selector de puerto por un nombre único.
4. Si está trabajando con una transformación de combinación o de búsqueda, seleccione el ámbito.
Los puertos disponibles varían según el grupo de puertos que elija.
5. En el área **Reglas de selección**, seleccione un **Operador**.
 - Incluir. Cree una regla que incluya puertos para el selector de puerto. Para poder excluir puertos, antes debe incluirlos.
 - Excluir. Cree una regla que excluya puertos específicos del selector de puerto.

6. Elija los **Criterios de selección**.

- Por nombre. Seleccione puertos concretos por su nombre. Puede seleccionar los nombres de puerto de una lista de puertos en el ámbito.
- Por tipo. Seleccione puertos por el tipo. Puede seleccionar uno o varios tipos de datos.
- Por patrón. Seleccione puertos siguiendo un patrón de caracteres en el nombre del puerto. Puede realizar las búsquedas con caracteres específicos o crear una expresión regular.

La siguiente imagen muestra la ficha Selector de puerto:



7. Haga clic en la columna **Detalle**.

Se abre el cuadro de diálogo **Detalles de la regla de entrada**.

8. Seleccione los valores por los que quiera filtrar los puertos.

- Por nombre. Decida si quiere crear una lista de puertos por valor o por un parámetro. Haga clic en **Elegir** para seleccionar los puertos de la lista.
- Por tipo. Seleccione uno o varios tipos de datos de una lista. El área **Vista previa del puerto** muestra los puertos de los tipos que ha seleccionado.
- Por patrón. Decida si quiere hallar un determinado patrón de caracteres en el prefijo o sufijo del nombre del puerto. También puede optar por crear una expresión regular con la que realizar la búsqueda. Configure un parámetro o el patrón de búsqueda.

El área **Vista previa del puerto** muestra los puertos en el selector de puerto a medida que configura las reglas.

9. Para reordenar los puertos en el selector de puerto, seleccione **Reordenar los puertos generados según el orden de regla de entrada**.

Definir una condición de combinación

La condición de combinación contiene puertos de ambos orígenes de entrada que el servicio de integración de datos utiliza para unir dos filas.

En función del tipo de combinación seleccionado, el servicio de integración de datos añade la fila al conjunto de resultados o la descarta. La transformación de combinación genera conjuntos de resultados en función del tipo unión, de la condición y de los orígenes de datos de entrada.

Antes de definir una condición de combinación, compruebe que los orígenes principal y detallado están configurados para un rendimiento óptimo. Durante la ejecución de una asignación, el servicio de integración de datos compara cada fila del origen principal con el detallado. Para mejorar el rendimiento de una transformación de combinación no ordenada, utilice como origen principal el que tiene menos filas. Para mejorar el rendimiento de una transformación de combinación no ordenada, utilice como origen principal el que tiene menos valores de clave duplicados.

En la condición de combinación, utilice uno o varios puertos de los orígenes de entrada de una transformación de combinación. Cuantos más puertos haya, mayor será el tiempo necesario para unir dos orígenes. El orden de los puertos de la condición puede afectar al rendimiento de la transformación de combinación. Si utiliza varios puertos en la condición de combinación, el servicio de integración de datos los compara en el orden en el que los especifica.

Si une los tipos de datos Char y Varchar, el servicio de integración de datos cuenta todos los espacios de los valores Char como parte de la cadena:

```
Char(40) = "abcd"  
Varchar(40) = "abcd"
```

El valor Char es "abcd" junto con 36 espacios en blanco; el servicio de integración de datos no unirá los dos campos porque el de Char contiene espacios al final.

Nota: La transformación de combinación no hace coincidir valores null. Por ejemplo, si tanto EMP_ID1 como EMP_ID2 contienen una fila que tiene un valor null, el servicio de integración de datos no los considera como una coincidencia y no une ambas filas. Para unir filas que tienen valores null, sustituya estos por valores predeterminados y, a continuación, únalos.

Puede definir un tipo de condición simple o avanzado. También puede definir un parámetro de expresión. Un parámetro de expresión es un parámetro que contiene una expresión de combinación. Puede cambiar el valor del parámetro en tiempo de ejecución por un parámetro de asignación.

Tipo de condición simple

Defina un tipo de condición simple para una transformación de combinación ordenada o no ordenada.

Una condición simple incluye una o varias condiciones que comparan los orígenes principal y detallado especificados. Las condiciones simples deben tener el siguiente formato:

```
<master_port> operator <detail_port>
```

Para las transformaciones de incorporación ordenadas, la condición debe utilizar el operador de igualdad.

En las transformaciones de combinación no ordenadas, la condición puede utilizar cualquiera de los siguientes operadores: =, !=, >, >=, <, <=.

Por ejemplo, si dos orígenes con tablas llamadas EMPLOYEE_AGE y EMPLOYEE_POSITION contienen números de ID de empleado, la siguiente condición hace coincidir las filas con los empleados que se enumeran en ambos orígenes:

```
EMP_ID1 = EMP_ID2
```

Developer Tool valida los tipos de datos en una condición simple. Los dos puertos de la condición deben tener el mismo tipo de datos. Si en la condición necesita utilizar dos puertos con tipos de datos no coincidentes, convierta los tipos de datos para hacerlos coincidir.

Puede configurar una lista de condiciones de combinación en una condición simple. Al configurar varias condiciones de combinación, todas las condiciones deben ser TRUE para que la combinación sea posible.

Por ejemplo, puede configurar las siguientes instrucciones en una condición simple:

```
StoreID = StoreNO  
Dept = Department  
Salary > Commission
```

Si ve las mismas instrucciones como una condición avanzada, la condición de combinación aparece como la siguiente expresión:

```
StoreID = StoreNO AND Dept = Department AND (Salary > Commission)
```

Tipo de condición avanzada

Defina un tipo de condición avanzada para una transformación de combinación no ordenada.

Una condición avanzada puede incluir cualquier expresión que evalúe un valor numérico o booleano. Una condición avanzada puede incluir cualquiera de los siguientes operadores: =, !=, >, >=, <, <=.

Puede especificar una constante para la condición de combinación. El equivalente numérico de FALSE es cero (0). Todo valor distinto de cero equivale a TRUE. Por ejemplo, la transformación contiene un puerto llamado NUMBER_OF_UNITS con un tipo de datos numérico. Configure una condición de filtro para que devuelva FALSE si el valor de NUMBER_OF_UNITS es igual a cero. De lo contrario, la condición devuelve TRUE.

Nota: No se puede utilizar un solo puerto dinámico o un selector de puerto como valor booleano para una condición de combinación.

Para introducir una expresión en la condición de combinación, elija el tipo de condición Avanzadas en la ficha **Unión**. Utilice el editor de expresiones para incluir puertos, parámetros, expresiones, selectores de puerto y operadores en la condición. Puede utilizar puertos generados. Puede especificar un solo puerto en el editor de expresiones si el tipo de puerto es numérico. Sin embargo, no se puede introducir un selector de puerto como una expresión.

Por ejemplo, imagine que desea unir orígenes de forma que coincida el nombre completo de un empleado. El origen principal incluye un puerto de nombre FirstName y otro de apellido LastName. El origen de detalle incluye un puerto de nombre completo FullName. Defina la siguiente condición para concadenar los puertos principales y hacer coincidir el nombre completo en los dos orígenes:

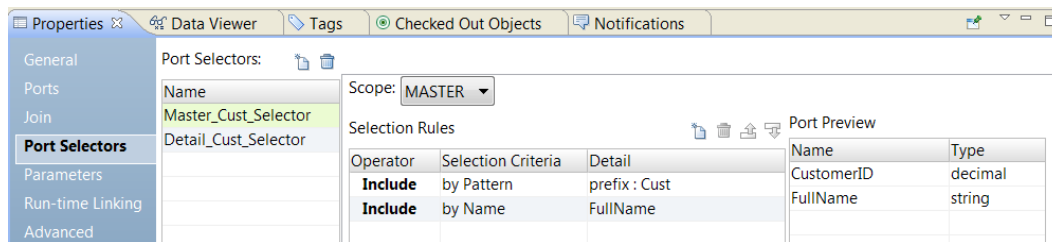
```
CONCAT(FirstName, LastName) = FullName
```

Selectores de puerto en condiciones de combinación

En una condición de combinación puede incluir selectores de puerto. La condición de combinación debe hacer referencia a un selector de puerto del grupo principal y a un selector de puerto del grupo de detalles.

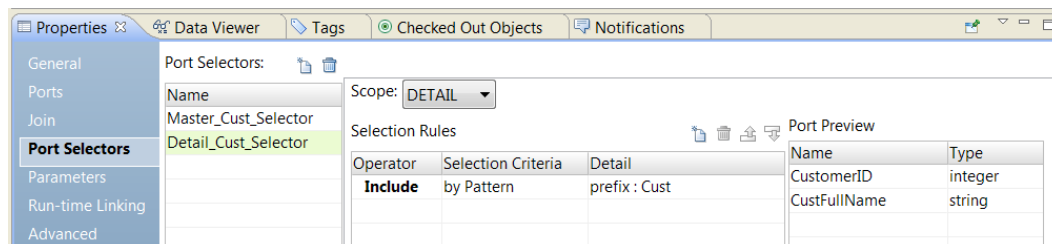
Por ejemplo, la transformación de combinación tiene puertos dinámicos. Puede que necesite comparar varios puertos generados en la condición de combinación.

La siguiente imagen muestra los campos del selector de puerto del grupo principal:

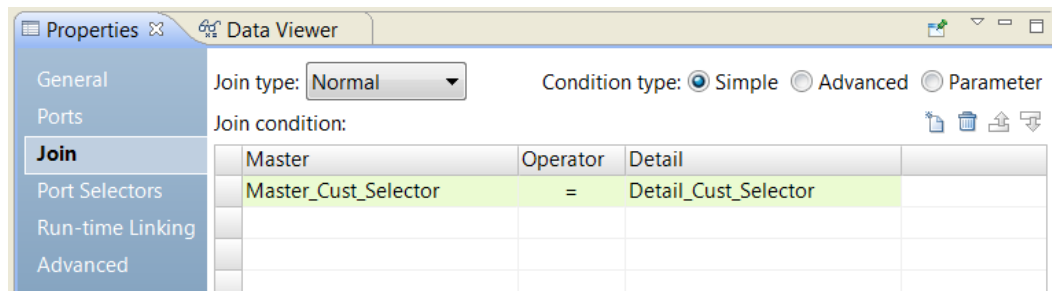


Master_Cust_Selector contiene los puertos CustomerID y FullName.

La siguiente imagen muestra los campos del selector de puerto del grupo de detalles:



Detail_Cust_Selector contiene los puertos CustomerNo y CustFullName. Estos puertos tienen el prefijo Cust. Cree la siguiente condición de combinación simple:



La condición de combinación compara cada puerto de Master_Cust_Selector con los de Detail_Cust_Selector. La condición de combinación es: CustomerID = CustomerNo AND FullName = CustFullName.

Cada selector de puerto debe contener el mismo número de puertos. Los puertos deben ser del mismo tipo.

Nota: Si cambia el ámbito de un selector de puerto y una condición de combinación de tipo simple ya no es válida, Developer tool podría cambiar el tipo de la condición a avanzado. Puede restaurar el tipo de la condición de combinación de nuevo a simple en la ficha **Combinación**.

Puertos dinámicos en condiciones de combinación

Puede hacer referencia a un puerto dinámico en un selector de puerto.

Un puerto dinámico puede contener uno o varios puertos generados. Si la condición de combinación contiene puertos dinámicos, el número de puertos principales debe ser el mismo que el número de puertos de detalle.

Por ejemplo, el puerto dinámico A tiene dos puertos generados:

```
CustomerID
OrderID
```

El puerto dinámico B también tiene dos puertos generados:

```
CustomerNo
OrderNo
```

La siguiente condición de combinación es válida:

```
DynamicPortA = DynamicPortB
```

La condición de combinación se expande a la siguiente expresión:

```
CustomerID = CustomerNo AND OrderID = OrderNo
```

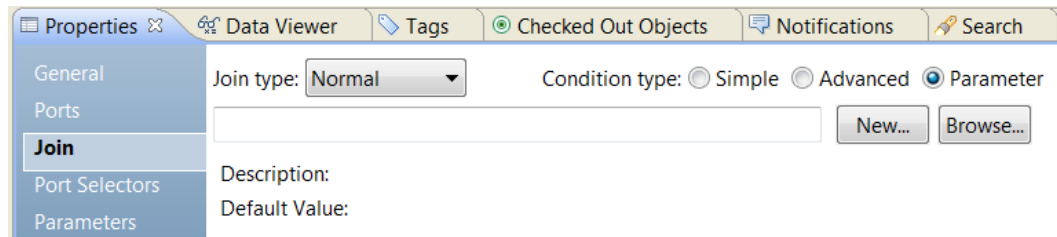
Puede hacer referencia a un selector de puerto y a un puerto dinámico en una condición de combinación si el selector de puerto contiene el mismo número de puertos que el puerto dinámico.

Parámetro de expresión

Puede definir un parámetro de expresión que contenga una condición de combinación. Puede elegir el parámetro como la condición de combinación en la transformación de combinación.

Para utilizar un parámetro en la condición de combinación, seleccione el tipo de condición de parámetro en la ficha Combinación.

La siguiente imagen muestra dónde se selecciona el tipo de condición de parámetro:



Puede buscar un parámetro existente o crear uno. Para crear un parámetro, haga clic en **Nuevo** y defina el parámetro. Crear la expresión en el editor de expresiones.

Nota: Un parámetro de expresión no puede contener otros parámetros. Si incrusta un parámetro en un parámetro de expresión, el Servicio de integración de datos generará un error de validación en tiempo de ejecución.

Tipos de unión

En una transformación de incorporación, la unión se puede originar desde distintos tipos de orígenes.

La transformación de incorporación es compatible con los siguientes tipos de unión:

- Normal
- Outer principal
- Outer detallado
- Outer completo

Nota: Un outer join normal o principal se ejecuta más rápidamente que una unión outer completo o una outer de detalle.

Si un conjunto de resultados incluye campos que no contienen datos en ninguno de los orígenes, las propiedades de la transformación de incorporación llena los campos vacíos con valores nulos. Si sabe que un campo devuelve NULL y no desea insertar valores NULL en el destino, puede definir un valor predeterminado para el puerto correspondiente.

Unión normal

Con una unión normal, el servicio de integración de datos descarta, de acuerdo con la condición, todas las filas de datos del origen principal y el origen de detalle que no coinciden.

Por ejemplo, tenemos dos orígenes de datos para recambios de automóvil llamados PARTS_SIZE y PARTS_COLOR.

El origen de los datos de PARTS_SIZE es el origen principal y contiene los siguientes datos:

PART_ID1	DESCRIPTION	SIZE
1	Seat Cover	Large
2	Ash Tray	Small
3	Floor Mat	Medium

El origen de los datos de PARTS_COLOR es el origen detallado y contiene los siguientes datos:

PART_ID2	DESCRIPTION	COLOR
1	Seat Cover	Blue
3	Floor Mat	Black
4	Fuzzy Dice	Yellow

Para unir las dos tablas haciendo coincidir PART_IDs en ambos orígenes, defina la condición del siguiente modo:

```
PART_ID1 = PART_ID2
```

Cuando se unen estas tablas de muestra con una unión normal, el conjunto de resultados incluye los siguientes datos:

PART_ID	DESCRIPTION	SIZE	COLOR
1	Seat Cover	Large	Blue
3	Floor Mat	Medium	Black

El siguiente ejemplo muestra la instrucción SQL equivalente:

```
SELECT * FROM PARTS_SIZE, PARTS_COLOR WHERE PARTS_SIZE.PART_ID1 = PARTS_COLOR.PART_ID2
```

Outer Join principal

Un outer join principal mantiene todas las filas de datos del origen de detalle y las filas coincidentes del origen principal. Descarta las filas no coincidentes del origen principal.

Cuando se unen las tablas de muestra con un outer join principal y la misma condición, el conjunto de resultados incluye los siguientes datos:

PART_ID	DESCRIPTION	SIZE	COLOR
1	Seat Cover	Large	Blue
3	Floor Mat	Medium	Black
4	Fuzzy Dice	NULL	Yellow

Puesto que no se ha especificado ningún tamaño para Datos de peluche, el servicio de integración de datos llena el campo con NULL.

El siguiente ejemplo muestra la instrucción SQL equivalente:

```
SELECT * FROM PARTS_SIZE RIGHT OUTER JOIN PARTS_COLOR ON (PARTS_COLOR.PART_ID2 = PARTS_SIZE.PART_ID1)
```

Outer Join de detalles

Un outer join de detalles mantiene todas las filas de datos del origen principal y las filas coincidentes del origen de detalle. Descarta las filas no coincidentes del origen de detalle.

Cuando se unen las tablas de muestra con un outer join de detalles y la misma condición, el conjunto de resultados incluye los siguientes datos:

PART_ID	DESCRIPTION	SIZE	COLOR
1	Seat Cover	Large	Blue
2	Ash Tray	Small	NULL
3	Floor Mat	Medium	Black

Puesto que no se ha especificado ningún color para Cenicero, el servicio de integración de datos llena el campo con NULL.

El siguiente ejemplo muestra la instrucción SQL equivalente:

```
SELECT * FROM PARTS_SIZE LEFT OUTER JOIN PARTS_COLOR ON (PARTS_SIZE.PART_ID1 = PARTS_COLOR.PART_ID2)
```

Outer Join completo

Un outer join completo mantiene todas las filas de datos de tanto orígenes principales como orígenes de detalle.

Cuando se unen las tablas de muestra con un outer join completo y la misma condición, el conjunto de resultados incluye los siguientes datos:

PARTED	DESCRIPTION	SIZE	Color
1	Seat Cover	Large	Blue
2	Ash Tray	Small	NULL
3	Floor Mat	Medium	Black
4	Fuzzy Dice	NULL	Yellow

Puesto que no se especifica ningún color para Cenicero y no se especifica ningún tamaño para los Dados de peluche, el servicio de integración de datos llena los campos con NULL.

El siguiente ejemplo muestra la instrucción SQL equivalente:

```
SELECT * FROM PARTS_SIZE FULL OUTER JOIN PARTS_COLOR ON (PARTS_SIZE.PART_ID1 = PARTS_COLOR.PART_ID2)
```


Entrada ordenada para una transformación de incorporación

Puede aumentar el rendimiento de una transformación de incorporación con la opción de entrada ordenada. Utilice la entrada ordenada cuando los datos estén ordenados.

Cuando se configura la transformación de incorporación para utilizar datos ordenados, el servicio de integración de datos aumenta el rendimiento minimizando la entrada y la salida del disco. Verá el mejor rendimiento cuando se trabaja con conjuntos de datos de gran tamaño.

Para configurar una asignación para que utilice datos ordenados, defina y mantenga un orden de clasificación en la asignación para que el servicio de integración de datos pueda utilizar los datos ordenados cuando procese la transformación de incorporación. Lleve a cabo los siguientes pasos para configurar la asignación:

1. Configure el orden de clasificación de los datos que desee unir.
2. Añada transformaciones que mantengan el orden de los datos ordenados.
3. Configure la transformación de incorporación para que utilice datos ordenados y configure la condición de unión para que utilice los puertos de origen de clasificación. El origen de la ordenación representa el origen de los datos ordenados.

Cómo configurar el orden de clasificación

Configure el orden de clasificación para asegurarse de que el servicio de integración de datos pasa datos ordenados a la transformación de incorporación.

Para configurar el orden de clasificación, utilice uno de los siguientes métodos:

- Utilice archivos sin formato ordenados. Cuando los archivos sin formato contienen datos ordenados, compruebe que el orden de las columnas de ordenación coincide en cada uno de los archivos de origen.
- Utilice datos relacionales ordenados. Utilice puertos ordenados en el objeto de datos relacionales para ordenar columnas de la base de datos de origen. Configure el orden de los puertos ordenados del mismo modo en cada objeto de datos relacionales.
- Utilice una transformación de incorporación para ordenar datos relacionales o de archivo sin formato. Ponga una transformación de incorporación en los canales principal y de detalle. Configure cada transformación de incorporación para que utilice el mismo orden que los puertos clave de ordenación y la misma dirección de ordenación.

Si se pasan datos desordenados u ordenados incorrectamente a una transformación de incorporación configurada para utilizar datos ordenados, la ejecución de la asignación generará un error. El servicio de integración de datos registra el error en el archivo de registro.

Cómo añadir transformaciones a la asignación

Añada transformaciones a la asignación que mantiene el orden de los datos ordenados en una transformación de incorporación.

Puede poner la transformación de incorporación directamente después del origen de ordenación para mantener los datos ordenados.

Cuando se añaden transformaciones entre el origen de ordenación y la transformación de incorporación, utilice las siguientes directrices para mantener los datos ordenados:

- No ponga ninguna de las siguientes transformaciones entre el origen de ordenación y la transformación de incorporación.
 - Rango
 - Unión
 - Agregación no ordenada
 - Mapplet que contiene una de las transformaciones anteriores
- Puede poner una transformación de agregación ordenada entre el origen de ordenación y la transformación de incorporación si utiliza las siguientes directrices:
 - Configure la transformación de agregación para una entrada ordenada.
 - Utilice los mismos puertos para las columnas agrupar por en la transformación de agregación que los puertos en el origen de ordenación.
 - Los puertos agrupar por deben estar en el mismo orden que los puertos en el origen de ordenación.
- Cuando una el conjunto de resultados de una transformación de incorporación con otro canal, compruebe que la salida de datos de la primera transformación de incorporación esté ordenada.

Reglas y directrices para las condiciones de combinación

Algunas reglas y directrices se aplican al crear condiciones de unión para una transformación de combinación ordenada.

Utilice las siguientes directrices cuando cree condiciones de unión:

- Debe definir un tipo de condición simple que utilice el operador de igualdad.
- Si utiliza una transformación de agregación ordenada entre el origen de la ordenación y la transformación de combinación, trate la transformación de agregación como el origen de ordenación cuando defina la condición de unión.
- Los puertos que se utilizan en la condición de unión deben coincidir con los puertos en el origen de ordenación.
- Cuando se configuran varias condiciones de unión, los puertos de la primera condición de unión deben coincidir con los primeros puertos en el origen de ordenación.
- Cuando se configuran varias condiciones, el orden de estas debe coincidir con el orden de los puertos en el origen de ordenación y no se debe omitir ningún puerto.
- El número de puertos ordenados en el origen de ordenación debe ser mayor o igual que el número de puertos en la condición de unión.
- Si una puertos con tipos de datos Decimal, la precisión de cada puerto debe pertenecer al mismo rango de precisión.

Puede utilizar uno de los siguientes rangos de precisión válidos:

- 0-18 decimal
- 19-28 decimal
- Decimal 29 a 38
- Decimal 39 y superior

Por ejemplo, si define la condición `DecimalA = DecimalB`, donde `DecimalA` tiene una precisión de 15 y `DecimalB` de 25, la condición no será válida.

Ejemplo de una condición de combinación y del orden de clasificación

Este ejemplo muestra una transformación de unión que une los canales principal y de detalle con puertos ordenados.

Las transformaciones de ordenación en los canales maestro y de detalle se configuran con los siguientes puertos ordenados:

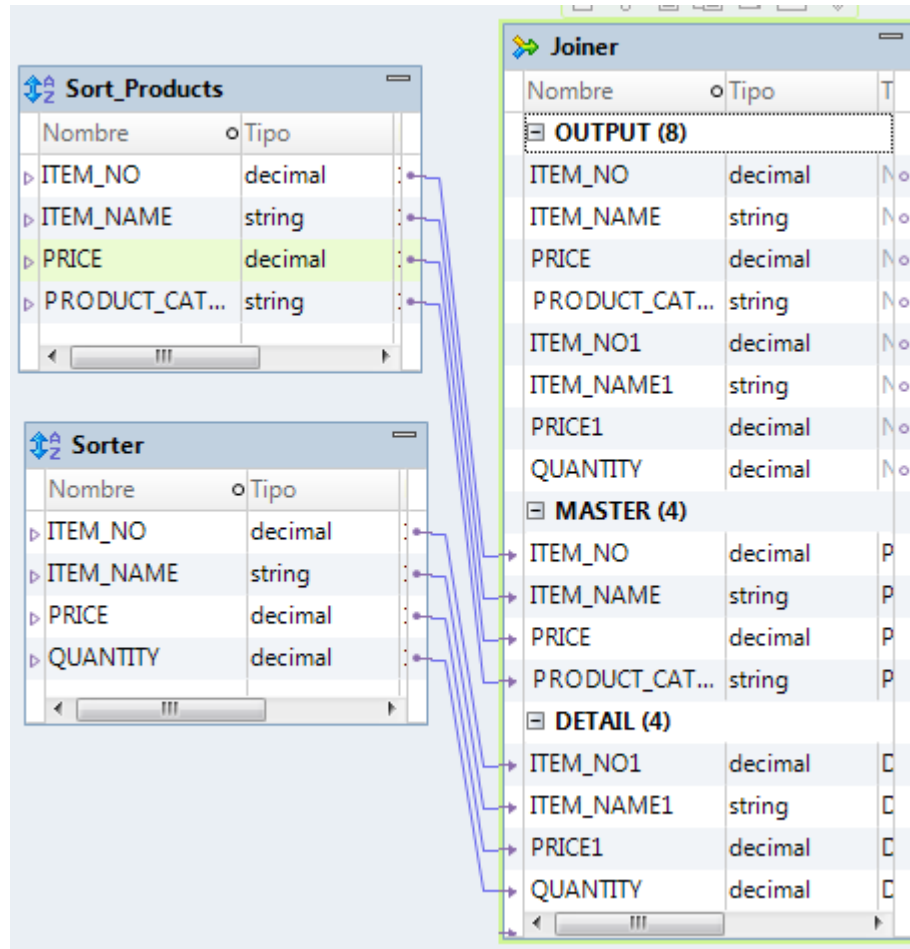
- ITEM_NO
- ITEM_NAME
- PRICE

Cuando configure la condición de combinación, utilice las siguientes directrices para mantener el orden de clasificación:

- Debe utilizar ITEM_NO en la primera condición de combinación.
- Si añade una segunda condición de combinación, debe utilizar ITEM_NAME.
- Si desea utilizar PRICE en una condición de combinación, también debe utilizar ITEM_NAME en la segunda condición de combinación.

Si omite ITEM_NAME y une ITEM_NO y PRICE, pierde el orden de clasificación y el servicio de integración de datos no ejecuta la asignación.

La siguiente figura muestra una asignación configurada para ordenar y unir los puertos ITEM_NO, ITEM_NAME, y PRICE:



Cuando se usa la transformación de unión para unir los canales principal y de detalle, se puede configurar cualquiera de las siguientes condiciones de combinación:

ITEM_NO = ITEM_NO

o

ITEM_NO = ITEM_NO1

ITEM_NAME = ITEM_NAME1

o

ITEM_NO = ITEM_NO1

ITEM_NAME = ITEM_NAME1

PRICE = PRICE1

Cómo unir datos de un mismo origen

Se pueden unir datos de un mismo origen si se desea realizar un cálculo en parte de los datos y unir los datos transformados a los datos originales.

Cuando se unen datos de un mismo origen, pueden mantener los datos originales y transformar partes de los mismos dentro de una asignación. Se pueden unir datos de un mismo origen de las siguientes maneras:

- Unir dos ramificaciones del mismo canal.
- Unir dos instancias del mismo origen.

Cómo unir dos ramificaciones del mismo canal

Cuando se unen datos de un mismo origen, puede crear dos ramificaciones del canal.

Cuando se ramifica un canal, debe añadirse una transformación entre la entrada de asignación y la transformación de incorporación en, como mínimo, una ramificación del canal. Debe unir datos ordenados y configurar la transformación de incorporación para la entrada ordenada.

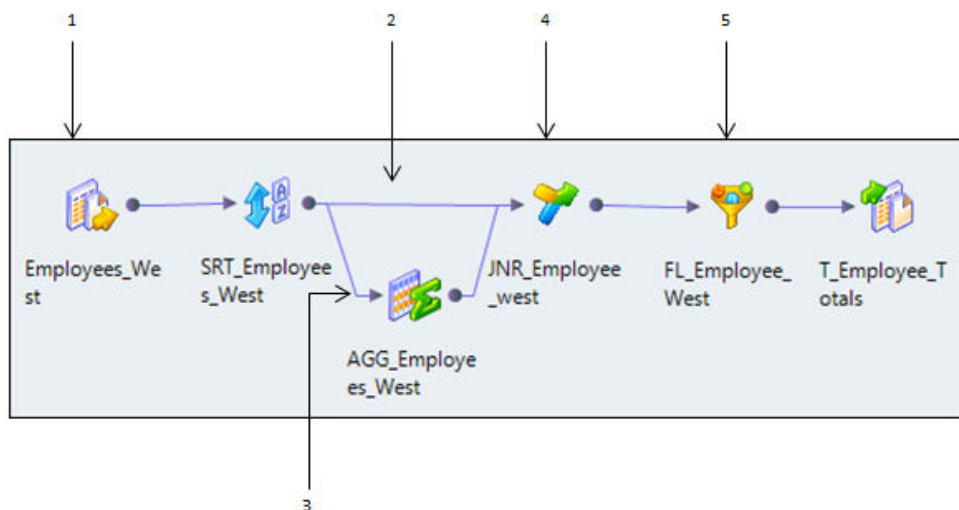
Por ejemplo, tiene un origen con los siguientes puertos:

- Empleado
- Departamento
- Total de ventas

En el destino, desea ver los empleados que han generado más ventas que la cifra promedio de ventas de sus departamentos. Para ello, cree una asignación con las siguientes transformaciones:

- Transformación de ordenación. Ordena los datos.
- Transformación de agregación ordenada. Produce un promedio de los datos de ventas y agrupa por departamento. Cuando se realiza esta agregación, se pierden los datos de cada uno de los empleados. Para conservar los datos de los empleados debe pasar una ramificación del canal a la transformación de agregación y pasar una ramificación con los mismos datos a la transformación de incorporación para mantener los datos originales. Cuando se unen ambas ramificaciones del canal se unen los datos agregados con los datos originales.
- Transformación de incorporación ordenada. Une los datos agregados ordenados a los datos originales.

- Transformación de filtro. Compara los datos del promedio de ventas con los datos de ventas de cada empleado y excluye los empleados con una cifra de ventas inferior al promedio de ventas indicado arriba.



1. Origen Employees_West
2. Ramificación de canal 1
3. Ramificación de canal 2
4. Transformación de incorporación ordenada
5. Filtrar empleados sin un promedio de ventas superior a la media

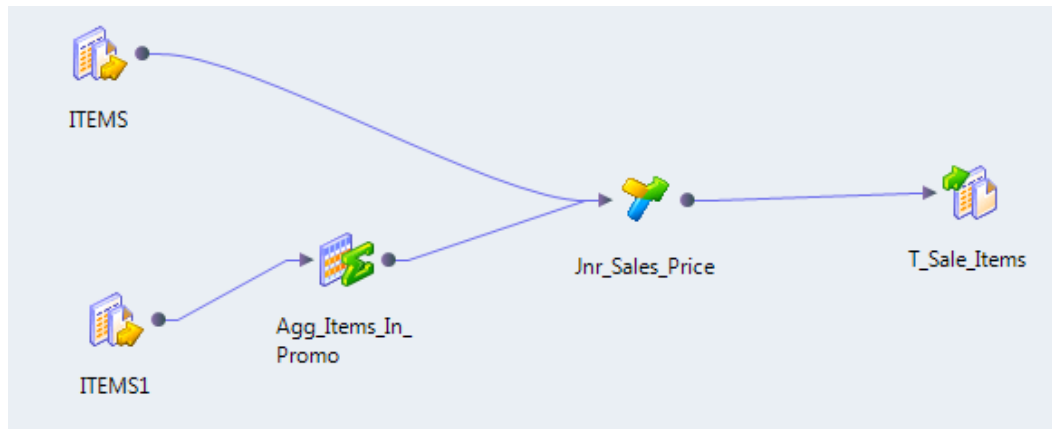
Unir dos ramificaciones puede reducir el rendimiento si la transformación de incorporación recibe datos de una ramificación mucho después de recibir los datos de la otra ramificación. La transformación de incorporación guarda en memoria caché todos los datos de la primera ramificación y escribe la memoria caché en disco en el caso de que esté llena. Luego, la transformación de incorporación debe leer los datos del disco cuando recibe los datos de la segunda ramificación.

Cómo unir dos instancias del mismo origen

Se pueden unir datos de un mismo origen creando una segunda instancia del origen.

Después de crear la segunda instancia del origen, puede unir los canales de las dos instancias del origen. Si desea unir datos no ordenados, debe crear dos instancias del mismo origen y unir los canales.

La siguiente figura muestra dos instancias del mismo origen unidas con una transformación de unión:



Cuando se unen dos instancias de un mismo origen, el servicio de integración de datos lee los datos de origen para cada instancia del origen. El rendimiento puede ser más lento que unir dos ramificaciones de un canal.

Pautas para unir datos de un mismo origen

Se aplican ciertas pautas en el momento de decidir si se deben unir ramificaciones de un canal o unir dos instancias de un origen.

Utilice las siguientes pautas en el momento de decidir si se deben unir ramificaciones de un canal o unir dos instancias de un origen.

- Una dos ramificaciones de un canal cuando el origen sea de gran tamaño o si solamente se pueden leer los datos de origen una sola vez.
- Una dos ramificaciones de un canal cuando utilice datos ordenados. Si los datos de origen no están ordenados y utiliza una transformación de ordenación para ordenar los datos, ramifique el canal después de ordenar los datos.
- Una dos instancias de un origen cuando tenga que añadir una transformación de bloqueo al canal entre el origen y la transformación de incorporación.
- Una dos instancias de un origen en el caso de que un canal se procese más lentamente que el otro canal.
- Una dos instancias de un origen si necesita unir datos no ordenados.

Cómo bloquear los canales de origen

Cuando se ejecuta una asignación con una transformación de incorporación, el servicio de integración de datos bloquea y desbloquea los datos de origen conforme la configuración de la asignación y si se ha configurado la transformación de incorporación para una entrada ordenada.

Transformación de incorporación no ordenada

Cuando el servicio de integración de datos procesa una transformación de incorporación no ordenada, lee todas las filas principales antes de leer las filas de detalle. El servicio de integración de datos bloquea el origen de detalle mientras recopila filas del origen principal.

Después de que el servicio de integración de datos lee y guarda en la memoria caché todas las filas principales, desbloquea el origen de detalle y lee las filas de detalle. Algunas asignaciones con transformaciones de incorporación no ordenadas infringen la validación del flujo de datos.

Transformación de incorporación ordenada

Cuando el servicio de integración de datos procesa una transformación de incorporación ordenada, bloquea datos según la configuración de la asignación. La lógica de bloqueo es posible tanto si la entrada principal como la de detalle de la transformación de incorporación proceden de orígenes diferentes.

El servicio de integración de datos utiliza lógica de bloqueo para procesar la transformación de incorporación, si puede hacerlo, sin bloquear simultáneamente todos los orígenes de un grupo de orden de carga de destino. De lo contrario, no utiliza lógica de bloqueo. En lugar de ello, almacena más filas en la memoria caché.

Cuando el servicio de integración de datos puede utilizar lógica de bloqueo para procesar la transformación de incorporación, almacena menos filas en la memoria caché, con lo que se aumenta el rendimiento.

Cómo guardar filas maestras en la memoria caché

Cuando el servicio de integración de datos procesa una transformación de incorporación, lee simultáneamente filas de ambos orígenes y construye el índice y la memoria caché de datos según las filas maestras.

Luego, el servicio de integración de datos realiza la unión según los datos de origen de detalle y los datos de la memoria caché. El número de filas que el servicio de integración de datos guarda en la memoria caché depende de los datos de origen y si se ha configurado la transformación de incorporación para una entrada ordenada.

Para aumentar el rendimiento de una transformación de incorporación no ordenada, utilice el origen con menos filas que el origen principal. Para aumentar el rendimiento de una transformación de incorporación ordenada, utilice el origen con menos valores clave duplicados que el origen principal.

Consejos sobre el rendimiento de la transformación de combinación

Siga estos consejos para aumentar el rendimiento de la transformación de combinación.

Las transformaciones de incorporación pueden reducir el rendimiento porque necesitan espacio adicional en tiempo de ejecución para guardar los resultados intermedios. Puede consultar la información del contador de rendimiento de combinación para determinar si es necesario optimizar las transformaciones de combinación.

Siga los consejos que se dan a continuación para aumentar el rendimiento de la transformación de combinación:

Designe el origen principal como origen con menos valores de clave duplicadas.

Cuando el servicio de integración de datos procesa una transformación de combinación ordenada, guarda simultáneamente en la memoria caché cien claves únicas. Si el origen principal contiene muchas filas con el mismo valor de clave, el servicio de integración de datos debe guardar más filas en la memoria caché, con lo cual se reduce el rendimiento.

Designe el origen principal como origen con menos filas.

La transformación de combinación compara cada fila del origen de detalle con el origen principal. A menor número de filas que haya en el origen principal se producen menos iteraciones de la comparación de uniones, lo que acelera el proceso de unión.

Realice uniones en una base de datos siempre que sea posible.

Las uniones en bases de datos se realizan más rápido que las que se realizan durante la ejecución de la asignación. El rendimiento puede variar en función del tipo de unión de base de datos que se utilice. Las uniones normales son más rápidas que los outer joins y producen menos filas. En ocasiones no es posible realizar la unión en la base de datos, como unir tablas de dos bases de datos o sistemas de archivos sin formato diferentes.

Una datos ordenados siempre que sea posible.

Configure la transformación de combinación para que utilice la entrada ordenada. El servicio de integración de datos aumenta el rendimiento reduciendo al mínimo la entrada y la salida de disco. Se

observa el mayor rendimiento al trabajar con conjuntos de datos de gran tamaño. Para una transformación de combinación no ordenada, designe como origen principal el origen con menos filas.

Optimizar la condición de combinación.

El servicio de integración de datos intenta reducir el tamaño del conjunto de datos de un operando de combinación leyendo las filas del grupo más pequeño, buscando las filas coincidentes en el grupo más grande y luego llevando a cabo la operación de combinación. La reducción del tamaño del conjunto de datos mejora el rendimiento de la asignación, ya que el servicio de integración de datos ya no necesita leer filas innecesarias desde el origen del grupo más grande. El servicio de integración de datos mueve la condición de combinación al origen del grupo más grande y solamente lee las filas que coinciden con el grupo más pequeño.

Utilice el método de optimización semi-join.

Utilice el método de optimización semi-join para mejorar el rendimiento de la asignación cuando un grupo de entrada tiene muchas más filas que el otro y cuando el grupo más grande tiene muchas filas sin coincidencia en el grupo más pequeño en función de la condición de combinación.

Reglas y directrices para una transformación de combinación

Hay ciertas reglas y directrices que se aplican cuando se usa una transformación de combinación.

La transformación de combinación acepta entradas de la mayoría de transformaciones. Sin embargo, no se puede usar una transformación de combinación cuando uno de los dos canales contiene una transformación de estrategia de actualización.

Transformación de combinación en un entorno no nativo

El procesamiento de transformaciones de combinación en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de combinación en el motor de Blaze

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación contiene una combinación de desigualdad y la combinación de asignación está deshabilitada.
- La expresión de la transformación de combinación hace referencia a una transformación de búsqueda no conectada.

La combinación de asignación está deshabilitada cuando la transformación de combinación está configurada para una combinación externa detallada o para una combinación interna completa.

Transformación de combinación en el motor de Spark

La validación de asignaciones genera un error en la siguiente situación:

- La condición de combinación es de tipo de datos binario o contiene expresiones binarias.

fuelle: Transformaciones en el entorno no nativo (Conrefs) conref: Databricks y Spark. Esta referencia de contenido se reutiliza en el tema Transformaciones de la Guía de transformación del desarrollador y la Guía del usuario de BDM.

Transformación de combinación en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- Una transformación de combinación está en el nivel inferior de una transformación de agregación.
- Una transformación de combinación está en el nivel inferior de una transformación de rango.
- Un canal de transmisión contiene más de una transformación de combinación.
- Una transformación de combinación une datos procedentes de canales de transmisión y de no transmisión.

Directrices generales

Para especificar una condición de combinación, seleccione la función TIME_RANGE en el tipo de condición Avanzada en la ficha de combinación e introduzca una expresión de condición de combinación. La función TIME_RANGE determina el intervalo de tiempo para los eventos de transmisión a los que se va a unir.

Transformación de combinación en el motor de Databricks Spark

La validación de asignaciones genera un error en la siguiente situación:

- La condición de combinación es de tipo de datos binario o contiene expresiones binarias.

fuelle: Transformaciones en el entorno no nativo (Conrefs) conref: Databricks y Spark. Esta referencia de contenido se reutiliza en el tema Transformaciones de la Guía de transformación del desarrollador y la Guía del usuario de BDM.

CAPÍTULO 23

Transformación de generador de claves

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de generador de claves, 395](#)
- [Estrategia Soundex, 396](#)
- [Estrategia de cadena, 397](#)
- [Estrategia NYSIIS, 397](#)
- [Puertos de salida del generador de claves, 398](#)
- [Cómo configurar una estrategia de agrupación, 398](#)
- [Propiedades de la creación de claves, 399](#)
- [Propiedades avanzadas de Transformación de generador de claves, 399](#)
- [Transformación del generador de claves en un entorno no nativo, 400](#)

Resumen de la transformación de generador de claves

La transformación de generador de claves es una transformación activa que organiza registros en grupos según los valores de los datos de una columna seleccionada. Utilice esta transformación para ordenar registros antes de pasarlos a la transformación de coincidencia.

La transformación de generador de claves utiliza una estrategia de agrupación para crear claves de grupo para la columna seleccionada. Las estrategias son String, Soundex y NYSIIS. Los registros con valores comunes en el campo seleccionado tienen un valor de clave de grupo común. La transformación de coincidencia procesa simultáneamente los registros con valores de clave de grupo comunes. Esto permite un análisis de duplicados más rápido en la transformación de coincidencia.

El número de operaciones de comparación que debe realizar la transformación de coincidencia aumenta exponencialmente con el número de registros del conjunto de datos. Este aumento exponencial puede consumir cantidades considerables de recursos informáticos. Mediante la creación de claves de grupo, la transformación de generador de claves permite que la transformación de coincidencia compare registros en grupos más pequeños, con lo cual se reduce el tiempo de procesamiento.

Cuando se realice la coincidencia de campos, seleccione una columna para la generación de claves susceptible de proporcionar grupos útiles para las necesidades de coincidencia. Por ejemplo, una columna Apellido es susceptible de proporcionar datos de clave de grupo más relevantes que una columna Nombre.

Sin embargo, no utilice la columna Apellido si prevé seleccionar esa columna para un análisis de duplicados en la transformación de coincidencia.

La transformación de generador de claves también puede crear un ID único para cada registro. Cada registro que entra en la transformación de coincidencia debe contener un ID único. Utilice la transformación de generador de claves para crear ID para los datos en el caso de que no exista ninguno.

Estrategia Soundex

La estrategia Soundex analiza palabras y crea claves de grupo mediante códigos alfanuméricos que representan la pronunciación de las palabras.

Los códigos de Soundex comienzan por la primera letra de la palabra, seguida de una serie de números que representan consonantes sucesivas. Utilice la estrategia Soundex para asignar el mismo código a palabras que suenan de forma similar. Configure la profundidad de Soundex para definir el número de caracteres alfanuméricos que devuelve la estrategia.

Esta estrategia se centra en el sonido de las palabras, más que en la ortografía, y puede agrupar pronunciaciones alternativas y variaciones mínimas de la pronunciación. Por ejemplo, los códigos de Soundex para `Smyth` y `Smith` son los mismos.

La estrategia Soundex también puede agrupar palabras mal pronunciadas. Por ejemplo, los códigos de Soundex para `Edmonton` y `Edmonson` son los mismos.

Propiedades de estrategia Soundex

Configure las propiedades de la estrategia Soundex para determinar los ajustes Soundex que utiliza la transformación de generador de claves para crear una clave de grupo.

En la siguiente tabla se describen las propiedades de la estrategia Soundex:

Propiedad	Descripción
Profundidad de Soundex	Determina el número de caracteres alfanuméricos devueltos por la estrategia Soundex. La profundidad predeterminada es 3. Esta profundidad crea un código Soundex que consta de la primera letra de la cadena y dos números que representan los dos siguientes sonidos consonánticos diferentes.

TEMAS RELACIONADOS

- [“Propiedades de la estrategia de cadena” en la página 397](#)
- [“Propiedades de la creación de claves” en la página 399](#)
- [“Cómo configurar una estrategia de agrupación” en la página 398](#)

Estrategia de cadena

La estrategia de cadena crea claves de grupo desde subcadenas en datos de entrada.

Se puede especificar la longitud y la ubicación de una subcadena dentro de la columna de entrada. Por ejemplo, puede configurar esta estrategia para crear una clave a partir de los cuatro primeros caracteres de la cadena de entrada.

Propiedades de la estrategia de cadena

Configure las propiedades de la estrategia de cadena para determinar las subcadenas que utiliza la transformación de generador de claves para crear una clave de grupo.

En la siguiente tabla se describen las propiedades de la estrategia de cadena:

Propiedad	Descripción
Empezar por la izquierda	Configura la transformación para leer el campo de entrada de izquierda a derecha.
Empezar por la derecha	Configura la transformación para leer el campo de entrada de derecha a izquierda.
Posición de inicio	Especifica el número de caracteres que omitir. Por ejemplo, si especifica 3 para la Posición de inicio , la subcadena comienza por el cuarto carácter del campo de entrada, empezando por el lado especificado.
Longitud	Especifica la longitud de la cadena que utilizar como clave de grupo. Especifique 0 para utilizar el campo de entrada completo.

TEMAS RELACIONADOS

- [“Propiedades de estrategia Soundex” en la página 396](#)
- [“Propiedades de la creación de claves” en la página 399](#)
- [“Cómo configurar una estrategia de agrupación” en la página 398](#)

Estrategia NYSIIS

La estrategia NYSIIS analiza palabras y crea claves de grupo mediante letras que representan la pronunciación de las palabras.

Mientras que la estrategia Soundex solamente tiene en cuenta la primera vocal de una cadena, la estrategia NYSIIS analiza todas las vocales incluidas en una cadena. La estrategia NYSIIS convierte todas las letras a uno de seis caracteres y convierte la mayoría de vocales a la letra A.

Puertos de salida del generador de claves

Los puertos de salida de la transformación de generador de claves crean identificadores y claves de grupo que la transformación de coincidencia utiliza para procesar registros.

En la siguiente tabla se describen los puertos de salida para la transformación del generador de claves:

Propiedad	Descripción
SequenceID	Crea un ID que identifica cada registro en el conjunto de datos de origen.
GroupKey	Crea las claves de grupo que utiliza la transformación de coincidencia para procesar registros.

Cuando cree una transformación reutilizable del generador de claves, utilice la ficha **Resumen** para ver los puertos. Cuando añada una transformación no reutilizable a una asignación, utilice la ficha **Puertos** de la vista **Propiedades** para ver los puertos.

Cómo configurar una estrategia de agrupación

Para configurar una estrategia de agrupación, edite las propiedades en la vista **Estrategias**.

Antes de configurar una estrategia de generador de claves, añada puertos de entrada a la transformación de generador de claves.

1. Seleccione la vista **Estrategias**.
2. Haga clic en el botón **Nueva**.
3. Seleccione una estrategia de agrupación.
4. Haga clic en **Aceptar**.
5. En la columna **Entradas**, seleccione un puerto de entrada.
6. Configure las propiedades de la estrategia haciendo clic en la flecha de selección del campo de propiedades.
7. Configure las propiedades de la creación de claves.

TEMAS RELACIONADOS

- [“Propiedades de estrategia Soundex” en la página 396](#)
- [“Propiedades de la estrategia de cadena” en la página 397](#)
- [“Propiedades de la creación de claves” en la página 399](#)

Propiedades de la creación de claves

Configure las propiedades de creación de claves apropiadas para los datos que se desea analizar.

En la tabla siguiente se describen las propiedades de creación de claves:

Propiedad	Descripción
Ordenar resultados	Ordena la salida de la transformación de generador de claves mediante el campo GroupKey. En el caso de operaciones de coincidencia de campos, debe seleccionar esta opción o comprobar que se proporciona la transformación de coincidencia con datos ordenados. No seleccione esta opción para operaciones de coincidencia de identidad.
Generar clave de secuencia automáticamente	Genera un campo de clave de secuencia según el orden de los datos de entrada.
Usar campo como clave de secuencia	Genera un campo de secuencia para la columna especificada.
Campo de clave de secuencia	Especifica el nombre del campo de la clave de secuencia.

TEMAS RELACIONADOS

- [“Propiedades de estrategia Soundex” en la página 396](#)
- [“Propiedades de la estrategia de cadena” en la página 397](#)
- [“Cómo configurar una estrategia de agrupación” en la página 398](#)

Propiedades avanzadas de Transformación de generador de claves

La transformación de generador de claves contiene propiedades avanzadas para determinar el comportamiento de la memoria caché y el nivel de seguimiento.

Puede configurar las siguientes propiedades avanzadas:

Directorio de archivo de memoria caché

Especifica el directorio en el que el servicio de integración de datos escribe los datos temporales de la transformación actual. El servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos de entrada es mayor que la memoria disponible del sistema. El servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del servicio de integración de datos. El servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de archivo de memoria caché

Determina la cantidad de memoria del sistema que utiliza el servicio de integración de datos para ordenar los datos de entrada en la transformación. Puede utilizar un parámetro para especificar el tamaño del archivo de memoria caché.

Antes de ordenar los datos, el servicio de integración de datos asigna la cantidad de memoria que se haya especificado. Si la operación de ordenación genera una cantidad de datos mayor, el servicio de integración de datos escribirá los datos en exceso en el directorio de memoria caché. Si la operación de clasificación necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

Si no especifica un tamaño de archivo de memoria caché, la transformación aplicará el valor de memoria máximo a las opciones de ejecución del servicio de integración de datos.

La transformación lee el valor en bytes. El valor predeterminado es 400 000 bytes. El valor máximo es 2 147 483 647 bytes. Puede utilizar un parámetro para especificar el tamaño del archivo de memoria caché.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación del generador de claves en un entorno no nativo

El procesamiento de transformaciones del generador de claves en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de generador de claves en el motor de Blaze

La transformación del generador de claves es compatible con las siguientes restricciones:

- Cuando ejecuta una asignación con una transformación de generador de claves en más de un nodo, es posible que la opción de generación de secuencia en la transformación no produzca valores de ID únicos.

Transformación de generador de claves en el motor de Spark

La transformación del generador de claves es compatible con las siguientes restricciones:

- Cuando ejecuta una asignación con una transformación de generador de claves en más de un nodo, es posible que la opción de generación de secuencia en la transformación no produzca valores de ID únicos.

Transformación de generador de claves en el motor de Databricks Spark

La transformación del generador de claves es compatible con las siguientes restricciones:

- Cuando ejecuta una asignación con una transformación de generador de claves en más de un nodo, es posible que la opción de generación de secuencia en la transformación no produzca valores de ID únicos.

CAPÍTULO 24

Transformación de etiquetador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de etiquetador, 401](#)
- [Cuando utilizar una transformación de etiquetador, 402](#)
- [Uso de los datos de referencia en la transformación de etiquetador, 403](#)
- [Estrategias de transformación de etiquetador, 405](#)
- [Puertos de transformación de etiquetador, 406](#)
- [Propiedades de etiquetado de caracteres, 407](#)
- [Propiedades de etiquetado con tokens, 409](#)
- [Cómo configurar una estrategia de etiquetado de caracteres, 412](#)
- [Cómo configurar una estrategia de etiquetado de tokens, 413](#)
- [Propiedades avanzadas de Transformación de etiquetador, 413](#)
- [Transformación de etiquetador en un entorno no nativo, 414](#)

Resumen de la transformación de etiquetador

La transformación de etiquetador es una transformación pasiva que analiza los campos del puerto de entrada campos y escribe etiquetas de texto que describen los datos de cada campo.

Utilice una transformación de etiquetador si desea conocer los tipos de información que un puerto contiene. Utilice una transformación de etiquetador si no conoce los tipos de información en un puerto o si desea para identificar los registros que no contienen los tipos de información esperados en un puerto.

Una etiqueta es una cadena de uno o más caracteres que describe una cadena de entrada. Configure la transformación de etiquetador para asignar etiquetas a cadenas de entrada en función de los datos que contiene cada cadena.

Cuando configure la transformación, especifique los tipos de carácter o de cadena que buscar y especifique la etiqueta que la transformación escribe como salida cuando encuentra el carácter o cadena asociados. Especifique los tipos de carácter y cadena que buscar y las etiquetas que utilizar cuando configure una operación de etiquetado. También puede utilizar los objetos de datos de referencia para especificar los caracteres, las cadenas y las etiquetas.

Configure la transformación para realizar el etiquetado de caracteres o el etiquetado de tokens:

Etiquetado de caracteres

Escribe una etiqueta que describe la estructura de caracteres de la cadena de entrada, incluyendo puntuación y espacios. La transformación escribe una sola etiqueta para cada fila en una columna. Por ejemplo, la transformación de etiquetador puede etiquetar el código postal 10028 como "nnnnn", donde "n" representa un carácter numérico.

Etiquetado de tokens

Escribe una etiqueta que describe el tipo de información en la cadena de entrada. La transformación escribe una etiqueta para cada token identificado en los datos de entrada. Por ejemplo, puede configurar la transformación de etiquetador para que etiquete la cadena "John J. Smith" con los tokens "Word Init Word".

Un token es un valor delimitado en una cadena de entrada.

Cuando el etiquetador encuentra un carácter o una cadena que coincide con una etiqueta que haya especificado, escribe el nombre de la etiqueta en un nuevo puerto de salida.

La transformación de etiquetador utiliza datos de referencia para identificar caracteres y tokens. Seleccione el objeto de datos de referencia al configurar una operación en una estrategia del etiquetador.

Cuando utilizar una transformación de etiquetador

La transformación de etiquetador escribe una etiqueta descriptiva para cada valor de un puerto.

Los ejemplos siguientes describen algunos de los tipos de análisis que puede realizar con una transformación de etiquetador.

Buscar registros con datos de contacto

Configure la transformación con una tabla de referencia que contenga una lista de nombres de pila. Cree una estrategia de etiquetado de tokens para etiquetar cualquier cadena que coincida con un valor de la tabla de referencia. Cuando revise los datos de salida, cualquier registro que contiene la etiqueta es probable que identifique a una persona.

Buscar registros de negocio

Configure la transformación con un conjunto de tokens que contenga una lista de sufijos empresariales, tales como Inc, Corp y Ltd. Cree una estrategia de etiquetado de tokens para etiquetar cualquier cadena que coincida con un valor de la tabla de referencia. Cuando revise los datos de salida, cualquier registro que contiene la etiqueta es probable que identifique a una empresa.

Nota: Use un conjunto de tokens de sufijos empresariales que desee para identificar cualquier nombre empresarial. Puede usar una tabla de referencia de nombres empresariales si está seguro de que la tabla contiene todos las empresas que desea identificar. Por ejemplo, puede utilizar una tabla de referencia que enumere las corporaciones de la Bolsa de Nueva York.

Buscar datos de número de teléfono

Configure la transformación con un conjunto de caracteres que defina la estructura de caracteres de un número de teléfono. Por ejemplo, puede utilizar un conjunto de caracteres que reconozca diferentes patrones de símbolos de puntuación y dígitos como números de teléfono de Estados Unidos. Puede revisar los datos para buscar registros que no contengan los dígitos correctos para un número de teléfono.

Las etiquetas de caracteres pueden utilizar los siguientes caracteres para analizar los datos de columna:

c=punctuation character n=digit s=space

La siguiente tabla muestra estructuras de números de teléfono de muestra:

Estructura de caracteres	Número de teléfono
cnnncsnnncnnncnnnnn	(212) 555-1212
nnnnnnnnnn	2125551212
cnnncnnncnnnn	+212-555-1212

Uso de los datos de referencia en la transformación de etiquetador

Informatica Developer se instala con distintos tipos de objetos de datos de referencia que puede utilizar con la transformación de etiquetador. También puede crear objetos de datos de referencia.

Cuando añade un objeto de datos de referencia a la estrategia de transformación de etiquetador, la transformación busca en los datos de entrada de la estrategia los valores del objeto de datos de referencia. La transformación reemplaza cualquier valor que encuentra con un valor válido del objeto de datos de referencia, o con un valor que especifique.

La siguiente tabla describe los tipos de datos de referencia que puede utilizar:

Tipo de datos de referencia	Descripción
Conjuntos de caracteres	Identifica los distintos tipos de caracteres, tales como las letras, números, y símbolos de puntuación. Usar en las operaciones de etiquetado de caracteres.
Modelos probabilísticos	Añade capacidades de coincidencia difusa para las operaciones de etiquetas de tokens. La transformación puede utilizar un modelo probabilístico para inferir el tipo de información de una cadena. Para habilitar las capacidades de coincidencia difusa, puede compilar el modelo probabilístico en Developer Tool. Usar en las operaciones de etiquetado con tokens.
Tablas de referencia	Busca las cadenas que coinciden con las entradas en una tabla de base de datos. Usar en las operaciones de etiquetado de caracteres y de etiquetado con tokens.

Tipo de datos de referencia	Descripción
Expresiones regulares	Identifica las cadenas que coinciden con las condiciones que defina. Puede utilizar una expresión regular para buscar una cadena dentro de una cadena más grande. Usar en las operaciones de etiquetado con tokens.
Conjuntos de tokens	Identifica las cadenas según los tipos de información que contienen. Usar en las operaciones de etiquetado con tokens. Informatica se instala con conjuntos de tokens y distintos tipos de definiciones de tokens, tales como definiciones de palabra, número de teléfono, código postal y de códigos de producto.

Juegos de caracteres

Un juego de caracteres contiene expresiones que identifican caracteres e intervalos de caracteres específicos. Puede utilizar juegos de caracteres en transformaciones de etiquetador que utilizan el modo de análisis de tokens.

Los rangos de caracteres especifican un rango secuencial de los códigos de carácter. Por ejemplo, el rango de caracteres "[A-C]" coincide con los caracteres en mayúscula "A", "B" y "C". Este rango de caracteres no coincide con los caracteres en minúscula "a", "b", o "c".

Utilice juegos de caracteres para identificar un carácter o un rango de caracteres como parte de las operaciones de análisis o de etiquetado. Por ejemplo, puede etiquetar todos los números de una columna que contiene números de teléfono. Después de etiquetar los números, puede identificar patrones con una transformación de analizador y escribir patrones problemáticos en puertos de salida aparte.

Modelos probabilísticos

Un modelo probabilístico identifica tokens por los tipos de información que contienen y por las posiciones que ocupan en una cadena de entrada.

Un modelo probabilístico contiene valores de datos de referencia y valores de etiqueta. Los valores de datos de referencia representan los datos de un puerto de entrada que conecta a la transformación. Los valores de etiqueta describen los tipos de información que contienen los valores de datos de referencia. Puede asignar una etiqueta a cada valor de datos de referencia del modelo.

Para vincular los valores de datos de referencia a las etiquetas de un modelo probabilístico, debe compilar el modelo. El proceso de compilación genera una serie de asociaciones lógicas entre los valores de datos y las etiquetas. Cuando se ejecuta una asignación que lee el modelo, el Servicio de integración de datos aplica la lógica del modelo a los datos de entrada de la transformación. El Servicio de integración de datos devuelve la etiqueta que mejor describe la información de cada valor de datos.

Cree un modelo probabilístico en Developer tool. El repositorio de modelos almacena el objeto del modelo probabilístico. Developer tool escribe los valores de datos, las etiquetas y los datos de compilación en un archivo de la estructura del directorio de Informatica.

Tablas de referencia

Una tabla de referencia es una tabla de base de datos que contiene al menos dos columnas. Una columna contiene la versión estándar o requerida de un valor de datos y otras columnas contienen versiones alternativas del valor. Cuando añade una tabla de referencia a una transformación, la transformación busca

los datos de puertos de entrada para los valores que también aparecen en la tabla. Puede crear tablas con cualquier dato que sea útil para el proyecto de datos en el que esté trabajando.

Expresiones regulares

En el contexto de las operaciones de etiquetado, una expresión regular es una expresión que se puede usar para identificar una cadena específica en datos de entrada. Puede utilizar expresiones regulares en transformaciones de etiquetador que utilizan el modo de etiquetado de tokens.

Las transformaciones de etiquetador utilizan expresiones regulares para hacer coincidir un patrón de entrada y crear una etiqueta individual. Las expresiones regulares que tienen varias salidas no generan varias etiquetas.

Conjuntos de tokens

Un conjunto de tokens contiene expresiones que identifican tokens específicos. Puede utilizar conjuntos de tokens en transformaciones de etiquetador que utilizan el modo de etiquetado con tokens.

Emplee conjuntos de tokens para identificar tokens específicos como parte de las operaciones de etiquetado de tokens. Por ejemplo, puede utilizar un conjunto de tokens para etiquetar todas las direcciones de correo electrónico que utilizan un formato "AccountName@DomainName". Después de etiquetar los tokens, puede utilizar la transformación de analizador para escribir direcciones de correo electrónico en los puertos de salida que especifique.

Developer Tool incluye conjuntos de tokens definidos por el sistema que pueden usarse para identificar una amplia gama de patrones. Algunos ejemplos de conjuntos de tokens definidos por el sistema son:

- Palabras
- Números
- Números de teléfono
- Direcciones de correo electrónico
- Códigos postales
- Números de identificación nacional, como números de la seguridad social
- Números de tarjeta de crédito

Estrategias de transformación de etiquetador

Utilice estrategias de etiquetado para asignar etiquetas a los datos de entrada. Para configurar una estrategia del etiquetador, edite los ajustes en la vista **Estrategias** de una transformación del etiquetador.

Cuando se crea una estrategia de etiquetado, se añade una o más operaciones. Cada operación implementa una tarea de etiquetado específica.

La transformación de etiquetador proporciona un asistente que puede usarse para crear estrategias. Cuando se crea una estrategia de etiquetado, se elige entre el modo de etiquetado de caracteres o el modo de etiquetado de tokens. A continuación se añaden las operaciones específicas a ese modo de etiquetado.

Importante: Puede modificar el orden de las operaciones y de las estrategias. El orden de las operaciones dentro de una estrategia puede cambiar la salida de ésta, puesto que cada operación lee los resultados de la operación precedente.

Operaciones de etiquetado de caracteres

Utilice operaciones de etiquetado de caracteres para crear etiquetas que describan los patrones de carácter en los datos.

Se pueden añadir los siguientes tipos de operación a una estrategia de etiquetado de caracteres:

Caracteres de etiqueta que usan juegos de caracteres

Caracteres de etiqueta que usan juegos de caracteres predefinidos, como dígitos o caracteres alfabéticos. Se pueden seleccionar juegos de caracteres Unicode y distintos de Unicode.

Caracteres de etiqueta que usan una tabla de referencia

Caracteres de etiqueta con etiquetas personalizadas de una tabla de referencia.

Operaciones de etiquetado con tokens

Utilice operaciones de etiquetado con tokens para crear etiquetas que describan las cadenas de los datos.

La transformación de etiquetador puede identificar y etiquetar varios tokens de una cadena de entrada. Por ejemplo, puede configurar la transformación de etiquetador para que utilice los conjuntos de tokens Números de teléfono de Estados Unidos y Direcciones de correo electrónico. Cuando la transformación de etiquetador procesa la cadena de entrada "555-555-1212 someone@somewhere.com", la cadena de salida es "USPHONE EMAIL".

Se pueden añadir los siguientes tipos de operación de etiquetado con tokens a una estrategia de etiquetado:

Etiqueta con tabla de referencia

Cadenas de etiquetas que coinciden con entradas de tabla de referencia.

Tokens de etiqueta con conjunto de tokens

Patrones de cadenas de etiqueta que coinciden con datos del conjunto de tokens o con datos de modelos probabilísticos.

Puertos de transformación de etiquetador

Seleccione los puertos de entrada y salida necesarios para las operaciones de etiquetado que haya configurado en la transformación.

Las transformaciones de etiquetador utilizan el siguiente puerto:

Puertos de entrada

Lee entradas de cadena desde objetos de un nivel superior.

Puertos de salida etiquetada

Escribe las etiquetas definidas por las operaciones de transformación.

Puertos de salida agrupados

Hace pasar las cadenas de entrada que corresponden a cada etiqueta de la salida. Seleccione este puerto si va a añadir una transformación del analizador a un nivel inferior de la transformación de etiquetador en un maplet o asignación y va a configurar la transformación de analizador para ejecutar en modo de análisis basado en patrones. La transformación de analizador asocia la salida del etiquetado de tokens con los datos en los puertos de salida con tokens.

Puertos de salida de puntuación

Seleccione escribir los valores de puntuación generados por técnicas de coincidencia probabilística en una operación de etiquetado de tokens.

Al ejecutar una operación de etiquetado de tokens que utiliza un modelo probabilístico, la operación genera una puntuación numérica para cada cadena etiquetada. La puntuación representa el grado de similitud entre la cadena de entrada y los patrones definidos en el modelo probabilístico.

Propiedades de etiquetado de caracteres

Configure las propiedades para operaciones de etiquetado de caracteres en la vista **Estrategias** de la transformación de etiquetador.

Propiedades generales

Las propiedades generales se aplican a todas las operaciones de etiquetado de caracteres que haya definido en la estrategia. Utilice las propiedades generales para nombrar la estrategia y especifique los puertos de entrada y salida.

En la tabla siguiente, se describen las propiedades generales:

Propiedad	Descripción
Nombre	Proporciona un nombre para la estrategia.
Entradas	Identifica los puertos de entrada que las operaciones de estrategia pueden leer.
Salidas	Identifica los puertos de salida en los que las operaciones de estrategia pueden escribir.
Descripción	Proporciona una descripción de texto de la estrategia. Esta propiedad es opcional.

Propiedades de la tabla de referencia

Cuando defina una estrategia de etiquetado de caracteres, puede añadir operaciones para etiquetar con conjuntos de caracteres y con tablas de referencia. Utilice las propiedades de la tabla de referencia para especificar cómo la transformación utiliza las tablas de referencia.

La tabla siguiente describe las propiedades de la tabla de referencia:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Tabla de referencia	Especifica las tablas de referencia que utiliza la transformación para etiquetar caracteres

Propiedad	Descripción
Etiqueta	Especifica el texto de reemplazo para caracteres de entrada que coinciden con las entradas de la tabla de referencia.
Reemplazar otras etiquetas de la estrategia	Determina si esta operación de etiquetado reemplaza otras operaciones de etiquetado.

Propiedades de los conjuntos de caracteres

Cuando defina una estrategia de etiquetado de caracteres, puede añadir operaciones para etiquetar con conjuntos de caracteres y con tablas de referencia. Utilice las propiedades del conjunto de caracteres para especificar cómo utiliza la transformación los conjuntos de caracteres.

La siguiente tabla describe las propiedades del conjunto de caracteres:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Seleccionar conjuntos de caracteres	Especifica los conjuntos de caracteres que utiliza la transformación para etiquetar cadenas. Puede reemplazar el texto de sustitución en las cadenas de entrada que coinciden con el conjunto de caracteres. Haga clic en la flecha de selección de la columna Etiqueta para introducir texto de sustitución personalizado.
Filtrar texto	Utiliza los caracteres o los comodines que haya especificado para filtrar la lista de conjuntos de caracteres.
Añadir conjunto de caracteres	Seleccione esta opción para definir un conjunto de caracteres personalizado.
Editar	Edite el contenido de un conjunto de caracteres personalizado.
Importar	Permite crear copias no reutilizables de conjuntos de caracteres almacenados en los conjuntos de contenido. Los cambios en el conjunto de caracteres original no actualizan las copias almacenadas en la transformación de etiquetador.
Quitar	Elimina un conjunto de caracteres personalizado.
Especificar orden de ejecución	Establece el orden en el que la operación aplica los conjuntos de tokens a los datos. Utilice las flechas arriba y abajo para cambiar el orden.

Propiedades de filtro

Puede especificar los valores para omitirlos durante una operación de etiquetado. Utilice las propiedades **Omitir texto** para especificar valores que no se aplicarán a las operaciones de etiquetado.

La siguiente tabla describe las propiedades de filtro:

Propiedad	Descripción
Término de búsqueda	Especifica las cadenas que filtra la transformación antes de realizar el etiquetado. Utilice esta función para especificar las excepciones para la estrategia de etiquetado definida.
Distinguir mayúsculas de minúsculas	Determina si las cadenas filtradas deben coincidir con el formato de mayúsculas y minúsculas del término de búsqueda.
Mayúsculas	Convierte las cadenas filtradas a mayúscula.
Iniciar	Especifica la posición de carácter a partir de la cual comenzar a buscar cadenas filtradas.
Finalización	Especifica la posición de carácter en la que dejar de buscar cadenas filtradas.

Propiedades de etiquetado con tokens

Configure las propiedades para operaciones de etiquetado con tokens en la vista **Estrategias** de la transformación de etiquetador.

Propiedades generales

Las propiedades generales se aplican a todas las operaciones de etiquetado con tokens que haya definido en la estrategia. Utilice las propiedades generales para dar un nombre a la estrategia, especificar los puertos de entrada y salida y especificar si la estrategia permite técnicas de coincidencia probabilística.

En la tabla siguiente, se describen las propiedades generales:

Propiedad	Descripción
Nombre	Proporciona un nombre para la estrategia.
Entradas	Identifica los puertos de salida que las operaciones de estrategia pueden leer.
Salidas	Identifica el puerto de salida en los que las operaciones de estrategia pueden escribir.
Descripción	Describe la estrategia. La propiedad es opcional.
Utilizar técnicas de coincidencias probabilísticas	Especifica que la estrategia puede utilizar un modelo probabilístico para identificar los tipos de tokens.
Inversión habilitada	Indica que la estrategia lee datos de entrada de derecha a la izquierda. Esta propiedad está deshabilitada para la coincidencia probabilística.

Propiedad	Descripción
Delimitadores	Especifica los caracteres que utiliza la transformación para evaluar subcadenas en los datos de entrada. El valor predeterminado es espacio. La propiedad está deshabilitada en el etiquetado probabilístico.
Campo de salida con tokens	Indica que la estrategia escribe varias etiquetas en un puerto de salida. Seleccione este campo para crear datos de entrada para los análisis basados en patrones en la transformación de analizador.
Campo de puntuación de salida	Identifica el campo que contienen los valores de puntuación generados en la coincidencia probabilística. Establezca el campo de salida de la puntuación cuando seleccione la opción para utilizar técnicas de coincidencias probabilísticas.
Delimitador de salida	Especifica un carácter para separar valores de datos del puerto de salida. El valor predeterminado es un signo de dos puntos.

Propiedades de conjuntos de tokens

Las propiedades del conjunto de tokens se aplican cuando configura una operación de etiquetado para utilizar conjuntos de tokens.

En la tabla siguiente, se describen las propiedades generales:

Propiedad	Descripción
Seleccionar conjuntos de tokens	Especifica los conjuntos de tokens que utiliza la transformación para etiquetar cadenas.
Filtrar texto	Filtra la lista de conjuntos de tokens o de expresiones regulares. Utilice caracteres de texto y caracteres comodín como filtro.
Añadir conjunto de tokens	Utilícela para definir conjuntos de tokens personalizados.
Añadir expresión regular	Úsela para definir expresiones regulares que coincidan con un patrón de entrada.
Editar	Edita el contenido de un conjunto de tokens personalizado o una expresión regular.
Importar	Importa una copia no reutilizable de un conjunto de tokens o una expresión regular desde una carpeta del repositorio de modelos. Si actualiza el objeto de origen del conjunto de tokens o la expresión regular, el servicio de integración de datos no actualizará la copia no reutilizable.
Eliminar	Elimina un conjunto de tokens personalizados o una expresión regular.
Especificar orden de ejecución	Establece el orden en el que la operación aplica los conjuntos de tokens o las expresiones regulares en los datos. Utilice las flechas ascendentes y descendentes para cambiar el orden.

Propiedades de etiqueta personalizada

Cuando configura una operación de etiqueta de tokens, puede seleccionar la vista **Etiqueta personalizada** para crear etiquetas para términos de búsqueda específicos.

La tabla siguiente describe las propiedades de etiquetas personalizadas:

Propiedad	Descripción
Término de búsqueda	Identifica la cadena que desea buscar.
Distinguir mayúsculas de minúsculas	Especifica si el formato de mayúsculas y minúsculas de las cadenas de entrada debe coincidir con el del término de búsqueda.
Etiqueta personalizada	Identifica la etiqueta personalizada que aplicar.

Propiedades de coincidencia probabilística

Cuando selecciona las opciones para utilizar las técnicas de coincidencias probabilísticas, puede añadir un modelo probabilístico a la operación de etiquetado. No puede añadir un modelo probabilístico a una estrategia que utilice un conjunto de tokens o una tabla de referencia.

La siguiente tabla describe las propiedades asociadas con coincidencias probabilísticas:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Filtrar texto	Utiliza los caracteres o comodines que especifique para filtrar la lista de modelos probabilísticos en el repositorio.
Modelo probabilístico	Identifica el modelo probabilístico que utilizar en la operación.

Propiedades de la tabla de referencia

Las propiedades de la tabla de referencia se aplican al configurar una operación de etiquetado para usar una tabla de referencia.

La tabla siguiente describe las propiedades de la tabla de referencia:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Tabla de referencia	Especifica la tabla de referencia que utiliza la operación para etiquetar tokens.
Etiqueta	Especifica el texto que la operación escribe en un nuevo puerto si una cadena de entrada coincide con una tabla de referencia entrada.
Distinguir mayúsculas de minúsculas	Determina si el formato de mayúsculas y minúsculas de las cadenas de entrada debe coincidir con el de las entradas de la tabla de referencia.

Propiedad	Descripción
Reemplazar coincidencias por valores válidos	Reemplaza cadenas etiquetadas con la entrada de una columna Válida en la tabla de referencia.
Modo	Determina el método de etiquetado de tokens. Seleccione Inclusivo para etiquetar cadenas que coinciden con entradas de la tabla de referencia. Seleccione Exclusivo para etiquetar cadenas que no coinciden con entradas de la tabla de referencia.
Establecer prioridad	Determina si las operaciones de etiquetado de la tabla de referencia tienen prioridad sobre las operaciones de etiquetado del conjunto de tokens en una estrategia. Si establece esta propiedad, la transformación lleva a cabo el etiquetado de la tabla de referencia antes del etiquetado del conjunto de tokens y el análisis del conjunto de tokens no puede sobrescribir el análisis de la etiqueta de la tabla de referencia.

Cómo configurar una estrategia de etiquetado de caracteres

Para configurar una estrategia del etiquetador, edite los ajustes en la vista **Estrategias** de la transformación del etiquetador.

1. Seleccione la vista **Estrategias** y haga clic en **Nuevo** para crear una estrategia.
Se abre el asistente **Estrategia**.
2. Especifique un nombre para la estrategia.
3. Haga clic en los campos **Entradas** y **Salidas** para definir los puertos de la estrategia.
4. Opcionalmente puede introducir una descripción de la estrategia.
5. Seleccione el modo de etiquetado de los caracteres.
6. Haga clic en **Siguiente**.
7. Seleccione el tipo de operación de etiquetado de caracteres que desee configurar. Puede configurar las siguientes operaciones:
 - Etiquetar caracteres usando tablas de referencia.
 - Etiquetar caracteres usando conjuntos de caracteres.
8. Haga clic en **Siguiente**.
9. Configure las propiedades de la operación y haga clic en **Siguiente**.
10. Opcionalmente puede configurar propiedades **Omitir texto**.
11. Haga clic en **Siguiente** para añadir más operaciones a la estrategia, o haga clic en **Finalizar**.

Puede cambiar el orden en el que la transformación procesa las estrategias y las operaciones. En la vista **Estrategias**, seleccione una estrategia u operación y haga clic en **Mover hacia arriba** o **Mover hacia abajo**.

Cómo configurar una estrategia de etiquetado de tokens

Para configurar una estrategia del etiquetador, edite los ajustes en la vista **Estrategias** de la transformación del etiquetador.

1. Seleccione la vista **Estrategias** y haga clic en **Nuevo** para crear una estrategia.
Se abre el asistente **Estrategia**.
2. Especifique un nombre para la estrategia.
3. Haga clic en los campos **Entradas** y **Salidas** para definir los puertos de la estrategia.
4. Opcionalmente puede introducir una descripción de la estrategia.
5. Seleccione el modo de etiquetado de tokens.
Compruebe o edite las propiedades para el modo que ha seleccionado.
6. Haga clic en **Siguiente**.
7. Seleccione el tipo de operación de etiquetado de tokens que desee configurar. Puede configurar las siguientes operaciones:
 - Etiquetar tokens con un conjunto de tokens.
 - Etiquetar tokens con una tabla de referencia.
 - Etiquetar tokens usando la coincidencia probabilística.
8. Haga clic en **Siguiente**.
9. Configure las propiedades de la operación y haga clic en **Siguiente**.
Si configura la estrategia para utilizar la coincidencia probabilística, especifique una etiqueta para utilizar para los tokens identificados como datos de residuos.
10. Opcionalmente puede configurar las propiedades de **Etiqueta personalizada**.
11. Haga clic en **Siguiente** para añadir más operaciones a la estrategia o haga clic en **Finalizar**.

Puede cambiar el orden en el que la transformación procesa las estrategias y las operaciones. En la vista **Estrategias**, seleccione una estrategia u operación y haga clic en **Mover hacia arriba** o **Mover hacia abajo**.

Propiedades avanzadas de Transformación de etiquetador

Configure propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de etiquetador.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de etiquetador en un entorno no nativo

El procesamiento de transformaciones de etiquetador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 25

Transformación de búsqueda

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de búsqueda, 415](#)
- [Búsquedas conectadas y no conectadas, 416](#)
- [Desarrollar una transformación de búsqueda, 419](#)
- [Consulta de búsqueda, 419](#)
- [Filtro de origen de búsqueda, 422](#)
- [Condición de búsqueda, 422](#)
- [Memoria caché de búsqueda, 425](#)
- [Propiedades de consulta, 426](#)
- [Transformaciones de búsqueda en asignaciones dinámicas, 426](#)
- [Definir puertos dinámicos, 427](#)
- [Cambiar el origen de búsqueda, 427](#)
- [Selectores de puerto, 432](#)
- [Propiedades de tiempo de ejecución, 438](#)
- [Propiedades avanzadas, 439](#)
- [Crear una transformación de búsqueda reutilizable, 441](#)
- [Crear una transformación de búsqueda no reutilizable, 442](#)
- [Crear una transformación de búsqueda no conectada, 443](#)
- [Ejemplo de búsqueda no conectada, 444](#)
- [Transformación de búsqueda en un entorno no nativo, 446](#)

Resumen de la transformación de búsqueda

La transformación de búsqueda es una transformación pasiva o activa que busca datos en un archivo sin formato, objeto de datos lógicos, tabla de referencia o tabla relacional. La transformación de búsqueda puede devolver una o varias filas desde una búsqueda.

Antes de crear una transformación de búsqueda, cree el origen de búsqueda. Importe un archivo sin formato o una tabla de base de datos relacional como un objeto de datos físicos. También puede crear un objeto de datos lógicos o una tabla de referencia para utilizar como un origen de búsqueda. Cuando se crea una transformación de búsqueda, Developer tool añade las columnas del objeto de datos o tabla de referencia como puertos de búsqueda en la transformación. Después de crear la transformación, configure uno o más

puertos de salida para que devuelvan los resultados de la búsqueda. Configure las condiciones de búsqueda, así como otras propiedades de búsqueda.

Cuando se ejecuta una asignación o se obtiene una vista previa de datos, el Servicio de integración consulta el origen de búsqueda. El Servicio de integración consulta el origen de búsqueda en función de los puertos de búsqueda de la transformación, las propiedades de búsqueda y la condición de búsqueda. La transformación de búsqueda devuelve el resultado de la búsqueda al destino o a otra transformación.

Se puede configurar una transformación de búsqueda conectada o no conectada. Una transformación conectada se conecta a otra transformación en la asignación. Una transformación no conectada recibe la entrada desde una expresión :LKP de otra transformación. Si la transformación de búsqueda realiza una búsqueda en un objeto de datos lógicos, debe configurar una transformación de búsqueda conectada. Conecte los puertos de entrada de la transformación de búsqueda a una transformación de nivel superior o a un origen de nivel superior. Conecte los puertos de salida a una transformación de nivel inferior o a un destino de nivel inferior.

Puede usar varias transformaciones de búsqueda en una asignación.

Puede realizar las siguientes tareas con una transformación de búsqueda:

- Obtener un valor relacionado. Recupere un valor del origen de búsqueda basado en un valor de los datos de entrada. Por ejemplo, los datos de entrada contienen un ID de empleado. Recupere el nombre de empleado del origen de búsqueda por ID de empleado.
- Recuperar varias filas desde un origen de búsqueda.
- Realizar un cálculo. Recuperar un valor desde una tabla de búsquedas y usarlo en un cálculo. Por ejemplo, puede buscar un porcentaje de impuesto de venta, calcular un impuesto y devolver el impuesto a un destino.
- Realizar una búsqueda no conectada con una expresión :LKP en una transformación que acepta expresiones. Filtrar los resultados con otra expresión en la transformación.
- Parametrizar el origen de búsqueda y la condición de búsqueda para utilizar una transformación de búsqueda en una asignación dinámica.

Búsquedas conectadas y no conectadas

Puede configurar una transformación de búsqueda conectada o una transformación de búsqueda no conectada. Una transformación de búsqueda conectada es una transformación que tiene puertos de entrada y de salida que se conectan a otras transformaciones en una asignación. Una transformación de búsqueda no conectada aparece en la asignación, pero no está conectada a otras transformaciones.

Una transformación de búsqueda no conectada recibe entradas desde el resultado de una expresión :LKP en una transformación de expresión o de agregación. La expresión :LKP pasa argumentos a la transformación de búsqueda y recibe un resultado de dicha transformación. La expresión :LKP puede pasar resultados de búsqueda a otra expresión de la transformación de expresión o agregación a fin de filtrar resultados.

La siguiente tabla enumera las diferencias entre las búsquedas conectadas y las búsquedas desconectadas:

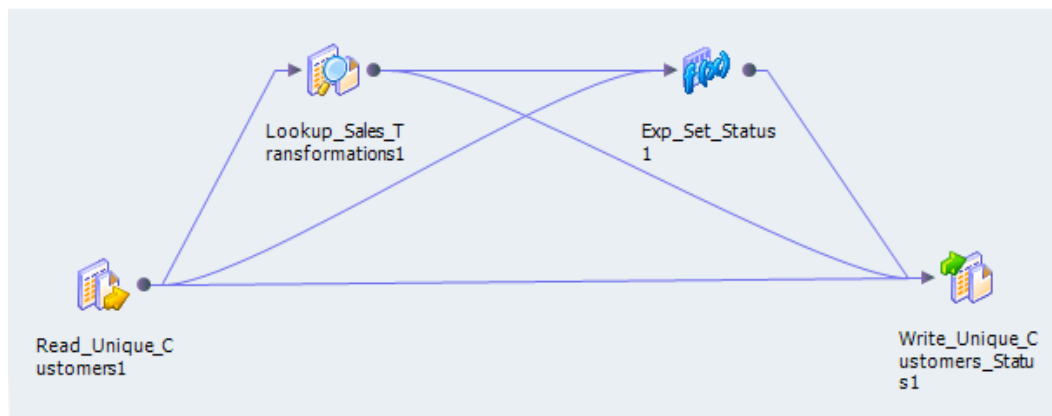
Búsqueda conectada	Búsqueda no conectada
Recibe valores de entrada directamente del canal.	Recibe valores de entrada desde el resultado de una expresión :LKP de otra transformación.
Utilice una memoria caché dinámica o estática.	Utilice una memoria caché estática.

Búsqueda conectada	Búsqueda no conectada
La memoria caché incluye las columnas de origen de búsqueda en la condición de búsqueda y las columnas de origen de búsqueda que son puertos de salida.	La memoria caché incluye todos los puertos de búsqueda y de salida en la condición de búsqueda y en el puerto de búsqueda o devolución.
Devuelve varias columnas desde la misma fila o las inserta en la memoria caché de búsqueda dinámica.	Devuelve una columna de cada fila a un puerto de devolución.
Si no hay ninguna coincidencia para la condición de búsqueda, el Servicio de integración devuelve el valor predeterminado de todos los puertos de salida. Si configura la memoria caché dinámica, el Servicio de integración inserta filas en la memoria caché o las deja sin cambios.	Si no hay ninguna coincidencia para la condición de búsqueda, el Servicio de integración devuelve NULL.
Si hay una coincidencia para la condición de búsqueda, el Servicio de integración devuelve el resultado de la condición de búsqueda para todos los puertos de búsqueda/salida. Si configura la memoria caché dinámica, el Servicio de integración actualiza la fila en la memoria caché o la deja sin cambios.	Si hay una coincidencia para la condición de búsqueda, el Servicio de integración devuelve el resultado de la condición de búsqueda a través del puerto de devolución.
Pasa varios valores de salida a otra transformación. Vincule puertos de búsqueda/salida a otra transformación.	Devuelve un valor de salida a otra transformación. El puerto de devolución de la transformación de búsqueda pasa el valor al puerto que contiene la expresión :LKP en la otra transformación.
Admite valores predeterminados definidos por el usuario.	No admite valores predeterminados definidos por el usuario.

Búsquedas conectadas

Una transformación de búsqueda conectada es una transformación de búsqueda que está conectada a un origen o un destino en una asignación.

La siguiente figura muestra una asignación con una transformación de búsqueda conectada:



Cuando se ejecuta una asignación que contiene una transformación de búsqueda conectada, el Servicio de integración realiza los siguientes pasos:

1. El Servicio de integración pasa valores desde otra transformación a los puertos de entrada de la transformación de búsqueda.
2. Para cada fila de entrada, el Servicio de integración consulta el origen de búsqueda o la memoria caché en función de los puertos de búsqueda y la condición de búsqueda en la transformación.
3. Si la transformación no está guardada en la memoria caché o utiliza una memoria caché estática, el Servicio de integración devuelve valores desde la consulta de búsqueda.

Si la transformación utiliza una memoria caché dinámica, el Servicio de integración inserta la fila en la memoria caché si no encuentra la fila en ella. Cuando el Servicio de integración busca la fila en la memoria caché, actualiza la fila en la memoria caché o la deja sin cambios. Marca la fila como insertada, actualizada o sin cambios.

4. El Servicio de integración devuelve datos desde la consulta y los pasa a la siguiente transformación de la asignación.

Si la transformación utiliza una memoria caché dinámica, puede pasar las filas a una transformación de filtro o de enrutador para filtrar filas nuevas en el destino.

Nota: Este capítulo analiza las transformaciones de búsqueda conectada a menos que se especifique lo contrario.

Búsquedas no conectadas

Una transformación de búsqueda no conectada es una transformación de búsqueda que no está conectada a un origen ni a un destino en la asignación. Invoque la búsqueda con una expresión :LKP en una transformación que permita expresiones.

La sintaxis de la expresión de búsqueda es `:LKP lookup_transformation_name(argumento, argumento, ...)`

El orden en que enumera cada argumento debe coincidir con el orden de las condiciones de búsqueda en la transformación de búsqueda. La transformación de búsqueda devuelve el resultado de la consulta mediante el puerto de devolución de la transformación de búsqueda. La transformación que invoca la búsqueda recibe el valor de resultado de búsqueda en el puerto que contiene la expresión :LKP. Si la consulta de búsqueda no consigue devolver un valor, el puerto recibe valores nulos.

Cuando se realiza una búsqueda no conectada, se puede realizar la misma búsqueda varias veces en una asignación. Puede comprobar los resultados de la búsqueda en otra expresión y filtrar las filas según los resultados.

Cuando se ejecuta una asignación que contiene una transformación de búsqueda no conectada, el servicio de integración realiza los siguientes pasos:

1. Una transformación de búsqueda no conectada recibe valores de entrada desde el resultado de una expresión: LKP en otra transformación, como puede ser una transformación de agregación, de expresión o de estrategia de actualización.
2. El servicio de integración consulta el origen de búsqueda o la memoria caché según los puertos de búsqueda y la condición establecida en la transformación de búsqueda.
3. El servicio de integración devuelve un valor a través del puerto de devolución de la transformación de búsqueda.
4. El servicio de integración pasa el valor de devolución al puerto que contiene la expresión :LKP.

Desarrollar una transformación de búsqueda

Cuando se desarrolla una transformación de búsqueda, es necesario tener en cuenta factores como el tipo de origen de búsqueda y la condición de búsqueda.

Tenga en cuenta los siguientes factores cuando desarrolle una transformación de búsqueda:

- Si desea o no crear la transformación a partir de un archivo sin formato, objeto de datos lógicos, tabla de referencia u objeto de datos relacionales. Antes de crear una transformación de búsqueda, cree el origen de búsqueda. Importe un archivo sin formato o una tabla de base de datos relacional como un objeto de datos físicos. También puede crear un objeto de datos lógicos o una tabla de referencia para utilizar como un origen de búsqueda.
- Los puertos de salida para la transformación.
- Las condiciones de búsqueda en la transformación.
- Si desea o no que el Servicio de integración guarde en la memoria caché los datos de la búsqueda. El Servicio de integración puede guardar datos en la memoria caché de archivos sin formato, tablas de referencia u objetos de datos relacionales.

Consulta de búsqueda

El Servicio de integración consulta la búsqueda según los puertos y las propiedades que se configuran en la transformación de búsqueda. El Servicio de integración ejecuta una consulta de búsqueda predeterminada cuando la primera fila entra en la transformación de búsqueda.

Si utiliza una búsqueda en una tabla relacional, podrá reemplazar la consulta de búsqueda. Puede utilizar el reemplazo para añadir una cláusula WHERE o transformar los datos de búsqueda antes de que se almacenen en la memoria caché.

Si configura un reemplazo de SQL y un filtro en la consulta de búsqueda, el Servicio de integración omite el filtro.

Consulta de búsqueda predeterminada

La consulta de búsqueda predeterminada contiene las siguientes instrucciones:

SELECT

La instrucción SELECT incluye todos los puertos de búsqueda en la asignación. Para ver la instrucción SELECT para la consulta de búsqueda, seleccione la propiedad Usar consulta personalizada.

ORDER BY

La cláusula ORDER BY ordena las columnas en el mismo orden que aparecen en la transformación de búsqueda. El Servicio de integración genera la cláusula ORDER BY. No se puede ver esta al generar el SQL predeterminado.

Reemplazo de SQL para una consulta de búsqueda

Puede reemplazar la consulta de búsqueda para una búsqueda relacional. Puede añadir una cláusula WHERE y transformar los datos de búsqueda antes de que se almacenen en la memoria caché.

Puede utilizar palabras reservadas y barras diagonales en los nombres de tablas y de columnas.

Puede introducir una consulta para reemplazar la consulta de búsqueda predeterminada por completo. O bien, puede ver y editar la consulta de búsqueda predeterminada. La consulta de búsqueda predeterminada incluye los puertos de búsqueda, los puertos de salida y el puerto de devolución.

Cuando se utiliza un reemplazo SQL, la consulta anexa la cláusula ORDER BY 1. La cláusula ordena los datos para que proporcionen de forma fiable los valores primero y último de otras cláusulas.

Nota: Puede validar el SQL manualmente ejecutando la siguiente consulta en una utilidad de línea de comandos de Hive:

```
CREATE VIEW <table name> (<port list>) AS <SQL>
```

Donde:

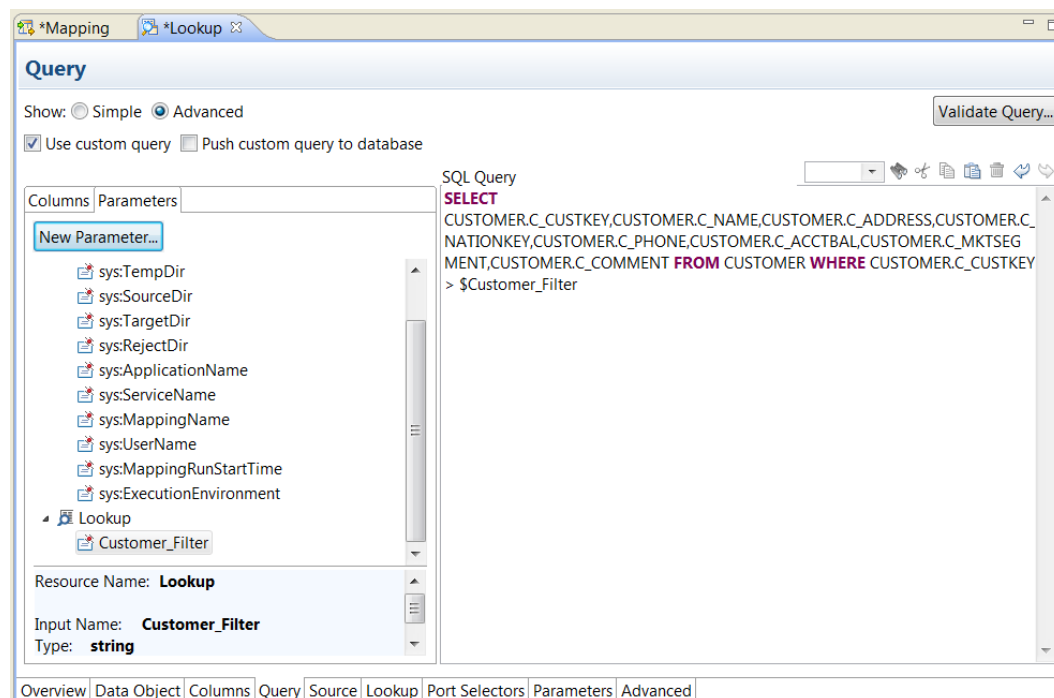
- <table name> es el nombre que elija.
- <port list> es la lista delimitada por comas de puertos en el origen.
- <SQL> es la consulta que se debe validar.

Parámetros de una consulta de reemplazo de SQL

Puede utilizar parámetros del sistema o parámetros definidos por el usuario en la consulta de búsqueda de una transformación de búsqueda. El editor de SQL proporciona una lista de los parámetros del sistema y los parámetros definidos por el usuario que puede seleccionar.

Puede buscar o crear estos parámetros definidos por el usuario en la ficha **Consulta** de una transformación de búsqueda. Defina un valor predeterminado para cada parámetro. Puede anular un valor predeterminado si enlaza un mapplet o un parámetro de asignación al parámetro de la transformación después de añadir la transformación de búsqueda a una asignación.

La siguiente imagen muestra la ficha Consulta de una transformación de búsqueda:



Palabras reservadas

Si ningún nombre de búsqueda o nombre de columna contiene una palabra reservada de base de datos, tales como MONTH o YEAR, la asignación genera errores de base de datos cuando el Servicio de integración ejecuta SQL en la base de datos.

Puede crear y mantener un archivo de palabras reservadas, reswords.txt, en el directorio de instalación del Servicio de integración. Cuando el Servicio de integración inicializa una asignación, busca el archivo reswords.txt y encierra entre comillas las palabras reservadas y, a continuación, ejecuta el SQL en las bases de datos de origen, destino y búsqueda.

Quizás necesite habilitar algunas bases de datos, como Microsoft SQL Server y Sybase, para utilizar los estándares de SQL-92 relacionados con identificadores entrecomillados. Utilice SQL de entorno para ejecutar el comando. Por ejemplo, con Microsoft SQL Server, utilice el comando siguiente:

```
SET QUOTED_IDENTIFIER ON
```

Directrices para reemplazar la consulta de búsqueda

Hay ciertas reglas y directrices que se aplican cuando se reemplaza una consulta de búsqueda.

Tenga en cuenta las siguientes directrices cuando se reemplaza la consulta SQL de búsqueda:

- Puede reemplazar la consulta SQL de búsqueda para búsquedas relacionales.
- Añade un filtro de búsqueda en el origen para filtrar las filas que se han añadido a la memoria caché de búsqueda. Esto garantiza que el servicio de integración inserte filas en la memoria caché dinámica y en la tabla de destino que coincidan con la cláusula WHERE.
- Si varias transformaciones de búsqueda comparten una memoria caché de búsqueda, utilice el mismo reemplazo de SQL de búsqueda para cada transformación de búsqueda.
- Si en la consulta de búsqueda un nombre de tabla o de columna contiene una palabra reservada, escriba la palabra reservada entre comillas.
- Para reemplazar la consulta de búsqueda para una búsqueda no almacenada en memoria caché, elija la opción de devolver cualquier valor cuando el servicio de integración encuentre varias coincidencias.
- No puede añadir ni eliminar ninguna de las columnas de la instrucción SQL predeterminada.
- Developer tool no valida la sintaxis de la consulta SQL. Si el reemplazo de SQL en una consulta de búsqueda sin conectar no es válida, se produce un error en la asignación.

Reemplazar la consulta de búsqueda

Puede reemplazar la consulta SQL de búsqueda predeterminada para crear una consulta personalizada en el origen de búsqueda.

1. En la vista **Propiedades**, seleccione la ficha **Consulta**.
2. Seleccione **Avanzadas**.
3. Seleccione **Usar consulta personalizada**.
4. Edite la consulta de búsqueda en el área Consulta SQL.
Puede hacer doble clic en un nombre de tabla, un nombre de columna o un parámetro para agregarlo a la consulta.
5. Haga clic en **Validar consulta** para validar la consulta de búsqueda.
6. Seleccione **Insertar una consulta personalizada en la base de datos** para ejecutar la consulta de búsqueda en la base de datos.

Filtro de origen de búsqueda

Puede configurar un filtro de origen de búsqueda para una transformación de búsqueda relacional que tiene la memoria caché habilitada. Añada el filtro de origen de búsqueda para limitar el número de búsquedas que el servicio de integración realiza en una tabla de origen de búsqueda.

Cuando configure un filtro de origen de búsqueda, el servicio de integración realiza búsquedas basadas en los resultados de la instrucción de filtro. Por ejemplo, quizás necesite para recuperar el apellido de todos los empleados cuyo ID es mayor que 510.

Configure el siguiente filtro de origen de búsqueda en la columna EmployeeID:

```
EmployeeID >= 510
```

EmployeeID es un puerto de entrada en la transformación de búsqueda. Cuando el servicio de integración lee la fila de origen, realiza una búsqueda en la memoria caché cuando el valor de EmployeeID es mayor que 510. Cuando EmployeeID es menor o igual a 510, la transformación de búsqueda no recupera el apellido.

Cuando añade un filtro de origen de búsqueda a la consulta de búsqueda para una asignación configurada para la optimización de inserciones, el servicio de integración crea una vista para representar el reemplazo de SQL. El servicio de integración ejecuta una consulta SQL en esta vista para insertar la lógica de transformación en la base de datos.

Filtrar filas de origen en una búsqueda

Puede configurar un filtro de origen de búsqueda para una transformación de búsqueda relacional que tiene la memoria caché habilitada. Filtrar filas de origen en una búsqueda para limitar el número de búsquedas que el servicio de integración realiza en una tabla de origen de búsqueda.

1. En la vista **Propiedades**, seleccione la ficha **Consulta**.
2. En la opción **Filtro**, haga clic en **Editar**.
3. En el editor de SQL, seleccione los puertos de entrada o especifique cualquier puerto de transformación de búsqueda que desea filtrar.
4. Especifique una condición de filtro.

No incluya la palabra clave WHERE en la condición de filtro. Escriba los parámetros y variables de la asignación de cadena en identificadores de cadena.

5. Haga clic en **Validar consulta** para validar la sintaxis de la condición de filtro.

Condición de búsqueda

El Servicio de integración de datos busca datos en el origen de búsqueda según una condición de búsqueda. Cuando se configura una condición de búsqueda en una transformación de búsqueda se compara el valor de una o más columnas en los datos de origen con valores en el origen o la memoria caché de búsqueda.

Por ejemplo, los datos de origen contienen un employee_number. La tabla de origen de búsqueda contiene employee_ID, first_name y last_name. Configure la siguiente condición de búsqueda:

```
employee_ID = employee_number
```

Para cada employee_number, el Servicio de integración de datos devuelve las columnas employee_ID, last_name y first_name del origen de búsqueda.

El Servicio de integración de datos puede devolver más de una fila desde el origen de búsqueda. Configure la siguiente condición de búsqueda:

```
employee_ID > employee_number
```

El Servicio de integración de datos devuelve filas para todos los números de employee_ID mayores que el número de empleado de origen.

Valores nulos en un objeto de datos de búsqueda

Cuando una entrada en la condición de búsqueda es nula, una transformación de búsqueda de objetos de datos devolverá una única fila con valores nulos para los puertos de solo salida y valores de las filas de entrada para los puertos de transferencia.

Por ejemplo, la siguiente condición de búsqueda realiza una búsqueda en un origen de datos que contiene una o más filas donde el valor de employee_ID es nulo:

```
employee_ID = employee_number
```

En este ejemplo, se usa una tabla de búsqueda con los siguientes datos:

EMPLOYEE_ID	LAST_NAME
1294765	Hara
1356356	Carver
1407207	NULL
1570348	Draper
NULL	Limonov

Puede comparar los siguientes valores de entrada del origen de datos con la tabla de búsqueda:

```
EMPLOYEE_NUMBER  
-----  
1294765  
1356356  
1407207  
1648246  
NULL
```

En este ejemplo, la condición de búsqueda genera los siguientes resultados:

```
1294765,Hara  
1356356,Carver  
1407207,NULL  
NULL,NULL  
NULL,NULL
```

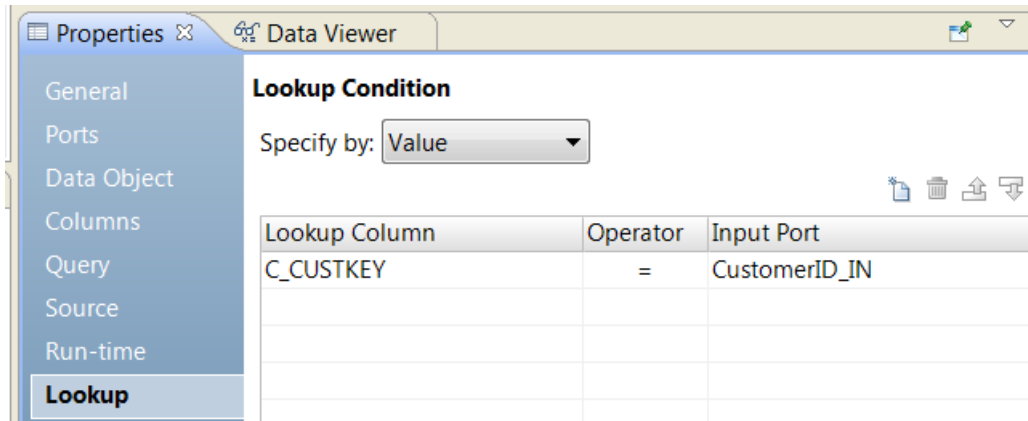
La condición de búsqueda halla coincidencias entre EMPLOYEE_ID y EMPLOYEE_NUMBER en las dos primeras filas. En la tercera fila, el origen de la búsqueda contiene una fila con un valor nulo que no participa en la condición de búsqueda. Coincide con la condición de búsqueda y devuelve un resultado con el valor nulo para la columna que no es de búsqueda.

En cuanto a las filas cuarta y quinta, la condición de búsqueda no halla ninguna coincidencia y devuelve un valor nulo para ambos valores. Con respecto a la quinta fila, observe que la condición de búsqueda no vuelve a encontrar una coincidencia, porque un valor nulo no coincide con ningún elemento, ni con uno nulo.

Configurar la condición de búsqueda

La condición de búsqueda es una expresión que determina qué filas se recuperan del origen de búsqueda. Configure la condición de búsqueda en la ficha **Búsqueda** de la vista **Propiedades**.

La siguiente imagen muestra una condición de búsqueda para realizar una búsqueda por número de cliente:



Puede configurar las siguientes opciones en la ficha **Búsqueda**:

Especificar por

Elija **Valor** para seleccionar los nombres de columna de búsqueda y de puerto de entrada. Seleccione **Parámetro** para configurar un parámetro de expresión que defina la condición de búsqueda.

Columna de búsqueda

Columna en el origen de búsqueda que debe cotejarse con una columna de la fila de entrada. Puede incluir varias columnas en la condición de búsqueda.

Operador

Operador que determina la condición de búsqueda entre la columna de búsqueda y el puerto de entrada. Los operadores pueden ser =, !=, >, <, >=, <=.

Puerto de entrada

Puerto de entrada que contiene el valor de búsqueda en el origen de búsqueda. Se puede comparar más de un puerto de entrada con los puertos en el origen de búsqueda.

Reglas y directrices para las condiciones de una transformación de búsqueda

Hay ciertas reglas y directrices que se aplican cuando se especifica una condición para una transformación de búsqueda.

Tenga en cuenta las siguientes reglas y directrices cuando especifique una condición para una transformación de búsqueda.

- Los tipos de datos para las columnas de una condición de búsqueda deben coincidir.
- Utilice un puerto de entrada por cada puerto de búsqueda en la condición de búsqueda. Puede utilizar el mismo puerto de entrada en más de una condición de una transformación.
- Si utiliza un selector de puerto o un puerto dinámico en una condición de búsqueda, la condición de búsqueda tendrá en cuenta todos los puertos en la expresión.

- Puede usar un puerto dinámico o un selector de puerto como puerto de entrada en una condición de búsqueda. El número de puertos generados en el puerto de entrada debe ser el mismo que el número de puertos en la columna de búsqueda.
- Cuando procese una transformación de búsqueda con varias condiciones de búsqueda, el Servicio de integración devuelve filas que coinciden con todas las condiciones de búsqueda.
- Puede crear un parámetro de expresión para parametrizar la condición de búsqueda en una transformación de búsqueda no reutilizable.
- Para aumentar el rendimiento de la búsqueda, especifique condiciones en el siguiente orden:
 - Igual a (=)
 - Menor que (<), mayor que (>), menor o igual que (<=), mayor o igual que o (>=)
 - No es igual que (!=)
- Utilice uno de los siguientes operadores cuando cree una condición de búsqueda: =, >, <, >=, <=, o !=, >, <, >=, <= o !=, =.
- El Servicio de integración procesa las coincidencias de la búsqueda de forma diferente en función de si se configura la transformación de búsqueda con una memoria caché de búsqueda dinámica, una memoria caché de búsqueda estática o una búsqueda sin almacenamiento en memoria caché.
- El Servicio de integración de datos busca coincidencias en valores nulos. Por ejemplo, el Servicio de integración considera un puerto de búsqueda y un puerto de entrada iguales si ambos tienen valores nulos.
- Si las columnas de una condición de búsqueda son de tipo Decimal, la precisión de cada una de ellas debe pertenecer al mismo rango de precisión. Los rangos de precisión válidos son:
 - 0-18 decimal
 - 19-28 decimal
 - Decimal 29 a 38
 - Decimal 39 y superior

Por ejemplo, si define la condición `DecimalA = DecimalB` donde `DecimalA` tiene una precisión de 15 y `DecimalB` de 25, la condición de búsqueda no es válida.

Memoria caché de búsqueda

Puede aumentar el rendimiento almacenando en la memoria caché un origen de búsqueda grande o tablas de búsqueda pequeñas. Cuando se almacena en la memoria caché el origen de búsqueda, el Servicio de integración consulta la memoria caché de búsqueda en lugar de consultar el origen de búsqueda de cada fila de entrada.

En función de sus requisitos empresariales, puede crear diferentes tipos de memorias caché de búsqueda. Puede crear una memoria caché estática o dinámica. Puede crear una memoria caché persistente o no persistente. Puede compartir una memoria caché entre varias transformaciones de búsqueda.

Si la transformación de búsqueda está en una asignación dinámica, puede tener una memoria caché persistente o no persistente. Cuando la memoria caché es persistente y cambia el origen de la búsqueda por un parámetro, la asignación fallará. La asignación también fallará si cambia el archivo de control de un origen de búsqueda de archivo sin formato.

Nota: No se puede utilizar una memoria caché de búsqueda dinámica o persistente cuando la transformación de búsqueda contiene un puerto dinámico o un origen de búsqueda parametrizado.

Propiedades de consulta

Configure las propiedades de consulta para ver o cambiar la consulta de búsqueda en una tabla de búsqueda de datos relacionales. Puede aplicar un filtro en la búsqueda o personalizar la consulta de búsqueda.

La siguiente tabla describe las propiedades de consulta para las transformaciones de búsqueda que realizan búsquedas de datos relacionales:

Propiedad	Descripción
Simple	Selecciónelo para ver la consulta de búsqueda y aplicar un filtro en la búsqueda.
Avanzadas	Selecciónelo para ver la consulta, personalizar la consulta o ejecutar la consulta en la base de datos que contiene la tabla de búsqueda de datos relacionales.
Filtro	Especifique un filtro para reducir el número de filas consultadas por el servicio de integración. Deberá seleccionar la opción Simple para ver esta opción.
Usar consulta personalizada	Selecciónelo para reemplazar la consulta de búsqueda. Deberá seleccionar la opción Avanzadas para ver esta opción.
Insertar una consulta personalizada en la base de datos	Selecciónelo para ejecutar la consulta en la base de datos que contiene la tabla de búsqueda de datos relacionales. Deberá seleccionar la opción Avanzadas para ver esta opción.
Consulta SQL	Muestra la consulta SQL utilizada para realizar la búsqueda. Puede personalizar la consulta SQL. Deberá seleccionar la opción Avanzadas para ver esta opción.

Transformaciones de búsqueda en asignaciones dinámicas

Puede utilizar una transformación de búsqueda en una asignación dinámica. Puede configurar puertos dinámicos para recibir y devolver diferentes puertos en función de los datos de origen. Puede parametrizar el origen y la condición de búsqueda para llevar a cabo una búsqueda basada en los distintos puertos.

Una asignación dinámica es una asignación en la que se pueden cambiar orígenes, destinos y lógicas de transformación en tiempo de ejecución. Puede establecer los parámetros y reglas con los que cambiar la estructura de los datos. Al utilizar una transformación de búsqueda en una asignación dinámica, los puertos de entrada de esa transformación pueden cambiar en función de los datos de origen. La estructura del origen de búsqueda y los puertos en la condición búsqueda pueden cambiar.

Nota: Cuando la transformación de búsqueda contiene un puerto dinámico o un origen de búsqueda parametrizado, la memoria caché de búsqueda no se puede conservar. Una memoria caché dinámica tampoco se puede configurar.

Puede realizar las siguientes tareas relativas a una transformación de búsqueda para utilizar la transformación en una asignación dinámica:

Definir puertos dinámicos

Defina puertos dinámicos y puertos generados para dar cabida a los cambios que se produzcan en las columnas de entrada.

Parametrizar el origen de la búsqueda

Asigne un parámetro para el objeto de datos que defina el origen de la búsqueda. Puede parametrizar el origen de búsqueda en una transformación de búsqueda no reutilizable.

Definir selectores de puerto

Defina un selector de puerto que especifique los puertos que se van a utilizar en la condición de búsqueda. Puede parametrizar los puertos del selector de puerto en una transformación de búsqueda no reutilizable.

Parametrizar la condición de búsqueda

Cree un parámetro de expresión y defina un valor predeterminado que contenga una expresión completa.

Para obtener más información sobre las asignaciones dinámicas, consulte la *Guía de asignación de Informática Developer*.

Definir puertos dinámicos

Puede definir puertos dinámicos en la transformación de búsqueda.

Puede hacer referencia a un puerto dinámico en la columna de entrada de una condición de búsqueda. Si el puerto dinámico contiene varios puertos generados, puede utilizar un selector de puerto para el elemento de columna de búsqueda de la condición de búsqueda. El puerto de entrada dinámico debe contener el mismo número de puertos que el selector de puerto de la condición de búsqueda.

Si el puerto dinámico contiene un valor, puede utilizar un único puerto en el elemento de columna de búsqueda de la condición de búsqueda.

Puede hacer referencia a puertos generados en la condición de búsqueda. Sin embargo, si cambia un origen de la asignación dinámica, el puerto generado podría no existir. La asignación fallará.

Cambiar el origen de búsqueda

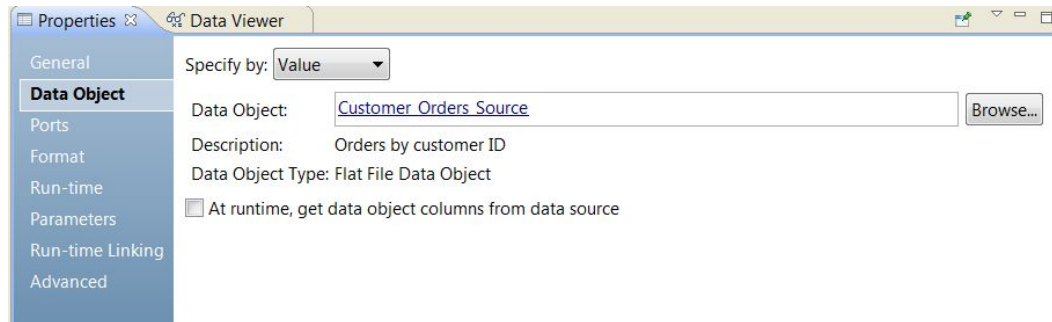
Puede cambiar el objeto de datos que constituye el origen de búsqueda de una transformación de búsqueda reutilizable. Puede configurar un parámetro que determine el objeto de datos que se debe utilizar como origen de búsqueda en tiempo de ejecución.

Al crear una transformación a partir de un objeto de datos físicos, la información sobre el objeto de datos aparece recogida en la ficha **Objeto de datos** de las propiedades de la transformación. Puede hacer clic en el nombre del objeto de datos para ver la definición del objeto de datos físicos del repositorio de modelos.

Si desea cambiar el objeto de datos de la transformación, busque otro objeto de datos físicos en el repositorio de modelos. Cuando se cambia el objeto de datos, la transformación utiliza las propiedades de tiempo de ejecución y propiedades avanzadas del objeto de datos que seleccione.

La estructura del objeto de datos se puede actualizar en tiempo de ejecución según los cambios en el origen de datos. El origen de datos es el archivo físico o la tabla de base de datos que el objeto de datos representa. Cuando se habilita el servicio de integración de datos para obtener columnas de datos del origen de datos, dicho servicio examina la estructura del origen de datos. Luego, actualizará los puertos de objeto de datos en la instancia de transformación según el origen de datos. El servicio de integración de datos no cambia la definición del objeto de datos físicos del repositorio de modelos.

En la siguiente imagen se muestra la ficha **Objeto de datos**:



La ficha **Objeto de datos** tiene los siguientes campos:

Especificar por

Elija **Valor** para introducir un nombre de objeto de datos específico. Elija **Parámetro** para parametrizar el objeto de datos.

Objeto de datos

Nombre del objeto de datos en el repositorio de modelos. Puede hacer clic en el vínculo **Objeto de datos** para abrir la definición del objeto de datos del repositorio de modelos. También puede hacer clic en **Examinar** para buscar otro objeto de datos en el repositorio de modelos.

Descripción

Descripción del objeto de datos en el repositorio. Solo lectura.

Tipo de objeto de datos

Describe el tipo del objeto de datos, como archivo sin formato, relacional o personalizado.

En tiempo de ejecución, obtener las columnas de objetos de datos del origen de datos

El servicio de integración de datos obtiene los cambios de definición de datos y metadatos del archivo de datos o de la tabla a la que hace referencia el objeto de datos, y actualiza la estructura del objeto de datos de la instancia de transformación en tiempo de ejecución.

Para obtener una vista previa del modo en que el servicio de integración de datos obtiene los cambios de definición de datos y metadatos en tiempo de ejecución, vea la asignación con los parámetros resueltos.

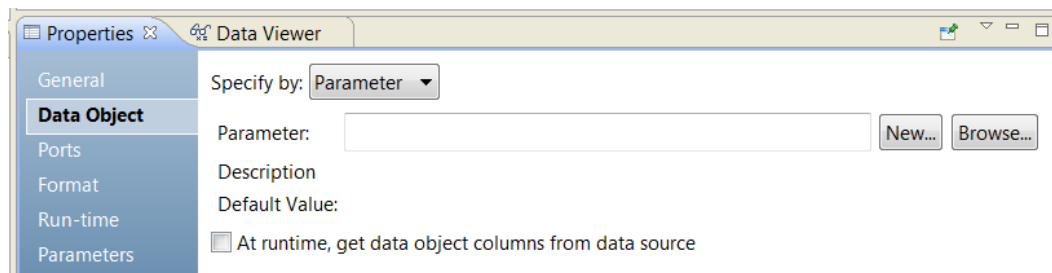
Parametrizar el origen de la búsqueda

Puede configurar un parámetro del origen de búsqueda en una transformación de búsqueda no reutilizable.

Para parametrizar un objeto de datos, elija **Especificar por parámetro** en la ficha **Objeto de datos**. Las propiedades en la ficha **Objeto de datos** cambian.

Para parametrizar el objeto de datos, cree un parámetro de tipo de recurso o busque un parámetro de recurso que ya haya creado. El valor predeterminado de parámetro es el nombre del objeto de datos físicos en el repositorio de modelos. Al crear un valor de parámetro predeterminado, hay que seleccionar un nombre de objeto de datos físicos de una lista de objetos de datos en el repositorio.

La siguiente imagen muestra la ficha **Objeto de datos** cuando el objeto de datos se especifica mediante un parámetro:



La ficha **Objeto de datos** tiene las siguientes opciones por parámetro:

Parámetro

Nombre de un parámetro de recurso que ha configurado como el objeto de datos. Solo lectura.

Descripción

Descripción del parámetro. Solo lectura.

Nuevo

Cree un parámetro de recurso. Busque y seleccione un objeto de datos en el repositorio de modelos como el valor predeterminado de parámetro.

Examinar

Busque un parámetro de recurso y seleccione el parámetro.

Valor predeterminado

Valor predeterminado del parámetro de recurso configurado para el objeto de datos. El valor predeterminado es un nombre de objeto de datos físicos y la ruta a ese objeto en el repositorio de modelos. Solo lectura.

Conflictos de nombres de puerto con el puerto de búsqueda

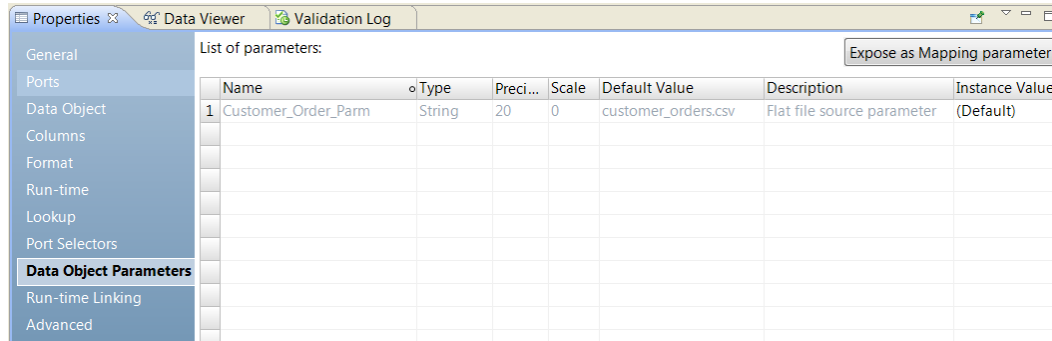
Cuando se parametriza el origen de una búsqueda, un puerto de entrada de la transformación de búsqueda podría tener un conflicto de nombre con un puerto en el origen de la consulta.

Cuando un puerto de entrada de la transformación de búsqueda tiene un conflicto de nombre con el puerto de búsqueda en el origen de la búsqueda, Developer tool no cambia el nombre de ninguno de los puertos. Developer tool muestra un error de validación. Debe cambiar el nombre del puerto de entrada en la transformación de búsqueda o quitar el puerto de la transformación.

Orígenes de búsqueda que contienen parámetros

Puede crear un origen de búsqueda a partir de un objeto de datos físicos que contiene parámetros. Al añadir el objeto de datos físicos a una asignación, los parámetros aparecen en la ficha **Parámetros de objeto de datos**.

En la siguiente imagen se muestra la ficha **Parámetros de objeto de datos** de la transformación de búsqueda:



La imagen muestra Customer_Order_Parm. Customer_Order_Parm es un parámetro del nombre de archivo de origen en un objeto de datos de archivo sin formato. Para reemplazar el nombre de archivo de origen en la asignación, enlace Customer_Order_Parm a un parámetro en la asignación. Haga clic en **Exponer como parámetro de asignación** para crear un parámetro duplicado en la asignación.

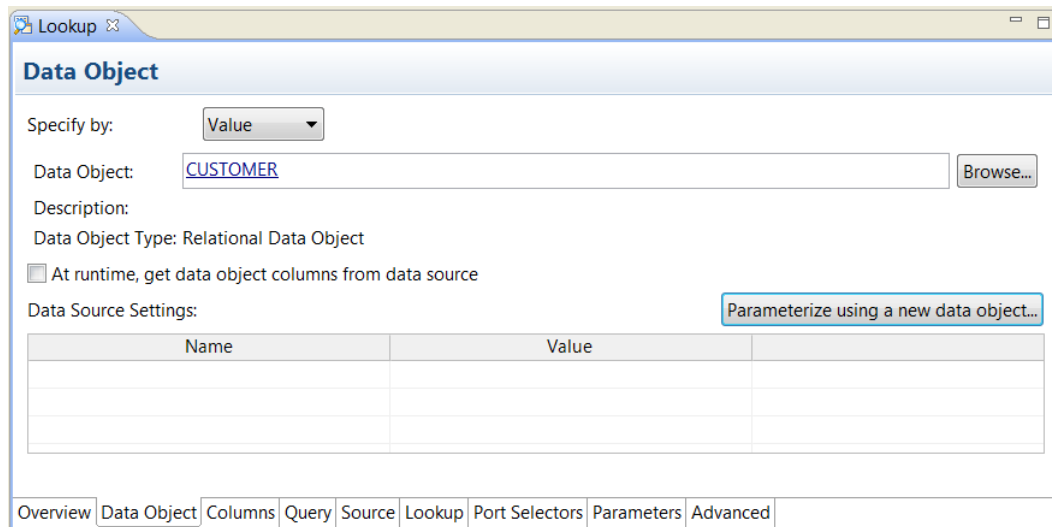
Configurar parámetros en un objeto de datos duplicado

Puede crear un objeto de datos duplicado en el repositorio y parametrizar las propiedades de ese objeto de datos físicos. Defina los valores predeterminados de propiedades como la conexión, el nombre del recurso, el propietario de la tabla o el nombre del archivo de control.

Puede crear un objeto de datos duplicado relativo a objetos de datos relacionales y a objetos de datos de archivo sin formato. También se puede crear un objeto de datos duplicado en transformaciones de búsqueda tanto reutilizables como no reutilizables.

Cree un objeto de datos duplicado en la ficha **Objeto de datos** de la transformación de búsqueda. Puede crearlo si especifica el objeto de datos como un valor. Al crear el objeto de datos duplicado, hay que reemplazar el nombre del objeto de datos en la transformación de búsqueda por el nombre del objeto de datos duplicado. Developer tool crea parámetros correspondientes a las propiedades del objeto de datos. Esta herramienta le pedirá que introduzca valores predeterminados para los parámetros. La sintaxis del nombre de un objeto de datos duplicado es la siguiente: <nombre de objeto original>_Param.

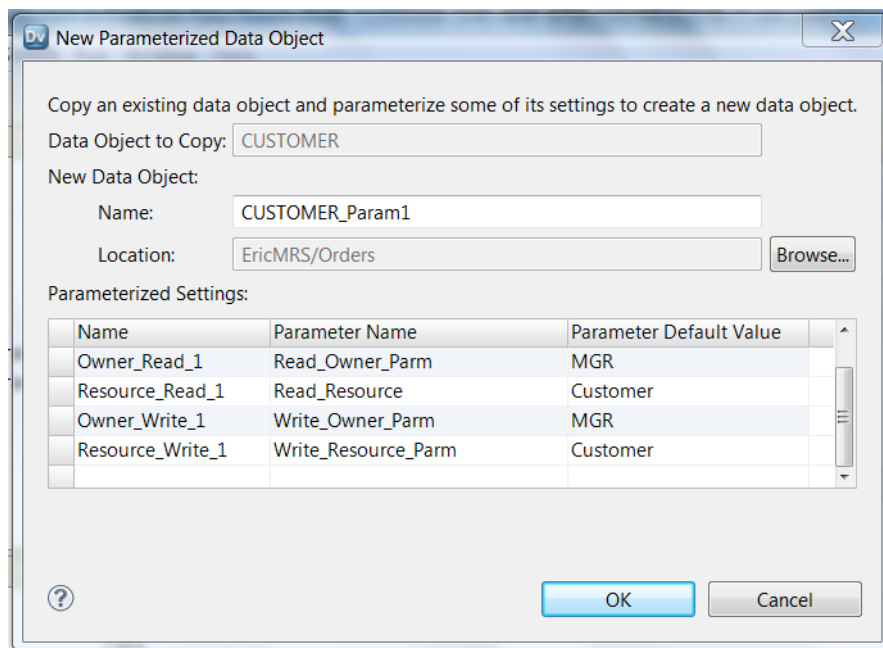
En la siguiente imagen se muestra el botón **Parametrice mediante un nuevo objeto de datos** de la ficha **Objeto de datos** para objetos de datos relacionales y de archivo sin formato:



Cuando se crea un objeto de datos duplicado para parametrizar un objeto de datos, Developer tool crea el conjunto de parámetros de ese objeto de datos. Developer tool crea parámetros diferentes dependiendo de si el objeto de datos es un objeto relacional o de archivo sin formato.

Al crear el objeto de datos duplicado, configure los valores de parámetro predeterminados en el cuadro de diálogo **Nuevo objeto de datos parametrizado**.

En la siguiente imagen se muestra el cuadro de diálogo **Nuevo objeto de datos parametrizado** de un objeto de datos relacionales:

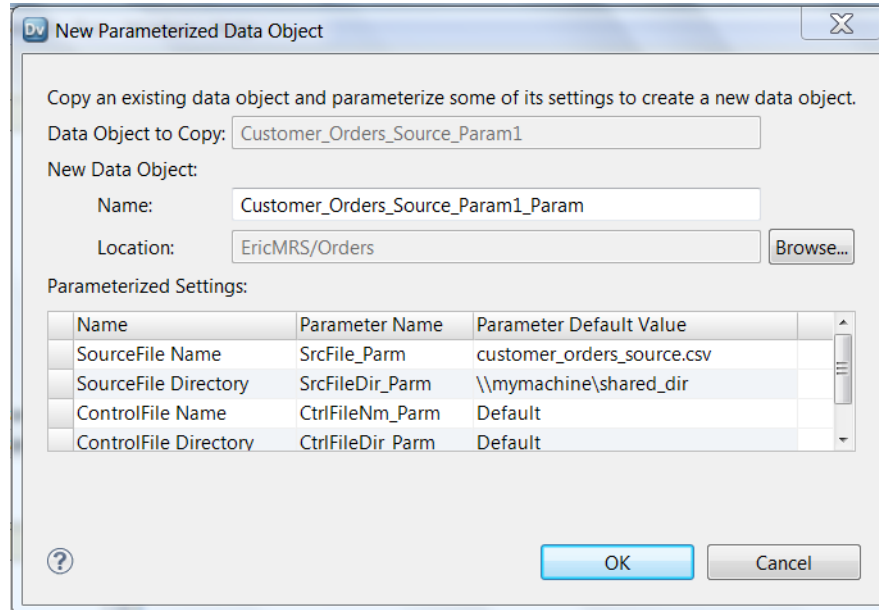


El nombre del objeto de datos se puede modificar. Introduzca los valores predeterminados de los parámetros de propietario y recurso.

Si el objeto de datos original está parametrizado, Developer tool copiará los parámetros de ese objeto de datos en el objeto de datos duplicado. Si una propiedad original no está parametrizada, Developer tool creará un parámetro para ella en el objeto de datos duplicado. Developer tool utiliza el valor de propiedad original como valor de parámetro predeterminado en el objeto de datos duplicado. En caso de que Developer tool no

pueda averiguar un valor de propiedad original, creará un parámetro con un valor predeterminado según el tipo de parámetro.

En la siguiente imagen se muestra el cuadro de diálogo **Nuevo objeto de datos parametrizado** de un objeto de datos de archivo sin formato:



Configure los valores de parámetro predeterminados del archivo de origen y del directorio del archivo de origen. Si el archivo sin formato tiene un archivo de control, configure el nombre y el directorio del archivo de control.

Después de configurar los valores predeterminados, Developer tool crea el objeto de datos duplicado. El nombre del objeto de datos duplicado aparece en la ficha **Objeto de datos** de la transformación de búsqueda. El objeto de datos duplicado aparece en el navegador de objetos.

Para cambiar los valores de parámetro del objeto de datos después de haberlo creado, abra el objeto de datos físicos en el navegador de objetos. Haga clic en la ficha **Parámetros**.

Selectores de puerto

Puede crear una condición de búsqueda cuando la transformación de búsqueda contiene puertos generados. Puede hacer referencia a un puerto dinámico o a un selector de puerto en una condición de búsqueda. También puede utilizar un parámetro de expresión para parametrizar la expresión de búsqueda completa.

Si el puerto dinámico contiene varios puertos generados, puede definir un selector de puerto que filtre esos puertos generados en la condición de búsqueda. El origen de búsqueda puede cambiar en una asignación dinámica. Puede configurar un selector de puerto que filtre los puertos que se van a usar en la columna de búsqueda. El selector de puerto del origen de búsqueda debe contener el mismo número de puertos que el selector de puerto de las columnas de entrada.

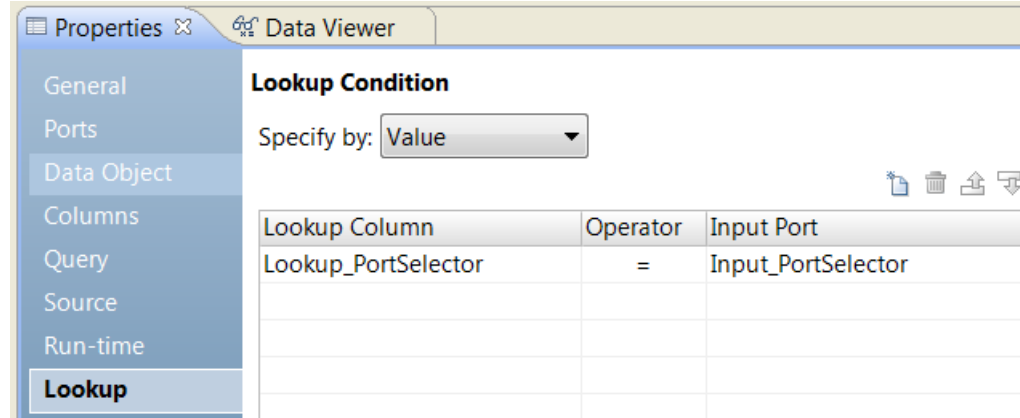
Por ejemplo, Lookup_PortSelector contiene los siguientes puertos:

```
C_CustKey  
C_OrderKey
```


Input_PortSelector contiene los siguientes puertos de entrada:

```
CustomerID_IN  
OrderID_IN
```

La siguiente imagen muestra una condición de búsqueda que contiene selectores de puerto:



La condición de búsqueda se expande a la siguiente expresión:

```
C_CustKey = CustomerID_IN AND C_OrderKey = OrderID_IN
```

Cuando la condición de búsqueda contiene varios puertos, puede configurar un operador. Así, puede cambiar el operador a mayor que (>). La condición de búsqueda se expande a la siguiente expresión:

```
C_CustKey > CustomerID_IN AND C_OrderKey > OrderID_IN
```

Puede crear una condición de búsqueda que contenga un puerto dinámico:

```
Lookup_PortSelector = Dynamic_Input_Port
```

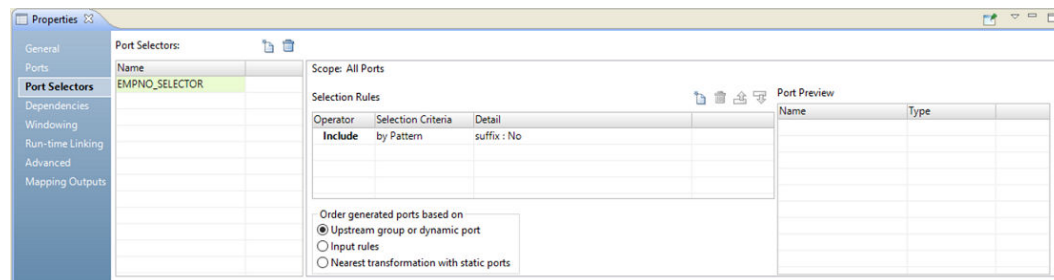
El puerto dinámico debe contener el mismo número de puertos que el selector de puerto.

Configuración del selector de puerto

Al configurar un selector de puerto, puede definir reglas de selección para saber qué puertos generados incluir. Las reglas de selección son similares a las reglas de entrada que se configuran para los puertos dinámicos.

Un selector de puerto puede incluir puertos estáticos o puertos generados. Configure un selector de puerto en la ficha **Selector de puerto**.

La siguiente imagen muestra la ficha **Selector de puerto**:



Configure las siguientes propiedades de un selector de puerto:

Nombre

Identifica el selector de puerto. Puede crear varios selectores de puerto en una transformación y hacer referencia a ellos en las expresiones.

Ámbito

Identifica un grupo de puertos al que el selector de puerto se aplica. Al crear un selector de puerto para una transformación de búsqueda o de unión, debe elegir el ámbito. Estas transformaciones tienen varios grupos de entrada. La transformación de unión tiene un ámbito Principal o Detalle. La transformación de búsqueda, por su parte, tiene un ámbito Importar o Búsqueda. La transformación de expresión tiene un grupo de entrada. El ámbito es siempre Todos los puertos.

Reglas de selección

Determina los puertos que se van a incluir en el selector de puerto. Al crear la selección reglas, el panel **Vista previa del puerto** muestra los puertos que cumplen los requisitos de los puertos de entrada actuales. Estos puertos pueden cambiar. Configure las reglas de selección para dar cabida a puertos de diferentes orígenes.

Reglas de selección

Las reglas de selección asociadas con un selector de puerto determinan los puertos a incluir en el selector de puerto.

Al crear la selección reglas, el panel **Vista previa del puerto** muestra los puertos que cumplen los requisitos de los puertos de entrada actuales. Estos puertos pueden cambiar. Configure las reglas de selección para dar cabida a puertos de diferentes orígenes.

Cree reglas de selección basadas en los criterios siguientes:

Operador

Incluye o excluye los puertos que las reglas de selección devuelven. El valor predeterminado es Incluir. Para poder excluir puertos, antes debe incluirlos.

Criterios de selección

Tipo de regla de selección que desea crear. Puede crear una regla basada en el nombre de columna, el tipo de puerto, el patrón o la definición del tipo de datos complejos. Para incluir puertos según el nombre de columna, busque nombres específicos o un patrón de caracteres en el nombre.

Detalle

Valores que se van a aplicar a los criterios de selección. Si los criterios de selección están definidos según el nombre de columna, configure la cadena o el nombre por el que realizar la búsqueda. Si los criterios de selección están definidos según el tipo de puerto, seleccione los tipos de puerto que quiera incluir.

En la siguiente tabla se describen los criterios de selección y cómo especificar los detalles de los criterios:

Criterios de selección	Descripción	Detalle
Todo	Incluye todos los puertos.	No se necesitan detalles.
Nombre	Filtra los puertos por el nombre del puerto.	Seleccione los nombres de puerto de una lista de valores o utilice un parámetro de tipo puerto o lista de puertos.
Tipo	Filtra los puertos por el tipo de datos de cada puerto.	Seleccione los tipos de datos de una lista.

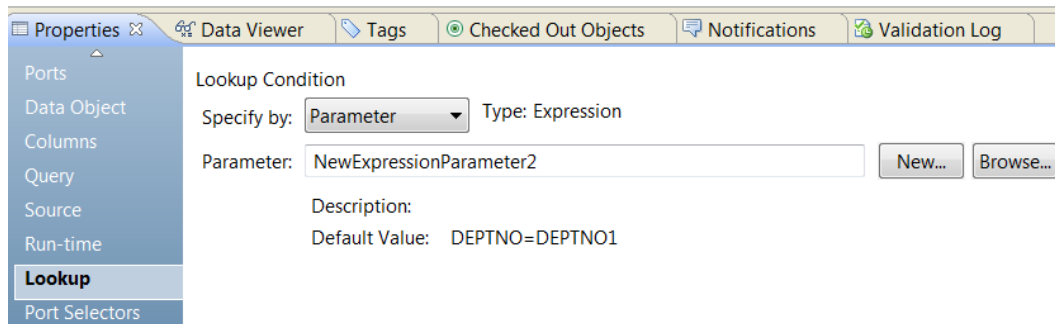
Criterios de selección	Descripción	Detalle
Patrón	Filtra los puertos por una cadena de caracteres en el nombre o por una expresión regular.	Elija Prefijo, Sufijo o Expresión regular como tipo de patrón para el nombre del puerto. Luego, introduzca un valor para el patrón o utilice un parámetro de tipo cadena.
Definición de tipo de datos complejo	Filtra los puertos a partir de una definición de tipo de datos complejo.	Elija un prefijo, un sufijo o una expresión regular como el tipo de patrón de la definición de tipo de datos complejo. Luego, introduzca un valor para el patrón o utilice un parámetro de tipo String.

Parametrizar la condición de búsqueda

Puede configurar un parámetro de expresión que defina la condición de búsqueda. Un parámetro de expresión contiene una expresión completa creada en un editor de expresiones. Puede definir un parámetro de asignación para anular el parámetro de expresión en tiempo de ejecución.

Si especifica una condición de búsqueda mediante un parámetro, puede buscar un parámetro de expresión o crear un parámetro.

La siguiente imagen muestra dónde se puede configurar un parámetro de expresión para la condición de búsqueda:



Para crear un parámetro, haga clic en **Nuevo**. Defina un nombre para el parámetro y modifique el valor predeterminado. El valor predeterminado del parámetro de expresión es la expresión completa para definir la condición de búsqueda. Puede utilizar puertos generados, puertos dinámicos y selectores de puerto en la expresión.

Nota: Al crear la expresión, la columna de búsqueda siempre es el primer valor y la columna de entrada, el segundo.

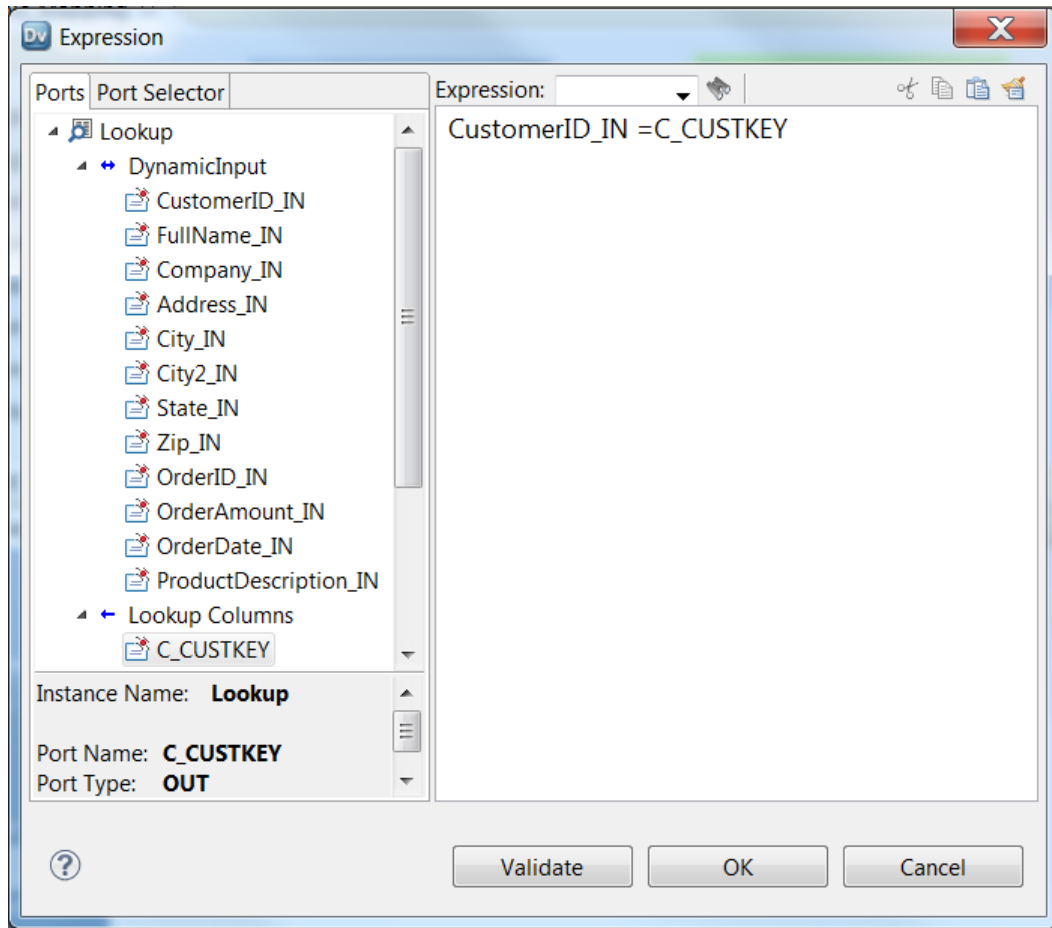
Por ejemplo, crea la siguiente condición de búsqueda en un parámetro de expresión:

```
CustomerID_IN = C_CUSTKEY
```

CustomerID_IN es la columna de búsqueda.

C_CUSTKEY es la columna de entrada.

La siguiente imagen muestra la expresión de búsqueda en el editor de expresiones:



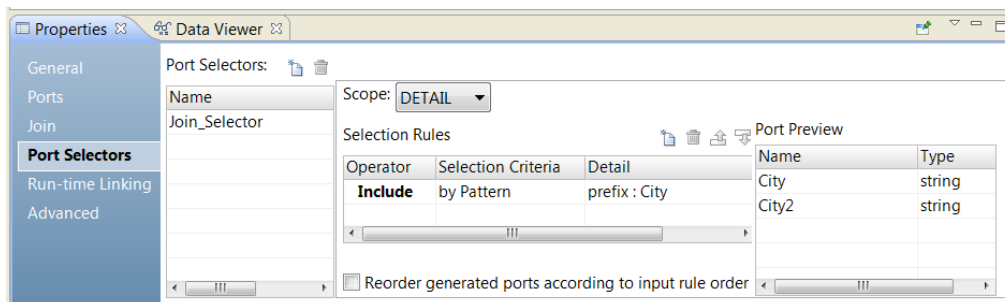
Crear un selector de puerto

Cree un selector de puerto para saber qué puertos hay que usar en una expresión dinámica, en una condición de búsqueda o en una condición de combinación.

1. Haga clic en la ficha **Selectores de puerto**.
2. En el área **Selectores de puerto**, haga clic en **Nuevo**.
Developer tool crea un selector de puerto con una regla de selección predeterminada que incluye todos los puertos.
3. En el área **Selectores de puerto**, cambie el nombre del selector de puerto por un nombre único.
4. Si está trabajando con una transformación de combinación o de búsqueda, seleccione el ámbito.
Los puertos disponibles varían según el grupo de puertos que elija.
5. En el área **Reglas de selección**, seleccione un **Operador**.
 - Incluir. Cree una regla que incluya puertos para el selector de puerto. Para poder excluir puertos, antes debe incluirlos.
 - Excluir. Cree una regla que excluya puertos específicos del selector de puerto.
6. Elija los **Criterios de selección**.
 - Por nombre. Seleccione puertos concretos por su nombre. Puede seleccionar los nombres de puerto de una lista de puertos en el ámbito.

- Por tipo. Seleccione puertos por el tipo. Puede seleccionar uno o varios tipos de datos.
- Por patrón. Seleccione puertos siguiendo un patrón de caracteres en el nombre del puerto. Puede realizar las búsquedas con caracteres específicos o crear una expresión regular.

La siguiente imagen muestra la ficha Selector de puerto:



7. Haga clic en la columna **Detalle**.

Se abre el cuadro de diálogo **Detalles de la regla de entrada**.

8. Seleccione los valores por los que quiera filtrar los puertos.

- Por nombre. Decida si quiere crear una lista de puertos por valor o por un parámetro. Haga clic en **Elegir** para seleccionar los puertos de la lista.
- Por tipo. Seleccione uno o varios tipos de datos de una lista. El área **Vista previa del puerto** muestra los puertos de los tipos que ha seleccionado.
- Por patrón. Decida si quiere hallar un determinado patrón de caracteres en el prefijo o sufijo del nombre del puerto. También puede optar por crear una expresión regular con la que realizar la búsqueda. Configure un parámetro o el patrón de búsqueda.

El área **Vista previa del puerto** muestra los puertos en el selector de puerto a medida que configura las reglas.

9. Para reordenar los puertos en el selector de puerto, seleccione **Reordenar los puertos generados según el orden de regla de entrada**.

Propiedades de tiempo de ejecución

Establezca las propiedades de tiempo de ejecución para habilitar y configurar la memoria caché de búsqueda. Deberá añadir la transformación de búsqueda a una asignación antes de poder configurar las propiedades de búsqueda en tiempo de ejecución.

En la siguiente tabla se describen las propiedades del tiempo de ejecución para las transformaciones de búsqueda que realizan búsquedas en archivos sin formato, tablas de referencia o datos relacionales:

Propiedad	Descripción
Almacenamiento en memoria caché de búsqueda habilitado	<p>Indica si el Servicio de integración almacena en memoria caché los valores de búsqueda. Cuando se habilita el almacenamiento de las búsquedas en la memoria caché, el Servicio de integración consulta una vez el origen, almacena los valores en la memoria caché y después consulta los valores en la memoria caché. El almacenamiento de valores de búsqueda en la memoria caché puede aumentar el rendimiento en tablas de búsqueda grandes.</p> <p>Cuando se deshabilita el almacenamiento en memoria caché, cada vez que pasa una fila a la transformación, el Servicio de integración emite una instrucción de selección al origen de la búsqueda para obtener los valores de búsqueda.</p> <p>El Servicio de integración siempre almacena en memoria caché las búsquedas de archivos sin formato.</p>
Tamaño de la memoria caché de datos de búsqueda	<p>Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de datos para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.</p>
Tamaño de la memoria caché de índice de búsqueda	<p>Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de índice para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.</p>
Prefijo de nombre de archivo de memoria caché	<p>El prefijo para el archivo de memoria caché. Puede especificar el prefijo del nombre de archivo de la memoria caché para una memoria caché de búsqueda persistente.</p>
Generar previamente la memoria caché de búsqueda	<p>Permite que el Servicio de integración genere la memoria caché de búsqueda antes de que la transformación de búsqueda reciba los datos. El Servicio de integración puede generar simultáneamente varios archivos de memoria caché de búsqueda para aumentar el rendimiento.</p> <p>Configure una de las siguientes opciones:</p> <ul style="list-style-type: none">- Auto. El Servicio de integración determina el valor.- Se permite siempre. Permite que el Servicio de integración genere la memoria caché de búsqueda antes de que la transformación de búsqueda reciba los datos. El Servicio de integración puede generar simultáneamente varios archivos de memoria caché de búsqueda para aumentar el rendimiento.- No se permite nunca. El Servicio de integración no puede generar la memoria caché de búsqueda antes de que la transformación de búsqueda reciba la primera fila.

Propiedad	Descripción
Nombre del directorio de la memoria caché de búsqueda	Directorio empleado para generar los archivos de memoria caché de búsqueda cuando se configura la transformación de búsqueda para almacenar en memoria caché el origen de búsqueda. El valor predeterminado es el parámetro del sistema CacheDir. Para esta propiedad, puede configurar otro parámetro del sistema o definido por el usuario.
Volver a almacenar en memoria caché desde el origen de búsqueda	Se vuelve a generar la memoria caché de búsqueda para sincronizar la memoria caché persistente con la tabla de búsqueda. Vuelva a almacenar en la memoria caché la búsqueda desde la base de datos cuando la transformación de búsqueda tenga una memoria caché de búsqueda persistente y la tabla de búsqueda cambie de forma ocasional.

TEMAS RELACIONADOS

- [“Tamaño de la memoria caché” en la página 74](#)

Propiedades avanzadas

Configure la memoria caché de búsqueda persistente y la conexión a una base de datos relacional en las propiedades avanzadas. Las propiedades que aparezcan están basadas en el tipo de origen de búsqueda.

La siguiente tabla describe las propiedades avanzadas para cada tipo de origen de búsqueda:

Propiedad	Tipo de origen de búsqueda	Descripción
Memoria caché de búsqueda persistente	Archivo sin formato, tabla de referencia, búsqueda relacional	Indica si el servicio de integración utiliza una memoria caché de búsqueda persistente, que consta de dos archivos de memoria caché como mínimo. Si una transformación de búsqueda está configurada para una memoria caché de búsqueda persistente y no existen archivos de memoria caché de búsqueda persistente, el servicio de integración creará esos archivos.
Comparación de cadenas con distinción de mayúsculas y minúsculas	Archivo sin formato	El servicio de integración utiliza comparaciones de cadenas que distinguen mayúsculas de minúsculas al realizar búsquedas en las columnas de cadenas.
Orden nulo	Archivo sin formato	Determina el modo en el que el servicio de integración ordena los valores nulos. Puede optar entre ordenar los valores nulos por arriba o por abajo. El servicio de integración ordena los valores nulos por arriba de forma predeterminada. Esto reemplaza la configuración del servicio de integración, según la cual los valores nulos en operadores de comparación se tratan como altos, bajos o nulos. En el caso de búsquedas relacionales, el orden nulo se basa en el valor predeterminado de la base de datos.

Propiedad	Tipo de origen de búsqueda	Descripción
Entrada ordenada	Archivo sin formato	Determina si los datos del origen ya están ordenados en los puertos clave y de datos de la transformación de búsqueda. Cuando se establece en TRUE, Informatica omite la clasificación de datos cuando crea una memoria caché para los resultados de búsqueda. El valor predeterminado es false.
Nivel de seguimiento	Archivo sin formato, objeto de datos lógicos, tabla de referencia, búsqueda relacional	Establece la cantidad de detalles que aparece en el registro de esta transformación. Puede elegir Simplificado, Normal, Inicialización detallada o Datos detallados. El valor predeterminado es Normal.
Directiva de búsqueda en caso de varias coincidencias	Archivo sin formato	Determina el resultado cuando la consulta de búsqueda encuentra varias coincidencias. Seleccione uno de los siguientes métodos: <ul style="list-style-type: none"> - Devolver la primera fila - Devolver la última fila - Informar sobre errores - Devolver cualquier fila - Devolver todas las filas El valor predeterminado es Devolver cualquier fila.
Omitir valor nulos que coincidan	Archivo sin formato	Omita los valores nulos que coincidan al realizar búsquedas.
Memoria caché de búsqueda dinámica	Archivo sin formato	Actualizar la memoria caché de búsqueda con valores de búsqueda no hallados o nuevos.
Actualizar o insertar	Tabla de referencia, búsqueda relacional	Se aplica únicamente a la memoria caché de búsqueda dinámica. El servicio de integración actualiza la fila en la memoria caché si el tipo de fila que entra en la transformación de búsqueda es UPDATE, la fila existe en la memoria caché de índice y los datos de la memoria caché son diferentes de la fila existente. El servicio de integración inserta la fila en la memoria caché si es nueva.
Insertar o actualizar	Tabla de referencia, búsqueda relacional	Se aplica únicamente a la memoria caché de búsqueda dinámica. El servicio de integración inserta la fila en la memoria caché si el tipo de fila que entra en la transformación de búsqueda es INSERT y es nuevo. Si la fila ya existe en la memoria caché de índice, pero la memoria caché de datos es distinta que la de la fila actual, el servicio de integración actualiza la fila en la memoria caché de datos.
Proporcionar como salida el valor anterior al actualizar	Tabla de referencia, búsqueda relacional	El servicio de integración proporciona como salida el valor que existía en la memoria caché antes de que la fila se actualizara. De lo contrario, proporcionará como salida el valor actualizado que escribe en la memoria caché.
Actualización de la condición de memoria caché dinámica	Tabla de referencia, búsqueda relacional	Se aplica únicamente a la memoria caché de búsqueda dinámica. Expresión que indica si desea actualizar la memoria caché dinámica. El servicio de integración actualiza la memoria caché cuando la condición es TRUE y los datos están en la memoria caché. El valor predeterminado es true.

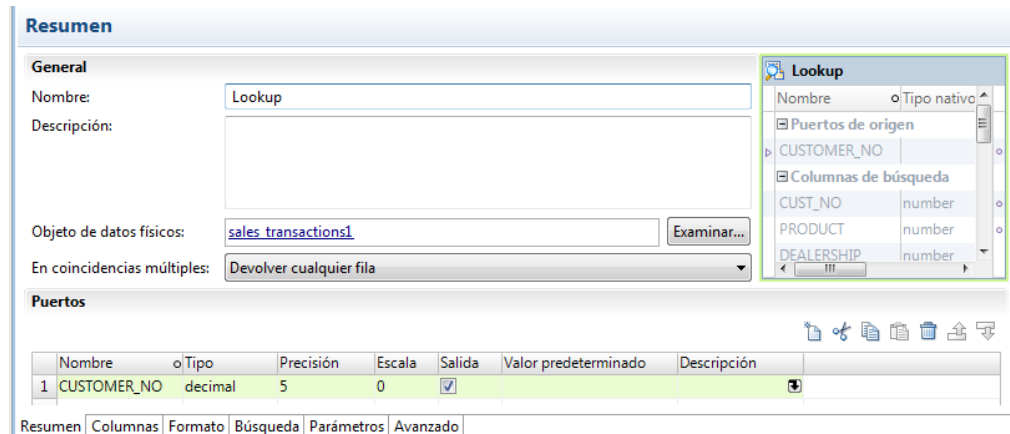
Propiedad	Tipo de origen de búsqueda	Descripción
Conexión	Tabla de referencia, búsqueda relacional	Conexión a la base de datos relacional que contiene el origen de búsqueda de datos relacionales. Puede utilizar un parámetro para la conexión. En las búsquedas en la tabla de referencia, este campo es de solo lectura.
Entrada ordenada	Archivo sin formato	Indica que los datos de entrada están ordenados previamente por grupos.
Formato de fecha y hora	Archivo sin formato	Defina un formato de fecha y hora y el ancho de campo. Los milisegundos, los microsegundos o los nanosegundos tienen un ancho de campo de 29. Si no se selecciona un formato de fecha y hora de un puerto, puede especificar cualquier formato de fecha y hora. El formato predeterminado es YYYY-MM-DD HH24:MI:SS. El formato de fecha y hora no cambia el tamaño del puerto. Este campo es de solo lectura.
Separador de millar	Archivo sin formato	El valor es Ninguno. Este campo es de solo lectura.
Separador decimal	Archivo sin formato	El valor es una coma. Este campo es de solo lectura.

Crear una transformación de búsqueda reutilizable

Cree una transformación de búsqueda para buscar datos en un archivo sin formato, objeto de datos lógicos, tabla de referencia u objeto de datos relacionales.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
3. Navegue hasta el asistente Búsqueda.
4. Seleccione **Búsqueda de objetos de datos de archivo sin formato**, **Búsqueda de objetos de datos lógicos**, **Búsqueda de tablas de referencia** o **Búsqueda de objetos de datos relacionales**.
5. Haga clic en **Siguiente**.
Aparece el cuadro de diálogo **Nueva transformación de búsqueda**.
6. Seleccione un objeto de datos físicos o una tabla de referencia en Developer Tool.
7. Especifique un nombre para la transformación.
8. En la lista desplegable **En coincidencias múltiples**, elija la estrategia que se usará cuando la transformación de búsqueda encuentre varias filas que coincidan con la condición de búsqueda.
9. Haga clic en **Finalizar**.
La transformación de búsqueda aparece en el editor.
10. En la sección **Puertos** de la vista **Resumen**, añada puertos de salida a la transformación.

La siguiente figura muestra el puerto de salida CUSTOMER_NO en una transformación de búsqueda:



11. En la ficha **Tiempo de ejecución** de la vista **Propiedades**, seleccione **Almacenamiento en memoria caché de búsqueda habilitado** para habilitar la memoria caché de búsqueda.
Nota: Deberá añadir la transformación de búsqueda a una asignación antes de poder configurar las propiedades de búsqueda en tiempo de ejecución.
12. En la ficha **Búsqueda** de la vista **Propiedades**, agregue una o más condiciones de búsqueda.
13. En la ficha **Avanzadas** de la vista **Propiedades**, configure el nivel de seguimiento, las propiedades de la memoria caché de búsqueda dinámica y la conexión en tiempo de ejecución.
14. Guarde la transformación.

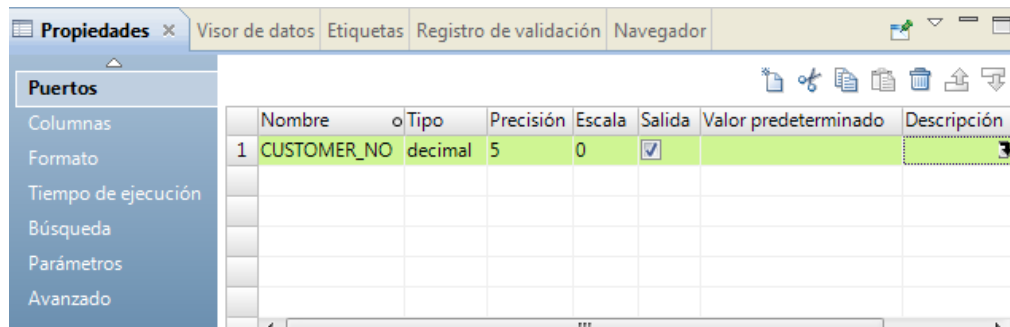
Crear una transformación de búsqueda no reutilizable

Cree una transformación de búsqueda no reutilizable como parte de una asignación o de un mapplet.

1. En una asignación o en un mapplet, arrastre una transformación de búsqueda desde la paleta Transformación hasta el editor.
A continuación, aparece el cuadro de diálogo **Nuevo**.
2. Seleccione **Búsqueda de objetos de datos de archivo sin formato**, **Búsqueda de objetos de datos lógicos**, **Búsqueda de tablas de referencia** o **Búsqueda de objetos de datos relacionales**.
3. Haga clic en **Siguiente**.
Aparece el cuadro de diálogo **Nueva transformación de búsqueda**.
4. Seleccione un objeto de datos físicos o una tabla de referencia en Developer Tool.
5. Especifique un nombre para la transformación.
6. En la lista desplegable **En coincidencias múltiples**, elija la estrategia que se usará cuando la transformación de búsqueda encuentre varias filas que coincidan con la condición de búsqueda.
7. Haga clic en **Finalizar**.
La transformación de búsqueda aparece en el editor.
8. Seleccione la transformación de búsqueda en el editor.
La barra de herramientas aparece encima de la transformación.

9. En la ficha **Puertos** de la vista **Propiedades**, añada puertos de salida a la transformación.

La siguiente figura muestra el puerto de salida CUSTOMER_NO en una transformación de búsqueda:



10. En la ficha **Tiempo de ejecución** de la vista **Propiedades**, seleccione **Almacenamiento en memoria caché de búsqueda habilitado** para habilitar la memoria caché de búsqueda.
Nota: Deberá añadir la transformación de búsqueda a una asignación antes de poder configurar las propiedades de búsqueda en tiempo de ejecución.
11. En la ficha **Búsqueda** de la vista **Propiedades**, agregue una o más condiciones de búsqueda.
12. En la ficha **Avanzadas** de la vista **Propiedades**, configure el nivel de seguimiento, las propiedades de la memoria caché de búsqueda dinámica y la conexión en tiempo de ejecución.
13. Guarde la transformación.

Crear una transformación de búsqueda no conectada

Cree una transformación de búsqueda no conectada cuando desee realizar una búsqueda desde una expresión. Puede crear una transformación de búsqueda no conectada reutilizable o no reutilizable en un objeto de datos de archivo sin formato, de tabla de referencia o relacionales.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
3. Navegue hasta el asistente Búsqueda.
4. Seleccione **Búsqueda de objetos de datos de archivos sin formato**, **Búsqueda de tablas de referencia** o **Búsqueda de objetos de datos relacionales**.
5. Haga clic en **Siguiente**.
Aparece el cuadro de diálogo **Nueva búsqueda**.
6. Seleccione un objeto de datos físicos o una tabla de referencia en la herramienta del desarrollador.
7. Especifique un nombre para la transformación.
8. En la lista desplegable **En coincidencias múltiples**, elija la estrategia que se usará cuando la transformación de búsqueda encuentre varias filas que coincidan con la condición de búsqueda.
Nota: No seleccione **Devolver todas** para una búsqueda no conectada.
9. Haga clic en **Finalizar**.
La transformación de búsqueda aparece en el editor.
10. En la sección **Puertos** de la vista **Resumen**, añada puertos a la transformación.

Cree un puerto de entrada para cada argumento en la expresión :LKP. Cree un puerto de entrada para cada condición de búsqueda que haya creado. Puede usar un puerto de entrada en varias condiciones.

11. En la sección **Puertos** de la vista **Resumen**, configure un puerto como puerto de devolución.
12. En la vista **Búsqueda**, añada una o más condiciones de búsqueda para comparar los valores de entrada de la transformación con valores del origen de la búsqueda o de la memoria caché.
Cuando la condición es verdadera, la búsqueda devuelve un valor en el puerto de devolución. Si la condición de búsqueda es falsa, la búsqueda devuelve NULL.
13. Cree una expresión :LKP para un puerto en una transformación que permita expresiones, como la transformación de agregación, la de expresión o la de estrategia de actualización.
14. Cuando cree una asignación, añada la transformación de búsqueda no conectada a la asignación en el editor, pero no conecte los puertos a las otras transformaciones de la asignación.

Ejemplo de búsqueda no conectada

Un comercio minorista de California añade un impuesto de venta estatal a cada precio de los artículos que vende a los clientes del estado. El importe del impuesto se basa en el condado donde reside el cliente. Para recuperar el impuesto de venta, cree una transformación de búsqueda que reciba el nombre del condado y que luego devuelva el importe del impuesto de venta para el condado. Si el condado no carga impuesto de venta, la transformación de búsqueda devuelve NULL. Invoque la búsqueda desde una transformación de expresión.

Realice los siguientes pasos para configurar una búsqueda no conectada del impuesto de venta por condado:

1. Importe un objeto de datos físicos de archivo sin formato que contenga los importes del impuesto de venta por condado.
2. Cree la transformación de búsqueda no conectada.
3. Añada puertos de entrada a la transformación de búsqueda.
4. Defina el puerto de devolución.
5. Cree la condición de búsqueda.
6. Invoque la búsqueda desde una transformación de expresión.

Paso 1. Importar el origen de la búsqueda del impuesto de venta en el repositorio de modelos.

El archivo de impuestos de venta debe encontrarse en el repositorio de modelos antes de crear la transformación de búsqueda. En este escenario, el archivo de impuestos de venta contiene dos campos: Sales_County y County_SalesTax. El condado es una cadena que contiene un nombre de condado. County_SalesTax es un campo decimal que contiene una tasa de impuesto para el condado. El archivo de impuestos de venta es el origen de la búsqueda.

Paso 2. Crear la transformación de búsqueda no conectada

Cree una transformación de búsqueda de archivo sin formato reutilizable con el objeto de datos de archivo sin formato de impuestos de venta. En este escenario, el nombre de la transformación es Sales_Tax_Lookup. Seleccione **Devolver la primera fila** en varias coincidencias.

Paso 3. Definir los puertos de la transformación de búsqueda

Defina los puertos de la transformación de búsqueda en la ficha **Puertos** de la vista **Propiedades**.

Tipo de puerto	Nombre	Tipo	Longitud	Escala
Entrada	In_County	String	25	
Salida	SalesTax	Decimal	3	3

Paso 4. Configurar el puerto de devolución de la transformación de búsqueda

El puerto de devolución es el campo del archivo sin formato que la búsqueda recupera. En la ficha **Columnas**, la columna `County_SalesTax` es el puerto de devolución.

Cuando la búsqueda da un resultado verdadero, el servicio de integración encuentra el condado en el origen de archivo sin formato. El servicio de integración devuelve un valor de impuesto de venta en el puerto de devolución. Si el servicio de integración no encuentra el condado, el resultado de la búsqueda será falso y el servicio de integración devolverá NULL en el puerto de devolución.

Paso 5. Definir la condición de búsqueda

En la vista **Búsqueda**, defina la condición de búsqueda para comparar el valor de entrada con valores del origen de la búsqueda.

Para añadir la condición de búsqueda, haga clic en **Columna de búsqueda**.

La condición de búsqueda tiene la siguiente sintaxis:

```
SALES_COUNTY = IN_COUNTY
```

Paso 6. Crear una transformación de expresión

Cree una transformación de expresión que reciba los registros de ventas desde un archivo sin formato. La transformación de expresión recibe un número de cliente, un importe de venta y el condado de la venta. Devuelve el número de cliente, el importe de venta y un importe de venta.

La transformación de expresión tiene los siguientes puertos:

Tipo de puerto	Nombre	Tipo	Longitud	Precisión	Valor predeterminado
Entrada	Condado	String	25	10	
Transferencia	Cliente	String	10		
Transferencia	SalesAmt	Decimal	10	2	
Salida	SalesTax	Decimal	10	2	0

El puerto `SalesTax` contiene una expresión `:LKP`. La expresión invoca la transformación `Sales_Tax_Lookup` y pasa el nombre del condado como parámetro. La transformación `Sales_Tax_Lookup` devuelve la tasa del impuesto de venta a la expresión. La transformación de expresión multiplica la tasa del impuesto por el importe de la venta.

Especifique la siguiente expresión para el puerto `SalesTax`:

```
(:LKP.Sales_Tax_Lookup(County) * SalesAmt)
```

El puerto `SalesTax` contiene el resultado de la expresión. Si la búsqueda es infructuosa, la transformación de búsqueda devuelve NULL y el puerto `SalesTax` contendrá valores nulos.

Puede añadir una expresión para comprobar si hay valores nulos en el puerto `SalesTax`. Si `SalesTax` es NULL puede configurar el puerto `SalesTax` para que devuelva cero. Añada el siguiente texto a la expresión de búsqueda para comprobar si hay valores nulos y devolver cero:

```
IIF(ISNULL(:LKP.Sales_Tax_Lookup(County) * SalesAmt),0, SalesTax)
```

Transformación de búsqueda en un entorno no nativo

El procesamiento de transformaciones de búsqueda en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de búsqueda en el motor de Blaze

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

Varias coincidencias

La transformación de búsqueda busca valores en función de la condición que configure en la transformación. Elija cómo manejar varias coincidencias en el origen de búsqueda.

Devolver la primera fila; Devolver la última fila

La transformación devuelve el primer resultado coincidente o el último resultado coincidente.

El servicio de integración de datos ordena los resultados para identificar la primera y la última fila. Las siguientes reglas determinan el orden de los resultados:

- El orden depende de los puertos de búsqueda presentes en la condición de búsqueda y los puertos de salida de búsqueda.
- Los valores numéricos se clasifican en orden ascendente.
- Los valores de cadena se clasifican de forma lexicográfica.
- Los valores de fecha se clasifican por la fecha más antigua primero.
- Cuando las asignaciones contienen una comparación que no es equijoin (\leq , \geq , $<$, $>$, \neq) y algunas filas contienen valores NULL, los resultados pueden ser diferentes en función del motor de tiempo de ejecución:
 - En el motor de Spark, los resultados NULL se tratan como si fueran TRUE. Para obtener más información sobre cómo procesa Spark los valores nulos, consulte la documentación de Apache: <https://spark.apache.org/docs/3.0.0-preview/sql-ref-null-semantics.html#comp-operators>
 - En los motores de Blaze o nativos, los resultados NULL se tratan como si fueran FALSE.

Devolver cualquier fila

La transformación devuelve cualquiera de las filas que coincidan con la condición de búsqueda. La transformación crea un índice basado en los puertos de clave en lugar de todos los puertos de la transformación de búsqueda. Cuando elige esta opción, el rendimiento puede mejorar porque el proceso de indexación de filas es más sencillo.

Devolver todas las filas

La transformación de búsqueda devuelve todas las filas que coinciden con la condición de búsqueda.

Informar sobre errores

Cuando la transformación de búsqueda utiliza una memoria caché estática o no utiliza memoria caché, el servicio de integración de datos marca la fila como un error. La transformación de búsqueda escribe la fila en el registro de la sesión de forma predeterminada y aumenta el recuento de errores en uno.

Cuando la transformación de búsqueda tiene una memoria caché dinámica, el servicio de integración de datos genera un error en la sesión cuando encuentra varias coincidencias. La sesión falla mientras el servicio de integración de datos almacena en caché la tabla de búsqueda o busca los valores de clave duplicados.

Además, si configura la transformación de búsqueda para generar valores antiguos en las actualizaciones, la transformación de búsqueda devuelve un error cuando encuentra varias coincidencias. La transformación crea un índice basado en los puertos de clave en lugar de todos los puertos de la transformación de búsqueda.

Reglas y directrices

La validación de asignaciones genera un error en las siguientes situaciones:

- La distinción entre mayúsculas y minúsculas está deshabilitada.
- La condición de búsqueda contiene el tipo de datos binario.
- La condición de búsqueda utiliza un campo con un tipo de datos complejo.
- La memoria caché está configurada para que sea compartida, con nombre, persistente, dinámica o no almacenada en caché. La memoria caché debe ser una caché estática.

Si se añade un objeto de datos que usa Sqoop como transformación de búsqueda en una asignación, el servicio de integración de datos no ejecuta la asignación mediante Sqoop, sino mediante JDBC.

Transformación de búsqueda en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Varias coincidencias

La transformación de búsqueda busca valores en función de la condición que configure en la transformación. Elija cómo manejar varias coincidencias en el origen de búsqueda.

Devolver la primera fila; Devolver la última fila

La transformación devuelve el primer resultado coincidente o el último resultado coincidente.

El servicio de integración de datos ordena los resultados para identificar la primera y la última fila. Las siguientes reglas determinan el orden de los resultados:

- El orden depende de los puertos de búsqueda presentes en la condición de búsqueda y los puertos de salida de búsqueda.
- Los valores numéricos se clasifican en orden ascendente.
- Los valores de cadena se clasifican de forma lexicográfica.
- Los valores de fecha se clasifican por la fecha más antigua primero.

- Cuando las asignaciones contienen una comparación que no es equijoin (<=,>=,<,>,<!=) y algunas filas contienen valores NULL, los resultados pueden ser diferentes en función del motor de tiempo de ejecución:
 - En el motor de Spark, los resultados NULL se tratan como si fueran TRUE. Para obtener más información sobre cómo procesa Spark los valores nulos, consulte la documentación de Apache: <https://spark.apache.org/docs/3.0.0-preview/sql-ref-null-semantics.html#comp-operators>
 - En los motores de Blaze o nativos, los resultados NULL se tratan como si fueran FALSE.

Devolver cualquier fila

La transformación devuelve cualquiera de las filas que coincidan con la condición de búsqueda. La transformación crea un índice basado en los puertos de clave en lugar de todos los puertos de la transformación de búsqueda. Cuando elige esta opción, el rendimiento puede mejorar porque el proceso de indexación de filas es más sencillo.

Devolver todas las filas

La transformación de búsqueda devuelve todas las filas que coinciden con la condición de búsqueda.

Informar sobre errores

Cuando la transformación de búsqueda utiliza una memoria caché estática o no utiliza memoria caché, el servicio de integración de datos marca la fila como un error. La transformación de búsqueda escribe la fila en el registro de la sesión de forma predeterminada y aumenta el recuento de errores en uno.

Cuando la transformación de búsqueda tiene una memoria caché dinámica, el servicio de integración de datos genera un error en la sesión cuando encuentra varias coincidencias. La sesión falla mientras el servicio de integración de datos almacena en caché la tabla de búsqueda o busca los valores de clave duplicados.

Además, si configura la transformación de búsqueda para generar valores antiguos en las actualizaciones, la transformación de búsqueda devuelve un error cuando encuentra varias coincidencias. La transformación crea un índice basado en los puertos de clave en lugar de todos los puertos de la transformación de búsqueda.

Reglas y directrices

La validación de asignaciones genera un error en las siguientes situaciones:

- La distinción entre mayúsculas y minúsculas está deshabilitada.
- La condición de búsqueda contiene el tipo de datos binario.
- La condición de búsqueda utiliza un campo con un tipo de datos complejo.
- La memoria caché está configurada para que sea compartida, con nombre, persistente, dinámica o no almacenada en caché. La memoria caché debe ser una caché estática.

La asignación genera un error en la siguiente situación:

- La transformación no está conectada y se usa con una transformación de combinación o Java.

Nota: Si una búsqueda HBase no devuelve una coincidencia, generará una fila con valores nulos para todas las columnas. Puede añadir una transformación de filtro después de la transformación de búsqueda para filtrar filas nulas.

Transformación de búsqueda en el motor de Databricks Spark

Algunas reglas de procesamiento del motor de Databricks Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Varias coincidencias

Si se elige devolver el primer valor, el último valor o cualquiera de ellos cuando existen varias coincidencias, la transformación de búsqueda devuelve cualquier valor.

Si configura la transformación para que notifique un error cuando haya varias coincidencias, el motor de Databricks Spark descarta las filas duplicadas y no incluye las filas en los registros.

Nota: Si una búsqueda HBase no devuelve una coincidencia, generará una fila con valores nulos para todas las columnas. Puede añadir una transformación de filtro después de la transformación de búsqueda para filtrar filas nulas.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La distinción entre mayúsculas y minúsculas está deshabilitada.
- La condición de búsqueda contiene el tipo de datos binario.
- La condición de búsqueda utiliza un campo con un tipo de datos complejo.
- La memoria caché está configurada para que sea compartida, con nombre, persistente, dinámica o no almacenada en caché. La memoria caché debe ser una caché estática.
- El origen de búsqueda no es Microsoft Azure SQL Data Warehouse.

Transformación de búsqueda en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La búsqueda es un objeto de datos.
- Una transformación de agregación está en el mismo canal de transmisión que una transformación de búsqueda pasiva configurada con una condición de búsqueda de desigualdad.
- Una transformación de rango está en el mismo canal de transmisión que una transformación de búsqueda pasiva configurada con una condición de búsqueda de desigualdad.
- Un canal contiene más de una transformación de búsqueda pasiva configurado con una condición de desigualdad.

La asignación genera un error en las siguientes situaciones:

- La transformación no está conectada.

Directrices generales

Tenga en cuenta las siguientes directrices generales:

- Si se utiliza un tipo de datos flotantes para buscar datos, se pueden devolver resultados no esperados.
- Utilice una transformación de búsqueda para buscar datos en un archivo sin formato, HDFS, Hive, relacional, JDBC V2 y HBase.

- Para evitar la combinación cruzada de DataFrames, configure la transformación de búsqueda para que ignore los valores nulos que coincidan.

Búsquedas de HBase

Para utilizar una transformación de búsqueda en tablas HBase que no están almacenadas en caché, lleve a cabo los siguientes pasos:

1. Cree un objeto de datos HBase. Cuando añada una tabla HBase como el recurso de un objeto de datos HBase, incluya la columna ROW ID.
2. Cree una operación de datos de lectura de HBase e impórtela en la asignación de transmisión.
3. Cuando importe la operación de datos en la asignación, seleccione la opción **Búsqueda**.
4. En la ficha Búsqueda, configure las siguientes opciones:
 - Columna de búsqueda. Especifique una condición de igualdad en ROW ID.
 - Operador. Especifique =.
5. Compruebe que el formato de cualquier valor de fecha de las tablas HBase corresponde a un formato de fecha Java válido. Especifique este formato en la propiedad **Formato de fecha y hora** de la ficha **Propiedades avanzadas** de la operación de lectura de objetos de datos.

Nota: Si una búsqueda HBase no devuelve una coincidencia, generará una fila con valores nulos para todas las columnas. Puede añadir una transformación de filtro después de la transformación de búsqueda para filtrar filas nulas.

La validación de asignaciones genera un error en las siguientes situaciones:

- La condición no contiene un ROW ID.
- La transformación contiene una condición de desigualdad.
- La transformación contiene varias condiciones.
- Una columna de entrada es de un tipo de fecha.

Búsquedas de JDBC V2

Puede añadir una operación de lectura de objeto de datos JDBC V2 como una búsqueda en una asignación de transmisión. Puede ejecutar asignaciones de transmisiones con búsquedas de JDBC V2 en el servicio Azure Databricks, en los servicios de nube de Microsoft Azure.

CAPÍTULO 26

Memorias caché de búsqueda

Este capítulo incluye los siguientes temas:

- [Introducción a las memorias caché de búsqueda, 451](#)
- [Tipos de memoria caché de búsqueda, 452](#)
- [Búsqueda no almacenada en caché, 453](#)
- [Memoria caché de búsqueda estática, 453](#)
- [Memoria caché de búsqueda persistente, 454](#)
- [Memoria caché de búsqueda dinámica, 455](#)
- [Memoria caché de búsqueda compartida, 455](#)
- [Comparación de memorias caché, 457](#)
- [Partición de la memoria caché para búsquedas, 457](#)

Introducción a las memorias caché de búsqueda

Puede configurar una transformación de búsqueda para almacenar en caché un origen de búsqueda de archivo sin formato o relacional. Habilite la memoria caché de búsqueda en un archivo o una tabla de búsqueda grandes para aumentar el rendimiento de la búsqueda.

El Servicio de integración genera una memoria caché en la memoria cuando procesa la primera fila de datos de una transformación de búsqueda en caché. El Servicio de integración crea memorias caché a medida que se introducen las filas de origen en la transformación de búsqueda. Asigna memoria para la memoria caché según la cantidad que haya configurado en la transformación. El Servicio de integración almacena los valores de condición en la memoria caché de índice y en los valores de salida de la memoria caché de datos. El Servicio de integración consulta en la memoria caché cada fila que entra en la transformación.

Si los datos no encajan en la memoria caché, el Servicio de integración almacena los valores de desbordamiento en los archivos de memoria caché. El Servicio de integración crea los archivos de memoria caché en el directorio de la memoria caché. De forma predeterminada, el Servicio de integración crea los archivos de la memoria caché en el directorio especificado en el parámetro de sistema CacheDir. Cuando se completa la asignación, el Servicio de integración de datos libera la memoria caché y elimina los archivos de ésta a menos que configure la transformación de búsqueda para que utilice una memoria caché persistente.

Si utiliza una búsqueda de archivos sin formato, el Servicio de integración almacena en caché el origen de la búsqueda. Si configura una búsqueda de archivo sin formato para la entrada ordenada, el Servicio de integración no puede almacenar en memoria caché la búsqueda si las columnas de la condición no están agrupadas. En caso de que lo estén y de que no estén ordenadas, el Servicio de integración procesa la búsqueda como si no se hubiera configurado la entrada ordenada.

Cuando no se configura la transformación de búsqueda para la memoria caché, el Servicio de integración consulta el origen de búsqueda para cada fila de entrada. Tanto si el origen de búsqueda se almacena en caché como si no, el resultado de la consulta de búsqueda y del procesamiento será el mismo. Sin embargo, puede aumentar el rendimiento de la búsqueda en un origen de búsqueda grande si habilita la memoria caché de búsqueda.

Tipos de memoria caché de búsqueda

Puede configurar distintos tipos de memorias caché de búsqueda. Por ejemplo, puede configurar una memoria caché compartida si desea compartir la memoria caché entre varias transformaciones de búsqueda en la misma asignación.

Puede configurar los siguientes tipos de memorias caché de búsqueda:

Memoria caché estática

Una memoria caché estática no cambia mientras el Servicio de integración procesa la búsqueda. El Servicio de integración reconstruye una memoria caché estática cada vez que procesa la búsqueda. De manera predeterminada, el Servicio de integración crea una memoria caché estática cuando se habilita el almacenamiento en caché para una transformación de búsqueda. El Servicio de integración genera la memoria caché cuando procesa la primera solicitud de búsqueda. Busca valores en la memoria caché para cada fila que entra en la transformación de búsqueda. Cuando la condición de búsqueda es verdadera, el Servicio de integración devuelve un valor de la memoria caché de búsqueda.

Utilice una memoria caché estática por las siguientes razones:

- El origen de la búsqueda no cambia mientras se ejecuta la asignación.
- La búsqueda es una búsqueda no conectada. Debe utilizar una memoria caché estática para una búsqueda no conectada.
- Para aumentar el rendimiento. Como el Servicio de integración no actualiza la memoria caché mientras procesa la transformación de búsqueda, el Servicio de integración procesa una transformación de búsqueda con una memoria caché estática más rápidamente que una transformación de búsqueda con una memoria caché dinámica.
- Cuando la condición de búsqueda es falsa, quiere que el Servicio de integración devuelva el valor predeterminado para las transformaciones conectadas o NULL para las transformaciones no conectadas.

Memoria caché persistente

Una memoria caché persistente no cambia cada vez que el Servicio de integración procesa la búsqueda. El Servicio de integración guarda los archivos de memoria caché de búsqueda y los reutiliza la siguiente vez que procesa una transformación de búsqueda configurada para utilizar la memoria caché. Utilice una memoria caché persistente cuando el origen de búsqueda no cambie.

Puede configurar la transformación de búsqueda para que regenere una memoria caché de búsqueda persistente si fuera necesario.

Memoria caché dinámica

Una memoria caché de búsqueda dinámica cambia mientras el Servicio de integración procesa la búsqueda. El Servicio de integración genera una memoria caché de búsqueda dinámica cuando procesa la primera solicitud de búsqueda. Cuando procesa cada fila, el Servicio de integración inserta o actualiza dinámicamente los datos en la memoria caché de búsqueda y pasa los datos al destino. La memoria caché dinámica se sincroniza con el destino.

Utilice una memoria caché dinámica cuando desee actualizar un destino de acuerdo con los registros nuevos y modificados. También puede utilizar una memoria caché dinámica cuando la asignación requiera una búsqueda en los datos de destino, pero la conexión con el destino sea lenta.

Memoria caché compartida

Varias transformaciones de búsqueda pueden utilizar una memoria caché compartida en la misma asignación. Utilice una memoria caché compartida para aumentar el rendimiento de la asignación. En lugar de generar una memoria caché de búsqueda aparte para cada transformación de búsqueda, el Servicio de integración genera una memoria caché.

Búsqueda no almacenada en caché

Una búsqueda no almacenada en memoria caché se produce cuando el Servicio de integración no almacena en caché el origen de búsqueda. De manera predeterminada, el Servicio de integración no utiliza una memoria caché de búsqueda para una transformación de búsqueda.

El Servicio de integración procesa una búsqueda no almacenada en memoria caché del mismo modo que procesa una búsqueda almacenada en caché, salvo que consulta el origen de búsqueda en lugar de generar y consultar la memoria caché de búsqueda.

Cuando la condición de búsqueda es verdadera, el Servicio de integración devuelve los valores del origen de búsqueda. Cuando el Servicio de integración procesa una transformación de búsqueda conectada, devuelve los valores representados por los puertos de búsqueda/salida. Cuando procesa una transformación de búsqueda no conectada, el Servicio de integración devuelve el valor representado por el puerto de devolución.

Cuando la condición no es verdadera, el Servicio de integración devuelve NULL o los valores predeterminados. Cuando el Servicio de integración procesa una transformación de búsqueda conectada, devuelve el valor predeterminado del puerto de salida cuando no se cumple la condición. Cuando procesa una transformación de búsqueda no conectada, el Servicio de integración devuelve NULL si no se cumple la condición.

Memoria caché de búsqueda estática

Una memoria caché de búsqueda estática es una memoria caché que el Servicio de integración no actualiza cuando procesa la transformación de búsqueda. De manera predeterminada, el Servicio de integración crea una memoria caché de búsqueda estática al configurar una transformación de búsqueda para el almacenamiento en caché.

El Servicio de integración genera la memoria caché cuando procesa la primera solicitud de búsqueda. Consulta la memoria caché según la condición de búsqueda de cada fila que pasa a la transformación.

Cuando la condición de búsqueda es verdadera, el Servicio de integración devuelve los valores de la memoria caché de búsqueda estática. Cuando el Servicio de integración procesa una transformación de búsqueda conectada, devuelve los valores representados por los puertos de búsqueda/salida. Cuando el Servicio de integración procesa una transformación de búsqueda no conectada, devuelve el valor representado por el puerto de devolución.

Cuando la condición no es verdadera, el Servicio de integración devuelve NULL o los valores predeterminados. Cuando el Servicio de integración procesa una transformación de búsqueda conectada, devuelve el valor predeterminado del puerto de salida cuando no se cumple la condición. Cuando el Servicio

de integración procesa una transformación de búsqueda no conectada, devuelve NULL si no se cumple la condición.

Memoria caché de búsqueda persistente

Una memoria caché de búsqueda persistente es una memoria caché que el Servicio de integración reutiliza para varias ejecuciones de la misma asignación. Utilice una memoria caché de búsqueda persistente cuando el origen de búsqueda no cambie entre las ejecuciones de asignaciones.

De forma predeterminada, el Servicio de integración utiliza una memoria caché no persistente cuando se habilita la memoria caché de búsqueda en una transformación de búsqueda. El Servicio de integración elimina los archivos de memoria caché cuando se completa la asignación. La próxima vez que ejecute la asignación, el Servicio de integración generará la memoria caché desde el origen de búsqueda.

Si configura la transformación de búsqueda para que utilice una memoria caché de búsqueda persistente, el Servicio de integración guardará y reutilizará los archivos de memoria caché para varias ejecuciones de la asignación. Una memoria caché persistente elimina el tiempo necesario para leer la tabla de búsqueda y regenerar la memoria caché de búsqueda.

La primera vez que el Servicio de integración ejecute una asignación con una memoria caché de búsqueda persistente, el Servicio de integración guardará los archivos de caché en disco. La próxima vez que el Servicio de integración ejecute la asignación, generará la memoria caché a partir de los archivos de memoria caché.

Puede configurar el Servicio de integración para que regenere una memoria caché de búsqueda persistente si cambia el origen de búsqueda. Cuando regenere una memoria caché, el Servicio de integración creará nuevos archivos de memoria caché y escribirá un mensaje en el registro del Servicio de integración.

Regeneración de una memoria caché de búsqueda persistente

Puede configurar el Servicio de integración para que regenere una memoria caché de búsqueda persistente. En algunos casos, el Servicio de integración regenerará la memoria caché de búsqueda persistente, incluso si no lo configura para que lo haga.

Tenga en cuenta las siguientes reglas y directrices cuando regenere una memoria caché de búsqueda persistente:

- Regenere una memoria caché de búsqueda persistente si el origen de búsqueda ha cambiado desde la última vez que el Servicio de integración generó la memoria caché.
- Puede regenerar la memoria caché cuando la asignación contenga una o más transformaciones de búsqueda que comparten una memoria caché.
- Si la tabla de búsqueda no cambia entre ejecuciones de sesión, configure la transformación de búsqueda para que utilice una memoria caché de búsqueda persistente. El Servicio de integración guarda y reutiliza los archivos de memoria caché, eliminando el tiempo necesario para leer la tabla de búsqueda.
- Si configura las transformaciones de búsqueda posteriores para que regeneren la memoria caché de búsqueda, el Servicio de integración compartirá la memoria caché en lugar de regenerar la memoria caché cuando procese la transformación de búsqueda posterior.
- Si una asignación contiene dos búsquedas persistentes y configura solo la segunda transformación de búsqueda para que regenere la memoria caché, el Servicio de integración regenerará la memoria caché de búsqueda persistente para las dos.

El Servicio de integración regenera la memoria caché de búsqueda persistente en los siguiente casos:

- El Servicio de integración no puede encontrar los archivos de memoria caché.
- El Servicio de integración no se puede reutilizar la memoria caché. En este caso, o bien regenera la memoria caché de búsqueda o genera un error en la asignación.
- Habilitar o deshabilita la alta precisión del Servicio de integración.
- Edita la transformación de búsqueda o la asignación.

Nota: El Servicio de integración no regenera la memoria caché si edita la descripción de la transformación.

- Cambia el número de particiones.
- Cambia la conexión de base de datos o la ubicación de archivo utilizadas para acceder al origen de búsqueda.
- Cambia el orden de clasificación en modo Unicode.
- Cambia la página de códigos del Servicio de integración.

Memoria caché de búsqueda dinámica

Una memoria caché dinámica es una memoria caché que el Servicio de integración actualiza cuando procesa cada fila. Utilice una memoria caché de búsqueda dinámica para mantener la memoria caché sincronizada con el destino.

Puede utilizar una memoria caché dinámica con una búsqueda de datos relacionales y una búsqueda de archivo sin formato. El Servicio de integración genera la memoria caché cuando procesa la primera solicitud de búsqueda. Consulta la memoria caché según la condición de búsqueda de cada fila que pasa a la transformación de búsqueda. El Servicio de integración actualiza la memoria caché de búsqueda cuando procesa cada fila.

Según los resultados de la consulta de búsqueda, el tipo de fila y las propiedades de transformación de búsqueda, el Servicio de integración inserta o actualiza las filas en la memoria caché, o no realiza ningún cambio en la memoria caché.

Memoria caché de búsqueda compartida

Una memoria caché de búsqueda compartida es una memoria caché de búsqueda estática que comparten varias transformaciones de búsqueda en una asignación. Utilice una memoria caché de búsqueda compartida para reducir la cantidad de tiempo necesario para generar la memoria caché.

De forma predeterminada, el Servicio de integración comparte la memoria caché para las transformaciones de búsqueda en una asignación que tengan estructuras compatibles de almacenamiento en caché. Por ejemplo, si una asignación contiene dos instancias de la misma transformación de búsqueda reutilizable en una asignación y utiliza los mismos puertos de salida para ambas instancias, las transformaciones de búsqueda compartirán la memoria caché de búsqueda de manera predeterminada.

El Servicio de integración genera la memoria caché cuando procesa la primera transformación de búsqueda. Utiliza la misma memoria caché para procesar las transformaciones de búsqueda posteriores que comparten la memoria caché. Cuando el Servicio de integración comparte una memoria caché de búsqueda, escribe un mensaje en el registro del Servicio de integración.

El Servicio de integración asigna la memoria caché de datos y la memoria caché de índice para la primera transformación de búsqueda. No asigna memoria adicional para las transformaciones de búsqueda posteriores que comparten la memoria caché de búsqueda.

Si la transformación o la estructura de la memoria caché no permiten el uso compartido, el Servicio de integración crea una nueva memoria caché.

Reglas y directrices para compartir una memoria caché de búsqueda

Tenga en cuenta las siguientes reglas y directrices cuando comparta una memoria caché de búsqueda:

- Puede compartir una o más memorias caché estáticas con una búsqueda dinámica. Si una búsqueda dinámica comparte la memoria caché con una búsqueda estática en la misma asignación, la búsqueda estática reutiliza la memoria caché creada por la búsqueda dinámica.
- No puede compartir una memoria caché entre búsquedas dinámicas.
- Si configura varias transformaciones de búsqueda para regenerar una memoria caché de búsqueda persistente, el Servicio de integración generará la memoria caché para la primera transformación de búsqueda y después compartirá la memoria caché de búsqueda persistente para las transformaciones de búsqueda posteriores.
- Si no configura la primera transformación de búsqueda para regenerar una memoria caché de búsqueda persistente, pero configura una transformación de búsqueda posterior para regenerar la memoria caché, las transformaciones no podrán compartir la memoria caché. El Servicio de integración genera la memoria caché cuando procesa cada transformación de búsqueda.
- Los puertos de búsqueda/salida de las transformaciones de búsqueda posteriores deben coincidir o ser un subconjunto de los puertos de la transformación de búsqueda utilizados por el Servicio de integración para generar la memoria caché. El orden de los puertos no es necesario que coincida.
- Las transformaciones de búsqueda que comparten la memoria caché deben tener las siguientes características:
 - Las transformaciones de búsqueda deben utilizar los mismos puertos en la condición de búsqueda.
 - Las transformaciones de búsqueda deben utilizar el mismo reemplazo de SQL si se utiliza uno.
 - El almacenamiento en memoria caché de búsqueda debe estar habilitado en todas las transformaciones de búsqueda.
 - Las transformaciones de búsqueda deben utilizar el mismo tipo de origen de búsqueda.
 - Todas las transformaciones de búsqueda relacionales deben utilizar la misma conexión de base de datos.
 - Las transformaciones de búsqueda deben utilizar el mismo nombre de tabla de búsqueda.
 - La estructura de la memoria caché de todas las transformaciones de búsqueda debe ser compatible.

Comparación de memorias caché

El Servicio de integración tiene un funcionamiento diferente según el tipo de memoria caché de búsqueda que se configure.

La siguiente tabla compara las transformaciones de búsqueda con una búsqueda no almacenada en caché, una memoria caché estática y una memoria caché dinámica:

No almacenada en caché	Memoria caché estática	Memoria caché dinámica
El Servicio de integración no inserta ni actualiza la memoria caché.	El Servicio de integración no inserta ni actualiza la memoria caché.	El Servicio de integración puede insertar o actualizar filas en la memoria caché a medida que pasa filas al destino.
Puede utilizar una búsqueda relacional.	Puede utilizar una búsqueda de datos relacionales o de archivo sin formato.	Puede utilizar una búsqueda de datos relacionales o de archivo sin formato.
<p>Cuando la condición es verdadera, el Servicio de integración devuelve un valor de la tabla de búsqueda o de la memoria caché.</p> <p>Cuando la condición no es verdadera, el Servicio de integración devuelve el valor predeterminado para las transformaciones conectadas y NULL para las transformaciones no conectadas.</p>	<p>Cuando la condición es verdadera, el Servicio de integración devuelve un valor de la tabla de búsqueda o de la memoria caché.</p> <p>Cuando la condición no es verdadera, el Servicio de integración devuelve el valor predeterminado para las transformaciones conectadas y NULL para las transformaciones no conectadas.</p>	<p>Cuando la condición es verdadera, el Servicio de integración actualiza las filas en la memoria caché o deja la memoria caché sin cambios, según el tipo de fila. Esto indica que la fila está en la memoria caché y en la tabla de destino. Puede pasar filas actualizadas a un destino.</p> <p>Cuando la condición no es verdadera, el Servicio de integración inserta filas en la memoria caché o deja la memoria caché sin cambios, según el tipo de fila. Esto indica que la fila no está en la memoria caché ni en la tabla de destino. Puede pasar filas insertadas a una tabla de destino.</p>

Partición de la memoria caché para búsquedas

La partición de la memoria caché crea una memoria caché diferente para cada partición que procese una transformación de agregación, combinación, rango o búsqueda. La partición de la memoria caché aumenta el rendimiento de la asignación porque cada partición consulta una memoria caché diferente en paralelo.

Cuando el Servicio de integración crea particiones para una asignación, este podría utilizar la partición de la memoria caché de las transformaciones de búsqueda con particiones.

El Servicio de integración utiliza la partición de la memoria caché para las transformaciones de búsqueda conectadas cuando se cumplen las siguientes condiciones:

- La condición de búsqueda contiene solo los operadores de igualdad.
- Cuando la transformación de búsqueda conectada busca datos en una tabla relacional y la base de datos está configurada para la comparación con distinción de mayúsculas y minúsculas.

El Servicio de integración no utiliza la partición de la memoria caché para las transformaciones de búsqueda no conectadas.

Cuando el Servicio de integración no utiliza la partición de la memoria caché para una transformación de búsqueda, todas las particiones de la transformación de búsqueda comparten la misma memoria caché. Cada partición consulta la misma memoria caché en serie.

CAPÍTULO 27

Memoria caché de búsqueda dinámica

Este capítulo incluye los siguientes temas:

- [Introducción a la memoria caché de búsqueda dinámica, 458](#)
- [Usos de la memoria caché de búsqueda dinámica, 459](#)
- [Propiedades de la memoria caché de búsqueda dinámica, 459](#)
- [Memoria caché de búsqueda dinámica y valores de salida, 461](#)
- [Valores de transformación de búsqueda, 462](#)
- [Reemplazo de SQL y memoria caché de búsqueda dinámica, 464](#)
- [Configuración de asignación de una memoria caché de búsqueda dinámica, 465](#)
- [Actualizaciones de memoria caché de búsqueda dinámica condicional, 468](#)
- [Actualización de memoria caché dinámica con resultados de expresión, 469](#)
- [Ejemplo de memoria caché de búsqueda dinámica, 471](#)
- [Reglas y directrices de las memorias caché de búsqueda dinámica, 472](#)

Introducción a la memoria caché de búsqueda dinámica

Utilice una memoria caché de búsqueda dinámica para mantener la memoria caché sincronizada con el destino. Puede utilizar una memoria caché dinámica con una búsqueda de datos relacionales o una búsqueda de archivo sin formato.

El Servicio de integración genera la memoria caché de búsqueda dinámica cuando procesa la primera solicitud de búsqueda. Consulta la memoria caché según la condición de búsqueda de cada fila que pasa a la transformación. El Servicio de integración actualiza la memoria caché de búsqueda cuando procesa cada fila.

Según los resultados de la consulta de búsqueda, el tipo de fila y las propiedades de transformación de búsqueda, el Servicio de integración realiza una de las siguientes acciones en la memoria caché de búsqueda dinámica cuando lee una fila del origen:

Inserta la fila en la memoria caché

El Servicio de integración inserta la fila cuando la fila no está en la memoria caché y ha configurado la transformación de búsqueda para que inserte filas en la memoria caché. Puede configurar la

transformación para que inserte filas en la memoria caché según los puertos de entrada o los ID de secuencia generados. El Servicio de integración marca la fila para inserción.

Actualiza la fila en la memoria caché

El Servicio de integración actualiza la fila cuando la fila existe en la memoria caché y ha configurado la transformación de búsqueda para que actualice filas en la memoria caché. El Servicio de integración actualiza la fila en la memoria caché según los puertos de entrada. El Servicio de integración marca la fila como fila para actualizar.

No realiza cambios en la memoria caché

El Servicio de integración no realiza cambios cuando la fila existe en la memoria caché y ha configurado la transformación de búsqueda para que solo inserte filas nuevas. O bien, la fila no está en la memoria caché y ha especificado que solo se actualicen las filas existentes. O bien, la fila está en la memoria caché, pero según la condición de búsqueda, no hay cambios. El Servicio de integración marca la fila como sin cambios.

Según el valor de NewLookupRow, también puede configurar una transformación de enrutador o de filtro con la transformación de búsqueda dinámica para enrutar inserciones o actualizar filas en la tabla de destino. Puede enrutar filas sin cambios a otra tabla de destino o archivo sin formato, o puede soltarlas.

Usos de la memoria caché de búsqueda dinámica

Puede configurar una transformación de búsqueda con una memoria caché de búsqueda dinámica para actualizar la memoria caché según los cambios en el origen de búsqueda.

Puede usar una memoria caché de búsqueda dinámica por las siguientes razones:

Actualizar una tabla de clientes principal con información nueva y actualizada de los clientes.

Por ejemplo, puede utilizar una transformación de búsqueda para realizar una búsqueda en la tabla de clientes para determinar si existe un cliente en el destino. La memoria caché representa la tabla de clientes. La transformación de búsqueda inserta y actualiza filas en la memoria caché a medida que pasa filas al destino.

Utilice un archivo sin formato exportado como origen de búsqueda en lugar de una tabla relacional.

Si la conexión con la base de datos es lenta, puede exportar el contenido de la tabla relacional a un archivo sin formato y utilizar el archivo como origen de búsqueda. Por ejemplo, es posible que necesite utilizar este método si una conexión ODBC con una base de datos es demasiado lenta. Puede configurar la tabla de base de datos como destino relacional en la asignación y volver a pasar los cambios de la memoria caché de búsqueda a la tabla de base de datos.

Propiedades de la memoria caché de búsqueda dinámica

Configure las propiedades de búsqueda dinámica para habilitar una memoria caché de búsqueda dinámica y configurar la forma en que se actualiza la memoria caché. Por ejemplo, puede configurar los valores que se insertan y se actualizan en la memoria caché dinámica.

Configure las siguientes propiedades cuando habilite una memoria caché de búsqueda dinámica:

En coincidencias múltiples

Establecida como Informar sobre errores.

Memoria caché de búsqueda dinámica

Habilita una memoria caché de búsqueda dinámica.

Esta opción está disponible después de habilitar el almacenamiento en memoria caché de búsqueda.

Actualizar o insertar

Se aplica a las filas que se introducen en la transformación de búsqueda cuando el tipo de fila es de actualización. Cuando está habilitada, el Servicio de integración actualiza las filas existentes en la memoria caché e inserta una fila si es nueva. Si está deshabilitada, el Servicio de integración no inserta filas nuevas.

Esta opción está disponible después de habilitar el almacenamiento en memoria caché dinámica.

Insertar o actualizar

Se aplica a las filas que se introducen en la transformación de búsqueda cuando el tipo de fila es de inserción. Cuando está habilitada, el Servicio de integración inserta filas en la memoria caché y actualiza las filas existentes. Si está deshabilitada, el Servicio de integración no actualiza las filas existentes.

Esta opción está disponible después de habilitar el almacenamiento en memoria caché dinámica.

Proporcionar como salida el valor anterior al actualizar

La transformación de búsqueda puede generar la salida de valores existentes o nuevos a partir de la memoria caché. Cuando está habilitado, el Servicio de integración genera la salida de los valores de existente desde los puertos de búsqueda/salida antes de actualizar el valor en la memoria caché. Cuando el Servicio de integración actualiza una fila en la memoria caché, genera la salida del valor en la memoria caché de búsqueda antes de actualizar la fila según los datos de entrada. Cuando el Servicio de integración inserta una fila en la memoria caché, genera valores nulos.

Deshabilite la propiedad para que el Servicio de integración pase los mismos valores desde los puertos de búsqueda/salida y de entrada/salida. Esta propiedad está habilitada de forma predeterminada.

Esta opción está disponible después de habilitar el almacenamiento en memoria caché dinámica.

Actualización de la condición de memoria caché dinámica

Cuando está habilitado, el Servicio de integración utiliza la expresión de condición para determinar si se debe actualizar la memoria caché dinámica. El Servicio de integración actualiza la memoria caché cuando la condición es verdadera y los datos están en la memoria caché.

Cree la expresión con puertos de búsqueda o con puertos de entrada. La expresión puede contener valores de entrada o valores en la memoria caché de búsqueda. El valor predeterminado es verdadero.

Esta opción está disponible después de habilitar el almacenamiento en memoria caché dinámica.

NewLookupRow

La Herramienta del desarrollador añade este puerto a una transformación de búsqueda que está configurada con una memoria caché dinámica.

La propiedad NewLookupRow puede contener uno de los siguientes valores:

- 0 = Sin actualización en la memoria caché.
- 1 = Insertar fila en la memoria caché.
- 2 = Actualizar fila en la memoria caché.

Para mantener sincronizadas la memoria caché de búsqueda y la tabla de destino, pase filas al destino cuando el valor de NewLookupRow sea igual a 1 o 2.

Puerto asociado

El Servicio de integración utiliza el valor del puerto asociado cuando actualiza los datos en la memoria caché. El Servicio de integración asocia los puertos de entrada y los puertos de origen de búsqueda especificados en la condición de búsqueda. Debe configurar puertos asociados para el resto de puertos de origen de búsqueda en la búsqueda dinámica. Si no configura un puerto asociado para todos los puertos de origen de búsqueda en una búsqueda dinámica, se produce un error de la validación de asignaciones.

Puede asociar un puerto de origen de búsqueda a los siguientes objetos:

Objeto	Descripción
Puerto de entrada	Actualiza la memoria caché según el valor de un puerto de entrada.
Expresión asociada	Selecciónelo para especificar una expresión. El Servicio de integración actualiza la memoria caché según el resultado de la expresión.
ID de secuencia	Genera una clave principal para las filas insertadas en la memoria caché de búsqueda. Puede asociar un ID de secuencia solo con columnas de bigint y de enteros.

Omitir las entradas NULL para las actualizaciones

La Herramienta del desarrollador activa esta propiedad de puerto para los puertos de búsqueda/salida cuando se configura la transformación de búsqueda para utilizar una memoria caché dinámica. Seleccione esta propiedad si no quiere que el Servicio de integración actualice la columna en la memoria caché con un valor de entrada NULL.

Omitir en la comparación

La Herramienta del desarrollador activa esta propiedad de puerto para los puertos de búsqueda/salida que no se usan en la condición de búsqueda cuando se configura la transformación de búsqueda para utilizar una memoria caché dinámica. De manera predeterminada, el Servicio de integración compara los valores de todos los puertos de búsqueda con los valores de sus puertos asociados. Seleccione esta propiedad si desea que el Servicio de integración omita el puerto cuando compara valores antes de actualizar una fila. Utilice esta propiedad para aumentar el rendimiento de la comparación.

Memoria caché de búsqueda dinámica y valores de salida

Si habilita una memoria caché de búsqueda dinámica, los valores del puerto de salida varían según la forma en que se configure la memoria caché de búsqueda dinámica. El valor de salida del puerto de búsqueda/salida depende de si se elige generar valores antiguos o nuevos cuando el servicio de integración actualiza una fila.

Puede configurar la propiedad **Proporcionar como salida el valor anterior al actualizar** para especificar uno de los siguientes tipos de valores de salida para un puerto de búsqueda/salida:

- Proporcionar como salida valores anteriores al actualizar. El servicio de integración genera el valor que existía en la memoria caché antes de actualizar la fila.
- Proporcionar como salida valores nuevos al actualizar. El servicio de integración genera el valor actualizado que escribe en la memoria caché. El valor del puerto de búsqueda/salida coincide con el valor del puerto de salida.

Valores de transformación de búsqueda

La transformación de búsqueda contiene los valores de los puertos de entrada, de búsqueda y de salida. Si habilita una memoria caché de búsqueda dinámica, los valores del puerto de salida varían según la forma en que se configure la memoria caché de búsqueda dinámica.

La transformación de búsqueda contiene los siguientes tipos de valores:

Valor de entrada

Valor que el Servicio de integración pasa a la transformación de búsqueda.

Valor de búsqueda

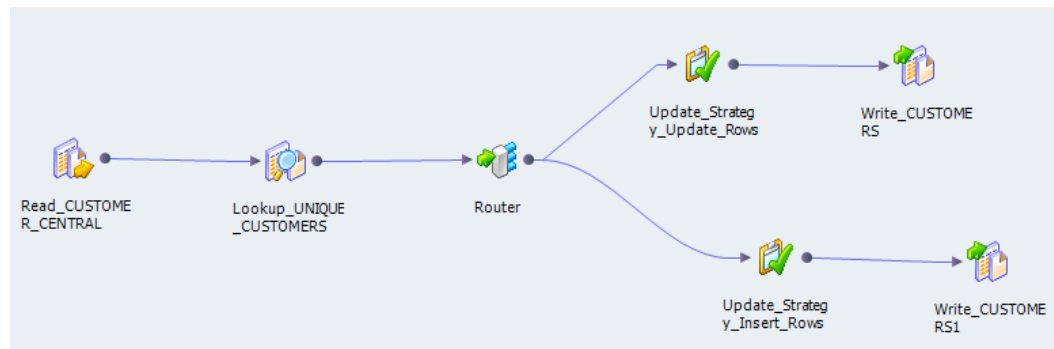
El valor que el Servicio de integración inserta en la memoria caché.

Valor de salida

Valor que el Servicio de integración pasa desde el puerto de salida de la transformación de búsqueda. El valor de salida del puerto de búsqueda/salida depende de si se elige generar valores antiguos o nuevos cuando el Servicio de integración actualiza una fila.

Ejemplo de valores de transformación de búsqueda

Por ejemplo, quiere crear una asignación con los siguientes objetos:



Para la transformación de búsqueda, habilite la memoria caché de búsqueda dinámica y defina la siguiente condición de búsqueda:

```
IN_CUST_ID = CUST_ID
```

De manera predeterminada, el tipo de fila de todas las filas que se introducen en la transformación de búsqueda es de inserción. Para realizar tanto inserciones como actualizaciones en la memoria caché y en la tabla de destino, seleccione la propiedad **Insertar o actualizar** en la transformación de búsqueda.

Valores de memoria caché inicial

Cuando ejecute la asignación, el Servicio de integración compilará la memoria caché de búsqueda desde la tabla de destino.

La siguiente tabla muestra los valores iniciales de la memoria caché de búsqueda:

PK_PRIMARYKEY	CUST_ID	CUST_NAME	ADDRESS
100001	80001	Marion James	100 Main St.
100002	80002	Laura Jones	510 Broadway Ave.

PK_PRIMARYKEY	CUST_ID	CUST_NAME	ADDRESS
100003	80003	Shelley Lau	220 Burnside Ave.

Valores de entrada

El origen contiene filas que existen y filas que no existen en la tabla de destino. El Servicio de integración pasa las filas de origen a la transformación de búsqueda.

La siguiente tabla muestra las filas de origen:

SQ_CUST_ID	SQ_CUST_NAME	SQ_ADDRESS
80001	Marion Atkins	100 Main St.
80002	Laura Gomez	510 Broadway Ave.
99001	Jon Freeman	555 6th Ave.

Valores de búsqueda

El Servicio de integración busca valores en la memoria caché según la condición de búsqueda. Actualiza las filas de la memoria caché para los ID de cliente existentes 80001 y 80002. Inserta una fila en la memoria caché para el ID de cliente 99001. El Servicio de integración genera una nueva clave (PK_PRIMARYKEY) para la nueva fila.

La siguiente tabla muestra las filas y los valores devueltos desde la búsqueda:

PK_PRIMARYKEY	CUST_ID	CUST_NAME	ADDRESS
100001	80001	Marion Atkins	100 Main St.
100002	80002	Laura Gomez	510 Broadway Ave.
100004	99001	Jon Freeman	555 6th Ave.

Valores de salida

El Servicio de integración marca las filas en la transformación de búsqueda según las inserciones y actualizaciones que realiza en la memoria caché dinámica. El Servicio de integración finalmente pasa las filas a una transformación de enrutador que crea una ramificación para insertar filas y otra ramificación para actualizar filas. Cada ramificación contiene una transformación de estrategia de actualización. Las transformaciones de estrategia de actualización marcan las filas para inserción o actualización según el valor del puerto NewLookupRow.

Los valores de salida de los puertos de búsqueda/salida y entrada/salida dependen de si se elige generar valores antiguos o nuevos cuando el Servicio de integración actualiza una fila. Sin embargo, los valores de salida del puerto NewLookupRow y cualquier puerto de búsqueda/salida que utiliza el ID de secuencia es el mismo para las filas nuevas y actualizadas.

Si elige generar la salida de valores nuevos, los puertos de búsqueda/salida generan los siguientes valores:

NewLookupRow	PK_PRIMARYKEY	CUST_ID	CUST_NAME	ADDRESS
2	100001	80001	Marion Atkins	100 Main St.
2	100002	80002	Laura Gomez	510 Broadway Ave.
1	100004	99001	Jon Freeman	555 6th Ave.

Si elige generar la salida de valores antiguos, los puertos de búsqueda/salida generan los siguientes valores:

NewLookupRow	PK_PRIMARYKEY	CUST_ID	CUST_NAME	ADDRESS
2	100001	80001	Marion James	100 Main St.
2	100002	80002	Laura Jones	510 Broadway Ave.
1	100004	99001	Jon Freeman	555 6th Ave.

Cuando el Servicio de integración actualiza las filas en la memoria caché de búsqueda, utiliza los valores de la clave principal (PK_PRIMARYKEY) para las filas en la memoria caché y en la tabla de destino.

El Servicio de integración utiliza el ID de secuencia para generar una clave principal para el cliente que no encuentra en la memoria caché. El Servicio de integración inserta el valor de clave principal en la memoria caché de búsqueda y devuelve el valor al puerto de búsqueda/salida.

El Servicio de integración genera la salida de los valores de los puertos de entrada/salida que coinciden con los valores de entrada.

Nota: Si el valor de entrada es NULL y selecciona la propiedad Omitir NULL para el puerto de entrada asociado, el valor de entrada no es igual al valor de búsqueda o al valor del puerto de entrada/salida. Cuando seleccione la propiedad Omitir NULL, la memoria caché de búsqueda y la tabla de destino podrían desincronizarse si pasa valores nulos al destino. Deberá comprobar que no haya pasado valores nulos al destino.

Reemplazo de SQL y memoria caché de búsqueda dinámica

Puede añadir una cláusula WHERE en la consulta de búsqueda para filtrar los registros utilizados para generar la memoria caché y para realizar una búsqueda en la tabla de base de datos para una búsqueda no almacenada en caché. Sin embargo, el servicio de integración no utiliza la cláusula WHERE al insertar filas en una memoria caché dinámica.

Cuando añada una cláusula WHERE en una transformación de búsqueda mediante una memoria caché dinámica, conecte una transformación de filtro antes de la transformación de búsqueda para filtrar las filas que no desea insertar en la memoria caché o en la tabla de destino. Si no incluye la transformación de filtro, podría obtener resultados inconsistentes entre la memoria caché y la tabla de destino.

Por ejemplo, quiere configurar una transformación de búsqueda para realizar una búsqueda dinámica en la tabla de empleados, EMP, de las filas coincidentes por EMP_ID. Puede definir el siguiente reemplazo de SQL de búsqueda:

```
SELECT EMP_ID, EMP_STATUS FROM EMP ORDER BY EMP_ID, EMP_STATUS WHERE EMP_STATUS = 4
```

Cuando se ejecuta por primera vez la asignación de , el servicio de integración genera la memoria caché de búsqueda a partir de la tabla de destino según el reemplazo de SQL de búsqueda. Todas las filas de la memoria caché coinciden con la condición de la cláusula WHERE, EMP_STATUS = 4.

Por ejemplo, el servicio de integración lee una fila de origen que cumple la condición de búsqueda que ha especificado, pero el valor de EMP_STATUS es de 2. Aunque el destino podría tener la fila en la que EMP_STATUS es 2, el servicio de integración no encuentra la fila en la memoria caché debido al reemplazo de SQL. El servicio de integración inserta la fila en la memoria caché y pasa la fila para la tabla de destino. Cuando el servicio de integración inserta esta fila en la tabla de destino, podría obtener resultados

inconsistentes cuando la fila ya existe. Además, no todas las filas de la memoria caché coinciden con la condición de la cláusula WHERE en el reemplazo de SQL.

Para comprobar que solo se insertan filas en la memoria caché que coinciden con la cláusula WHERE, añada una transformación de filtro antes de la transformación de búsqueda y defina la condición de filtro como la condición de la cláusula WHERE en el reemplazo de SQL de búsqueda.

Para el ejemplo anterior, introduzca la siguiente condición de filtro en la transformación de filtro y la cláusula WHERE en el reemplazo de SQL:

```
EMP_STATUS = 4
```

Configuración de asignación de una memoria caché de búsqueda dinámica

Si utiliza una búsqueda con una memoria caché dinámica, deberá configurar la asignación para actualizar la memoria caché de búsqueda dinámica y escribir los filas modificadas en el destino.

Realice los pasos siguientes para configurar una asignación con una memoria caché de búsqueda dinámica:

Marque las filas de entrada de la transformación de búsqueda para su inserción o actualización.

De manera predeterminada, el tipo de fila de todas las filas de entrada es de inserción. Añada una transformación de estrategia de actualización antes de la transformación de búsqueda para especificar tipos de filas diferentes para las filas de entrada.

Especifique la forma en que el Servicio de integración administra las filas de entrada para la memoria caché dinámica.

Seleccione las opciones **Insertar o actualizar** o **Actualizar o insertar** para procesar las filas marcadas para su inserción o actualización.

Cree canales de asignación distintos para las filas que se van a insertar y a actualizar en el destino.

Añada un filtro o una transformación de enrutador después de la transformación de búsqueda para enrutar las filas de inserción y actualización en ramificaciones de asignación distintas. Utilice el valor de NewLookupRow para determinar la ramificación correspondiente para cada fila.

Configure el tipo de fila para las filas de salida de la transformación de búsqueda.

Añada una transformación de estrategia de actualización para marcar las filas como de inserción o de actualización.

Insertar o actualizar

Utilice la propiedad **Insertar o actualizar** para actualizar las filas existentes en la memoria caché de búsqueda dinámica cuando el tipo de fila sea de inserción.

Esta propiedad solo se aplica a las filas que se introducen en la transformación de búsqueda cuando el tipo de fila es de inserción. Cuando una fila de cualquier otro tipo de fila, como las de actualización, entra en la transformación de búsqueda, la propiedad **Insertar o actualizar** no tiene ningún efecto en la forma en que el servicio de integración administra la fila.

Cuando seleccione **Insertar o actualizar** y el tipo de fila que se introduce en la transformación de búsqueda es de inserción, el servicio de integración inserta la fila en la memoria caché si es nueva. Si la fila ya existe en la memoria caché de índice, pero la memoria caché de datos es distinta que la de la fila actual, el servicio de integración actualiza la fila en la memoria caché de datos.

Si no selecciona **Insertar o actualizar** y el tipo de fila que se introduce en la transformación de búsqueda es de inserción, el servicio de integración inserta la fila en la memoria caché si es nueva y no realiza ningún cambio en la memoria caché si la fila ya existe.

La siguiente tabla describe la forma en que el servicio de integración cambia la memoria caché de búsqueda cuando el tipo de fila de las filas que se introducen en la transformación de búsqueda es de inserción:

Opción Insertar o actualizar	La fila se encuentra en la memoria caché	La memoria caché de datos es diferente	Resultado de la memoria caché de búsqueda	Valor de NewLookupRow
Borrado - Solo inserción	Sí	-	No hay cambios	0
Borrado - Solo inserción	No	-	Insertar	1
Seleccionado	Sí	Sí	Actualizar	2 ¹
Seleccionado	Sí	No	No hay cambios	0
Seleccionado	No	-	Insertar	1

¹. Si selecciona **Omitir NULL** para todos los puertos de búsqueda que no están en la condición de búsqueda y si todos los puertos contienen valores nulos, el servicio de integración no cambia la memoria caché y el valor de **NewLookupRow** es igual a 0.

Actualizar o insertar

Utilice la propiedad **Actualizar o insertar** para insertar filas nuevas en la memoria caché de búsqueda dinámica cuando el tipo de fila sea de actualización.

Puede seleccionar la propiedad **Actualizar o insertar** en la transformación de búsqueda. Esta propiedad solo se aplica a las filas que se introducen en la transformación de búsqueda cuando el tipo de fila es de actualización. Cuando una fila de cualquier otro tipo de fila, como las de inserción, entra en la transformación de búsqueda, esta propiedad no tiene ningún efecto en la forma en que el servicio de integración administra la fila.

Cuando seleccione esta propiedad y el tipo de fila que se introduce en la transformación de búsqueda sea de actualización, el servicio de integración actualizará la fila en la memoria caché si la fila ya existe en la memoria caché de índice y los datos de la memoria caché son diferentes de los de la fila existente. El servicio de integración inserta la fila en la memoria caché si es nueva.

Si no selecciona esta propiedad y el tipo de fila que se introduce en la transformación de búsqueda es de actualización, el servicio de integración actualiza la fila en la memoria caché si ya existe y no realiza ningún cambio en la memoria caché si la fila es nueva.

Si selecciona **Omitir NULL** para todos los puertos de búsqueda que no están en la condición de búsqueda y si todos los puertos contienen valores nulos, el servicio de integración no cambia la memoria caché y el valor de **NewLookupRow** es igual a 0.

La siguiente tabla describe la forma en que el servicio de integración cambia la memoria caché de búsqueda cuando el tipo de fila de las filas que se introducen en la transformación de búsqueda es de actualización:

Opción Actualizar o insertar	La fila se encuentra en la memoria caché	La memoria caché de datos es diferente	Resultado de la memoria caché de búsqueda	Valor de NewLookupRow
Borrado (Solo actualización)	Sí	Sí	Actualizar	2
Borrado (Solo actualización)	Sí	No	No hay cambios	0
Borrado (Solo actualización)	No	-	No hay cambios	0
Seleccionado	Sí	Sí	Actualizar	2
Seleccionado	Sí	No	No hay cambios	0
Seleccionado	No	-	Insertar	1

Memoria caché de búsqueda dinámica y destino de sincronización

Configure las transformaciones de nivel inferior para asegurarse de que la memoria caché de búsqueda dinámica y el destino estén sincronizados.

Cuando se utiliza una memoria caché de búsqueda dinámica, el Servicio de integración escribe en la memoria caché de búsqueda antes de escribir en la tabla de destino. La memoria caché de búsqueda y la tabla de destino pueden desincronizarse si el Servicio de integración no escribe los datos en el destino. Por ejemplo, la base de datos de destino podría rechazar los datos.

Tenga en cuenta las siguientes directrices para mantener la memoria caché de búsqueda sincronizada con la tabla de búsqueda:

- Utilice una transformación de enrutador para pasar las filas al destino de la memoria caché cuando el valor de NewLookupRow sea igual a uno o dos.
- Utilice la transformación de enrutador para eliminar filas cuando el valor de NewLookupRow sea igual a cero. O bien, puede enviar las filas a un destino diferente.
- Utilice transformaciones de estrategia de actualización después de la transformación de búsqueda para marcar las filas para inserción o actualización en el destino.
- Compruebe que los valores que la transformación de búsqueda envía al destino son los mismos que los que escribe el Servicio de integración en la memoria caché de búsqueda. Cuando elija enviar valores nuevos en la actualización, solo conecte los puertos de búsqueda/salida a la tabla de destino, en lugar de los puertos de salida de . Cuando elija enviar valores antiguos en la actualización, añada una transformación de expresión después de la transformación de búsqueda y antes de la transformación de enrutador. Añada puertos de salida en la transformación de expresión para cada puerto de la tabla de destino y cree expresiones para asegurarse de que no se envíen valores de entrada nulos al destino.
- Seleccione Insertar y actualizar como Actualizar cuando defina las opciones de la tabla de destino de la estrategia de actualización. Esto garantiza que el Servicio de integración actualice las filas marcadas para actualización e inserte filas marcadas para inserción.

Actualizaciones de memoria caché de búsqueda dinámica condicional

Puede actualizar una memoria caché de búsqueda dinámica según los resultados de una expresión booleana. El Servicio de integración actualiza la memoria caché cuando la expresión es verdadera.

Por ejemplo, puede tener un número de producto, la cantidad disponible y una columna de marca de tiempo en una tabla de destino. Deberá actualizar la cantidad disponible con los últimos valores de origen. Podrá actualizar la cantidad disponible cuando los datos de origen tengan una marca de tiempo mayor que la marca de tiempo de la memoria caché dinámica. Cree una expresión en la transformación de búsqueda similar a la siguiente expresión:

```
lookup_timestamp < input_timestamp
```

La expresión puede incluir los puertos de búsqueda y de entrada. Puede acceder a variables integradas, de asignaciones y de parámetros. Puede incluir funciones definidas por el usuario y hacer referencia a transformaciones no conectadas.

La expresión devuelve verdadero, falso o NULL. Si el resultado de la expresión es NULL, la expresión es falsa. El Servicio de integración no actualiza la memoria caché. Puede añadir una comprobación para los valores NULL de la expresión si tiene que cambiar el resultado de la expresión a verdadero. El valor predeterminado de la expresión es verdadero.

Procesamiento de la memoria caché de búsqueda dinámica condicional

Puede crear una condición que determine si el servicio de integración actualiza la memoria caché de búsqueda dinámica. Si la condición es falsa o NULL, el servicio de integración no actualiza la memoria caché de búsqueda dinámica.

Si la condición es falsa o NULL, independientemente de las propiedades de la transformación de búsqueda, el valor de NewLookupRow es 0 y los servicios de integración no actualizan la memoria caché de búsqueda dinámica con cualquier inserción o actualización de fila.

Si el valor de NewLookupRow es 2, es decir, la fila está marcada para actualización, y habilita Actualizar o insertar, el servicio de integración realiza una de las siguientes acciones:

- Si la fila existe en la memoria caché, el servicio de integración actualiza la fila existente en la memoria caché.
- Si la fila no existe en la memoria caché, el servicio de integración inserta la nueva fila en la memoria caché.

Si el valor de NewLookupRow es 1, es decir, la fila está marcada para inserción, y habilita Insertar o actualizar, el servicio de integración realiza una de las siguientes acciones:

- Si la fila no existe en la memoria caché, el servicio de integración inserta la nueva fila en la memoria caché.
- Si la fila existe en la memoria caché, el servicio de integración actualiza la fila existente en la memoria caché.

Configurar una memoria caché de búsqueda dinámica condicional

Puede configurar una expresión que determine si el Servicio de integración actualiza una memoria caché de búsqueda dinámica.

1. Cree la transformación de búsqueda.
2. En la ficha **Tiempo de ejecución** de la vista **Propiedades**, seleccione **Almacenamiento en memoria caché de búsqueda habilitado**.
3. En la ficha **Avanzadas** de la vista **Propiedades**, seleccione **Memoria caché de búsqueda dinámica**.
4. Para especificar la condición, haga clic en la flecha abajo de la propiedad **Actualizar condición de memoria caché dinámica**.
Aparece el Editor de expresiones.
5. Defina una condición de expresión.
Puede seleccionar puertos de entrada, puertos de búsqueda y funciones para la expresión.
6. Haga clic en **Validar** para comprobar que la expresión es válida.
7. Haga clic en **Aceptar**.
8. Si procede, configure las otras propiedades avanzadas que se apliquen a la memoria caché de búsqueda dinámica.

Actualización de memoria caché dinámica con resultados de expresión

La transformación de búsqueda puede actualizar los valores de la memoria caché de búsqueda dinámica con los resultados de una expresión.

Por ejemplo, un destino de una tabla de productos tiene una columna numérica que contiene un recuento de pedidos. Cada vez que la transformación de búsqueda recibe un pedido del producto, actualiza la memoria caché dinámica `order_count` con los resultados de la siguiente expresión:

```
order_count = order_count + 1
```

La transformación de búsqueda devuelve `order_count`.

Puede configurar la forma en que el Servicio de integración administra el caso en el que la expresión tiene como resultado NULL.

Valores de expresión nulos

La expresión devuelve NULL si uno de los valores en la expresión es nulo. Sin embargo, puede configurar una expresión para que devuelva un valor no nulo.

Si la expresión hace referencia a un puerto de búsqueda, pero los datos de origen son nuevos, el puerto de búsqueda contiene un valor predeterminado. El valor predeterminado puede ser NULL. Puede configurar una expresión `IsNull` para comprobar si hay valores nulos.

Por ejemplo, la siguiente expresión comprueba si `lookup_column` es NULL:

```
iif (isnull(lookup_column), input_port, user_expression)
```

Si la columna es nula, se devuelve el valor de `input_port`. De lo contrario, se devuelve el valor de la expresión.

Procesamiento de expresiones

El Servicio de integración puede insertar y actualizar las filas en la memoria caché de búsqueda dinámica en función de los resultados de la expresión. Los resultados de la expresión pueden variar según si el valor del puerto de búsqueda es NULL y está incluido en la expresión.

Cuando se habilita Insertar o actualizar, el servicio de integración inserta una fila con el resultado de la expresión si los datos no están en la memoria caché. El valor del puerto de búsqueda es NULL cuando los datos no existen en la memoria caché. Si la expresión hace referencia al valor del puerto de búsqueda, el servicio de integración sustituye el valor del puerto predeterminado en la expresión. Cuando se habilita Insertar o actualizar y los datos ya se encuentran en la memoria caché, el servicio de integración actualiza la memoria caché con el resultado de la expresión.

Cuando se habilita Actualizar o insertar, el servicio de integración actualiza la memoria caché con el resultado de la expresión si los datos ya se encuentran en la memoria caché. Cuando los datos no existen en la memoria caché, el servicio de integración inserta una fila que contiene el resultado de la expresión. Si la expresión hace referencia a un valor del puerto de búsqueda, el servicio de integración sustituye el valor del puerto predeterminado en la expresión.

Configurar una expresión para actualizaciones de memoria caché dinámica

Puede configurar una expresión para una actualización de búsqueda de memoria caché dinámica.

Deberá habilitar la transformación de búsqueda para realizar búsquedas dinámicas para poder crear una expresión condicional.

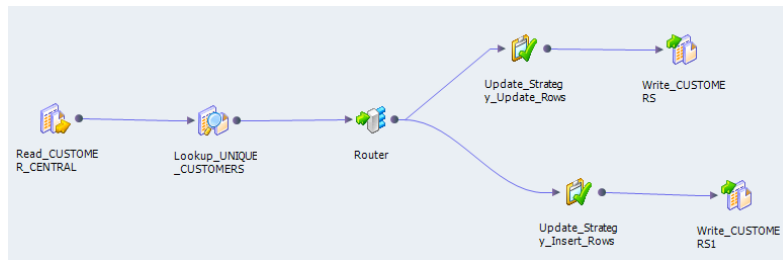
1. Cree la transformación de búsqueda.
2. En la ficha **Tiempo de ejecución** de la vista **Propiedades**, seleccione **Almacenamiento en memoria caché de búsqueda habilitado**.
3. En la ficha **Avanzadas** de la vista **Propiedades**, seleccione **Memoria caché de búsqueda dinámica**.
4. Si procede, configure las otras propiedades avanzadas que se apliquen a la memoria caché de búsqueda dinámica.
5. Para crear una expresión, seleccione la ficha **Columnas** en la vista **Propiedades**.
6. Haga clic en la flecha desplegable de la columna **Puerto asociado** del puerto de búsqueda que quiera actualizar.
7. Seleccione **Expresión asociada** en la lista desplegable y haga clic en **Introducir**.
Aparece el Editor de expresiones.
8. Defina la expresión.
Puede seleccionar puertos de entrada, puertos de búsqueda y funciones para la expresión. El valor de devolución de la expresión debe coincidir con el tipo de datos del puerto de búsqueda.
9. Haga clic en **Validar** para comprobar que la expresión es válida.
10. Haga clic en **Aceptar**.

Ejemplo de memoria caché de búsqueda dinámica

Puede usar una memoria caché de búsqueda dinámica para insertar y actualizar filas en el destino. Cuando utilice una memoria caché de búsqueda dinámica, podrá insertar y actualizar las mismas filas en la memoria caché que se inserten y se actualicen en el destino.

Por ejemplo, debe actualizar una tabla que contiene los datos de los clientes. Los datos de origen contienen filas de datos de clientes para insertar o actualizar en el destino. Cree una memoria caché dinámica que represente el destino. Configure una transformación de búsqueda para buscar clientes en la memoria caché dinámica.

La siguiente imagen muestra una asignación que tiene una transformación de búsqueda con una memoria caché de búsqueda dinámica:



La transformación de enrutador se divide en dos ramificaciones. La transformación de enrutador pasa las filas para inserción por una ramificación y las filas para actualización por la otra ramificación. Cada ramificación contiene una transformación de estrategia de actualización que escribe las filas en el destino. Las dos ramificaciones contienen el mismo destino.

Cuando se inicia la asignación, el Servicio de integración genera la memoria caché de búsqueda a partir de la tabla de destino. Cuando el Servicio de integración lee una fila que no está en la memoria caché de búsqueda, inserta la fila en la memoria caché.

La transformación de búsqueda devuelve cada fila a una transformación de enrutador. Dependiendo de si la fila está marcada para inserción o actualización, la transformación de enrutador dirige la fila a una de las transformaciones de estrategia de actualización. La transformación de enrutador determina si las filas están marcadas para inserción o actualización según la propiedad NewLookupRow. Las transformaciones de estrategia de actualización marcan cada fila como de inserción o actualización antes de pasarlas al destino.

La tabla de destino de clientes cambia a medida que se ejecuta la asignación. El Servicio de integración inserta filas nuevas y actualiza las filas existentes en la memoria caché de búsqueda. El Servicio de integración mantiene sincronizadas la memoria caché de búsqueda y la tabla de destino de clientes.

Para generar claves para el destino, utilice Sequence-ID en el puerto asociado. El Servicio de integración utiliza el ID de secuencia como clave principal para cada nueva fila insertada en la tabla de destino.

El rendimiento de la sesión aumenta cuando se utiliza una memoria caché de búsqueda dinámica porque la memoria caché se genera desde la base de datos solo una vez.

Reglas y directrices de las memorias caché de búsqueda dinámica

Tenga en cuenta las siguientes directrices cuando utilice una memoria caché de búsqueda dinámica:

- Deberá establecer la propiedad **En coincidencias múltiples** como Informe de error cuando utilice una memoria caché de búsqueda dinámica. Para restablecer la propiedad, cambie la búsqueda dinámica a una búsqueda estática, cambie la propiedad y después cambie la búsqueda estática a una búsqueda dinámica.
- No puede compartir la memoria caché entre una transformación de búsqueda dinámica y una transformación de búsqueda estática en el mismo grupo de orden de carga de destino.
- Puede habilitar la memoria caché de una búsqueda dinámica para una búsqueda relacional o de archivo sin formato.
- La transformación de búsqueda debe ser una transformación conectada.
- Puede utilizar una memoria caché persistente o no persistente.
- Si la memoria caché dinámica no es persistente, el Servicio de integración siempre se vuelve a generar la memoria caché desde la base de datos, aunque no haya habilitado **Volver a almacenar en memoria caché desde el origen de búsqueda**.
- Solamente puede crear una condición de búsqueda de igualdad. No se puede buscar un intervalo de datos en una memoria caché dinámica.
- Deberá asociar cada puerto de búsqueda que no esté en la condición de búsqueda con un puerto de entrada, un ID de secuencia o una expresión asociada.
- Utilice una transformación de enrutador para pasar las filas al destino de la memoria caché cuando el valor de NewLookupRow sea igual a uno o dos.
- Utilice la transformación de enrutador para eliminar filas cuando el valor de NewLookupRow sea igual a cero. O bien, puede enviar las filas a un destino diferente.
- Compruebe que el Servicio de integración envía los mismos valores al destino que escribe en la memoria caché de búsqueda. Cuando elija enviar valores nuevos en la actualización, solo conecte los puertos de búsqueda/salida a la tabla de destino, en lugar de los puertos de entrada/salida. Cuando elija enviar valores antiguos en la actualización, añada una transformación de expresión después de la transformación de búsqueda y antes de la transformación de enrutador. Añada puertos de salida en la transformación de expresión para cada puerto de la tabla de destino y cree expresiones para asegurarse de que no se envíen valores de entrada nulos al destino.
- Cuando utilice un reemplazo de SQL de búsqueda, asigne las columnas correctas a los destinos adecuados para la búsqueda.
- Cuando añada una cláusula WHERE al reemplazo de SQL de búsqueda, utilice una transformación de filtro antes de la transformación de búsqueda. Esto garantiza que el Servicio de integración inserte filas en la memoria caché dinámica y en la tabla de destino que coincidan con la cláusula WHERE.
- Cuando configure una transformación de búsqueda reutilizable para utilizar una memoria caché dinámica, no podrá editar la condición ni deshabilitar la propiedad **Memoria caché de búsqueda dinámica** en una asignación.
- Utilice transformaciones de estrategia de actualización después de la transformación de búsqueda para marcar las filas para inserción o actualización en el destino.
- Utilice una transformación de estrategia de actualización antes de la transformación de búsqueda para definir algunas o todas las filas como de actualización si desea utilizar la propiedad Actualizar o insertar en la transformación de búsqueda.

CAPÍTULO 28

Transformación de macro

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de macro, 473](#)
- [Instrucción macro, 473](#)
- [Parámetros de la transformación de macro, 474](#)
- [Transformaciones de macro en asignaciones dinámicas, 474](#)
- [Ejemplo de la transformación de macro, 475](#)
- [Transformación de macro en un entorno no nativo, 476](#)

Resumen de la transformación de macro

La transformación de macro es una transformación pasiva que habilita la función dinámica en cualquier lógica de transformación.

Utilice la transformación de macro en una asignación dinámica que pueda admitir cambios en orígenes y destinos en tiempo de ejecución. La transformación de macro propaga los cambios realizados en un mapplet al que hace referencia la transformación de macro, lo que permite al mapplet actuar como mapplet dinámico. La transformación de macro se asocia siempre con un mapplet, llamado "instrucción macro". El mapplet de instrucción macro puede contener transformaciones que no admiten la función dinámica.

La transformación de macro no se puede configurar como una transformación reutilizable.

Instrucción macro

La instrucción macro es el mapplet al que hace referencia la transformación de macro. Cada transformación de macro debe contener una instrucción macro.

Developer tool le pide que seleccione un mapplet como la instrucción macro cuando crea una transformación de macro. Una vez creada la transformación, no puede seleccionar un mapplet diferente como instrucción macro.

Tenga en cuenta las siguientes reglas y directrices relativas al mapplet de instrucción macro:

- El mapplet debe ser un mapplet de nivel intermedio.
- El mapplet debe tener exactamente una transformación de entrada.

- El mapplet no puede utilizar transformaciones activas.

Parámetros de la transformación de macro

Cuando añade una transformación de macro a una asignación, puede utilizar parámetros del mapplet de instrucción macro como parámetros de asignación, tal como lo haría si añadiera un mapplet a una asignación. Un parámetro de asignación representa un valor constante que se puede cambiar entre ejecuciones de asignaciones.

Al añadir un mapplet con parámetros de mapplet como una instrucción macro, puede establecer los valores de instancia de los parámetros. El valor de instancia de un parámetro de mapplet es el valor del parámetro para una determinada asignación. Establezca el valor de instancia en la ficha **Parámetros** de la vista **Propiedades** de la transformación de macro.

Para obtener más información, vea "Parámetros de mapplet en asignaciones" en la *Guía de asignaciones de Informatica® Developer*.

Transformaciones de macro en asignaciones dinámicas

Puede añadir una transformación de macro a una asignación dinámica. La transformación contiene puertos dinámicos que se pueden vincular con orígenes y destinos dinámicos.

Cuando se crea una transformación de macro y se asigna un mapplet de instrucción macro a ella, se crean puertos de entrada y salida dinámicos en función de los grupos de entrada y salida de la instrucción macro.

De forma predeterminada, los puertos de salida tienen el mismo tipo, precisión, escala y descripción que los puertos de entrada. Puede configurar los puertos de salida editando el puerto base asociado a cada puerto de salida en las propiedades de **Puertos**.

El número de puertos generados creados por un puerto dinámico en la transformación de macro debe ser uno o igual en número a los puertos generados por los otros puertos dinámicos de la transformación, excepto los puertos de transferencia. Los puertos de transferencia de la transformación de macro pueden tener cualquier número de puertos generados.

Por ejemplo, una transformación de macro tiene tres puertos dinámicos `transformed_input1`, `transformed_input2` y `passthrough_input`. Si `transformed_input1` genera tres puertos, `transformed_input2` debe generar uno o tres puertos. `Passthrough_input` puede generar cualquier número de puertos.

Ejemplo de la transformación de macro

Suponga que es el gestor de datos de una compañía de seguros sanitarios y necesita cifrar información personal sobre los titulares de las pólizas de seguro. El esquema del origen de datos puede variar. Por lo tanto, necesita utilizar una asignación dinámica.

Necesita una asignación dinámica que lea datos de un origen, que cifre todos los datos de cadena utilizando una transformación de enmascaramiento de datos y que escriba los datos modificados en un destino. Debido a que la transformación de enmascaramiento de datos no es compatible con la función dinámica, necesita utilizar una transformación de macro para habilitar la función dinámica en un mapplet que contenga la transformación de enmascaramiento de datos. La transformación de macro permite que la lógica de transformación del mapplet actúe como lógica de asignación dinámica al aceptar entradas con cualquier número de puertos de cadena y pasar cada puerto de entrada de cadena a través de una copia del mapplet de enmascaramiento de datos.

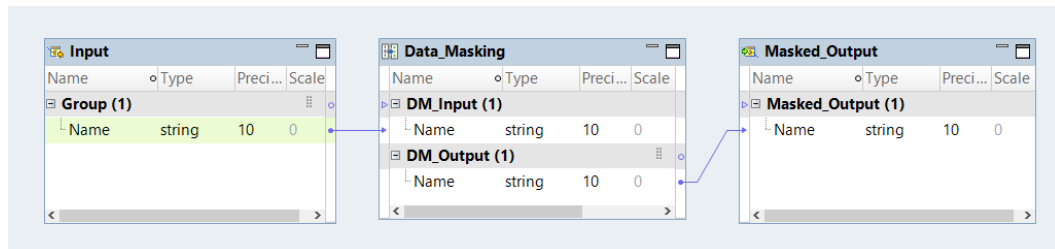
Para utilizar la transformación de macro, lleve a cabo los siguientes pasos:

1. Configure el mapplet.
2. Configure la asignación dinámica con la transformación de macro.
3. Ejecute la asignación.

Configurar el mapplet

Antes de configurar la transformación de macro, cree un mapplet que realice el cifrado de datos. El mapplet incluye una transformación de entrada, una transformación de enmascaramiento de datos y una transformación de salida. La transformación de entrada tiene un puerto de tipo cadena que se conecta con el grupo de entrada de la transformación de enmascaramiento de datos. En la transformación de enmascaramiento de datos, establezca un tipo de enmascaramiento para el puerto y, luego, conecte el puerto de salida a la transformación de salida.

La siguiente imagen muestra el mapplet:



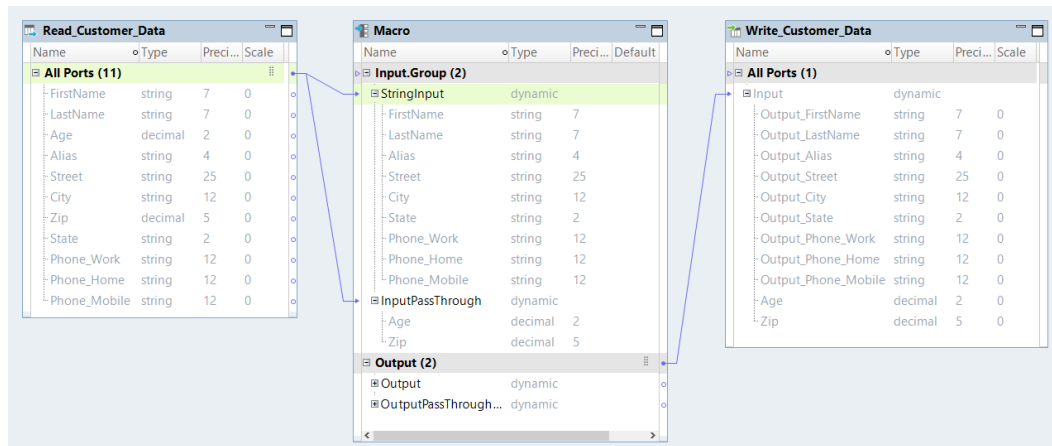
Configurar la asignación

Tras configurar el mapplet, hay que configurar la asignación dinámica. Cree una asignación que lea de un origen con parámetros y configúrela para obtener los nombres de columna en tiempo de ejecución. Luego, añada una transformación de macro a la asignación dinámica y seleccione el mapplet de enmascaramiento de datos como la instrucción macro.

Conecte el grupo de puertos de transformación de lectura al puerto de entrada *StringInput* y al puerto de transferencia de entrada *InputPassThrough*. Configure la regla de entrada de forma que *StringInput* incluya solo puertos de tipo cadena.

Por último, añada una transformación de escritura a la asignación y configúrela para definir los puertos según el flujo de asignación. Conecte el grupo de salida de la transformación de macro al puerto de entrada dinámica de la transformación de escritura.

La siguiente imagen muestra la asignación dinámica:



Ejecutar la asignación

Cuando la asignación se ejecuta, determina los puertos a partir del origen de datos. La transformación de macro crea una copia del mapplet por cada puerto de entrada de cadena y pasa cada puerto de cadena a través de la copia del mapplet. La transformación de enmascaramiento de datos cifra los datos de cadena y devuelve la salida a la asignación a través de la transformación de macro. La asignación escribe en el destino tanto los datos de cadena cifrados como los datos decimales sin modificar.

Transformación de macro en un entorno no nativo

El procesamiento de transformaciones de macro en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 29

Transformación de coincidencia

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de coincidencia, 477](#)
- [Análisis de coincidencia, 478](#)
- [Cálculos de puntuación de coincidencia, 482](#)
- [Análisis de datos principales, 485](#)
- [Datos del análisis de coincidencia de identidad y del índice persistente, 487](#)
- [Rendimiento de asignación de coincidencias, 488](#)
- [Rendimiento de coincidencia en el análisis de identidad, 490](#)
- [Vistas de transformación de coincidencia, 493](#)
- [Puertos de transformación de coincidencia, 494](#)
- [Maplets de coincidencia, 499](#)
- [Configuración de una operación de análisis de coincidencia, 501](#)
- [Transformación de coincidencia en un entorno no nativo, 502](#)

Resumen de la transformación de coincidencia

La transformación de coincidencia es una transformación activa que analiza los niveles de similitud entre registros. Utilice la transformación de coincidencia para hallar registros que contengan información duplicada en un conjunto de datos o entre dos conjuntos de datos.

La transformación de coincidencia analiza los valores en un puerto de entrada y genera un conjunto de puntuaciones numéricas que representan los grados de similitud entre los valores. Puede seleccionar varios puertos para averiguar los niveles generales entre los registros de entrada. Debe especificar una puntuación mínima como valor de umbral para identificar los registros que probablemente contengan información duplicada.

Puede utilizar la transformación de coincidencia en los siguientes proyectos de datos:

- Administración de relaciones con clientes. Por ejemplo, un almacén diseña una campaña de correo y debe comprobar la base de datos de clientes para buscar registros de clientes duplicados.
- Fusiones y compras. Por ejemplo, un banco compra a otro banco de la misma zona y los dos bancos tienen clientes común.
- Iniciativas de cumplimiento legal. Por ejemplo, una empresa opera de acuerdo con normas gubernamentales o industriales que insisten en que todos los sistemas de datos deben estar desprovistos de datos duplicados.

- Administración de riesgos financieros. Por ejemplo, un banco puede desear buscar relaciones entre titulares de cuentas.
- Administración de datos principales. Por ejemplo, una cadena minorista tiene una base de datos principal de registros de clientes y cada comercio minorista de la cadena envía registros a la base de datos principal de forma periódica.
- Cualquier proyecto que deba identificar registros duplicados en un conjunto de datos.

Análisis de coincidencia

Puede definir diferentes tipos de análisis de duplicados en la transformación de coincidencia. Las operaciones de análisis de duplicados que defina dependen del número de orígenes de datos en la asignación y del tipo de información que los orígenes contienen.

Tenga en cuenta los siguientes factores cuando configure una transformación de coincidencia:

- Puede seleccionar una única columna de un conjunto de datos o puede seleccionar varias columnas.
- Puede analizar columnas de un único origen de datos o puede analizar dos orígenes de datos.
- Puede configurar la transformación de coincidencia para analizar los datos sin formato en los campos del puerto de entrada, o puede configurar la transformación para analizar la información de las identidades en los datos.
- Puede configurar la transformación de coincidencia para escribir los distintos tipos de salida. El tipo de salida que seleccione determina el número de registros que la transformación escribe y el orden de los registros.
- Para aumentar el rendimiento, ordene los registros de entrada en grupos antes de realizar el análisis de coincidencia.

Análisis de columnas

Al configurar una transformación de coincidencia, seleccione una o más columnas para el análisis.

La transformación de coincidencia analiza las columnas en parejas. Cuando seleccione una sola columna para el análisis, la transformación crea una copia temporal de la columna y compara la columna de origen con la columna temporal. Cuando seleccione dos columnas para el análisis, la transformación compara los valores entre dos columnas que seleccione. La transformación compara cada valor de una columna con todos los valores de la otra columna. La transformación devuelve una puntuación de coincidencia para cada par de valores que analiza.

Las columnas que analizar se seleccionan al configurar una estrategia en la transformación de coincidencia. La estrategia especifica las columnas que analizar y el algoritmo que aplicar a las columnas. El algoritmo calcula los niveles de similitud entre cada par de valores. Los diferentes algoritmos de la transformación utilizan distintos criterios para medir los niveles de similitud entre los valores. Puede definir varias estrategias de una transformación, así como asignar distintas columnas a cada estrategia.

Ejemplo de análisis de columna

Desea comparar los valores en una columna con datos de apellidos. Cree una asignación que incluya un origen de datos y una transformación de coincidencia. Conecte el puerto *Apellido* a la transformación de coincidencia. La transformación crea una copia temporal de los datos en el puerto *Apellido* cuando la asignación se ejecuta.

La siguiente imagen muestra un fragmento de los datos de apellido:

	A	B
1	Surname	Surname_1
2	Annan	Annan
3	Baker	Baker
4	Barker	Barker
5	Edwards	Edwards
6	Parker	Parker
7	Smith	Smith
8	Smith	Smith
9	Zhang	Zhang

La asignación genera un conjunto de puntuaciones de coincidencia que indican que los siguientes valores podrían estar duplicados:

- Baker, Barker
- Barker, Parker
- Smith, Smith

Cuando revise los datos, puede decidir que *Baker*, *Barker* y *Parker* no son valores duplicados. Puede decidir que *Smith* y *Smith* son valores duplicados.

Análisis de un solo origen y análisis de origen dual

Puede configurar la transformación de coincidencia para analizar datos de uno o dos orígenes de datos. Seleccione los puertos de cada origen de datos al definir una estrategia en la transformación.

Cuando configure la transformación para realizar un análisis de un solo origen, seleccione uno o varios puertos de un único conjunto de datos. Cuando configure la transformación para realizar un análisis de origen dual, seleccione uno o varios puertos de cada conjunto de datos. Seleccione los puertos en pares. Para cada par de puertos que seleccione, la transformación compara todos los valores de un puerto con todos los valores del otro puerto. Si realiza análisis de un solo origen en datos de una sola columna, la transformación crea una copia temporal del puerto que seleccione.

Nota: Al realizar un análisis de coincidencia de identidad, puede comparar un origen de datos con un índice de datos de identidad persistente creado en una asignación anterior. Utilice las opciones de **Tipo de coincidencia** para especificar el análisis de identidad con un índice persistente.

Análisis de coincidencias de campos y análisis de coincidencia de identidades

Puede configurar una transformación de coincidencia para realizar un análisis de coincidencia de campos o un análisis de coincidencia de identidades.

En el análisis de coincidencia de campos, la transformación de coincidencia analiza los datos de origen que entran a la transformación. Puede realizar análisis de coincidencia de campos en cualquier tipo de datos. En el análisis de coincidencia de identidad, la transformación de coincidencia genera un índice de valores de datos alternativos a partir de los datos de entrada y analiza los datos del índice. Configure la transformación de coincidencia del análisis de coincidencia de identidad cuando los puertos de entrada contengan datos de identidad. Una identidad es un grupo de valores de datos que identifica a una persona o una organización.

Un conjunto de datos puede representar una única identidad de diferentes maneras. Por ejemplo, los siguientes valores de datos representan el nombre John Smith:

- John Smith
- Smith, John
- jsmith@email.com
- SMITHJMR

La transformación de coincidencia lee los datos de identidad en un registro y calcula las posibles versiones alternativas de la identidad. La transformación crea un índice que incluye las versiones actuales y las versiones alternativas de identidades. La transformación de coincidencia analiza los valores de índice y no los valores de los registros de entrada.

Archivos de llenado de identidad

Las operaciones de coincidencia de identidades leen archivos de datos de referencia llamados poblaciones. Los archivos de poblaciones definen las posibles variaciones en los datos de identidad. Estos archivos no se instalan junto con las aplicaciones de Informática. Puede adquirir y descargar los archivos de datos de población de Informática.

Instale los archivos en una ubicación a la que pueda acceder el servicio de administración de contenido. Utilice Informática Administrator para definir la ubicación del servicio de administración de contenido.

Grupos en el análisis de coincidencia

Una asignación de análisis de coincidencia puede tardar mucho tiempo en ejecutarse debido al número de comparaciones de datos que debe realizar la transformación. El número de comparaciones está relacionado con el número de valores de datos en los puertos que seleccione.

La siguiente tabla muestra el número de cálculos que una asignación realiza para diferentes números de valores de datos en un solo puerto:

Número de valores de datos	Número de comparaciones
10.000	50 millones
100.000	5.000 millones
1 millón	500.000 millones

Para reducir el tiempo que la asignación necesaria para ejecutarse, asigne los registros de datos de entrada a grupos. Un grupo es un conjunto de registros que contienen valores idénticos en un puerto que especifique. Al realizar el análisis de coincidencia en datos agrupados, la transformación de coincidencia analiza los registros dentro de cada grupo. La transformación no compara los registros de un grupo con los registros de otro grupo. Los grupos reducen el número total de comparaciones que la transformación debe realizar sin pérdida de precisión en el análisis de asignación.

Considere las siguientes reglas e instrucciones al organizar los datos en grupos:

- El puerto en el que agrupe los datos es el puerto de clave de grupo. Un puerto de clave de grupo debe contener un rango de valores duplicados, como un nombre de ciudad o de estado en un conjunto de datos de dirección. Si los datos de asignación no contienen un puerto de clave de grupo utilizable, utilice el generador de claves para crear el puerto a partir de los datos de asignación actuales. Conecte el puerto de salida de clave de grupo de la transformación del generador de claves a la transformación de coincidencia.
También puede utilizar la transformación del generador de claves para agregar identificadores de secuencia a los datos de asignación.
- Las operaciones de coincidencia de campos deben especificar un grupo de clave de puerto. Si configura la transformación de coincidencia para análisis de identidad, no seleccione un puerto de clave de grupo. El análisis de identidad genera claves de grupo para los datos de índice de identidad.
- No especifique un puerto de clave de grupo que vaya a utilizar en el análisis de coincidencia.
- Cuando cree grupos, debe comprobar que los grupos tienen un tamaño válido. Si los grupos son demasiado pequeños, es posible que el análisis de coincidencias no encuentre todos los datos duplicados del conjunto de datos. Si los grupos son demasiado grandes, el análisis de coincidencia podría devolver falsos duplicados. Seleccione claves de grupo que creen un tamaño medio de grupo de 10.000 registros.
- Los grupos no cambian la posición de los registros en el conjunto de datos de la asignación.

Pares de coincidencia y clústeres

La transformación de coincidencia puede leer y escribir distintos números de filas de entrada y filas de salida, y puede cambiar la secuencia de las filas de salida. Determina el formato de salida para los resultados del análisis de coincidencia.

La transformación puede escribir filas en los formatos siguientes:

Pares de coincidencia

La transformación escribe una fila por cada par de registros que coincide con una puntuación que cumple el umbral de coincidencia. La transformación escribe cada par de registros en una única fila.

Dado que un registro puede coincidir con más de un registro, es posible que aparezca en más de una fila de salida.

Mejor coincidencia

La transformación escribe una fila por cada registro en un conjunto de datos y añade a esa misma fila el registro más parecido de otro conjunto de datos.

Clústeres

La transformación asigna los registros de salida a clústeres según los niveles de parecido entre los registros. Un clúster es un conjunto de registros en el que cada registro coincide al menos con otro registro que tenga una puntuación que cumpla el umbral de coincidencia. La transformación escribe cada registro en una única fila.

Cada registro en un clúster debe coincidir con al menos un registro del clúster. En consecuencia, un clúster puede contener pares de registros que no coincidan entre sí. Un clúster puede contener un único registro si el registro no coincide con ningún otro registro.

Nota: La opción Clústeres del análisis de campos se corresponde con la opción Clústeres: coincidir con todos del análisis de identidad. La opción Clústeres: mejor coincidencia del análisis de identidad combina cálculos de clústeres y cálculos de pares de coincidencia.

Configure las opciones de salida en la vista **Salida de coincidencia** de la transformación.

TEMAS RELACIONADOS

- [“Opciones de salida del clúster” en la página 483](#)

Cálculos de puntuación de coincidencia

La puntuación de coincidencia es un valor numérico que representa el grado de similitud entre dos valores de columna. Un algoritmo calcula una puntuación de coincidencia como un valor decimal dentro del intervalo de 0 a 1. Un algoritmo asigna una puntuación de 1 cuando dos valores de columna son idénticos.

Cuando seleccione varios pares de columnas para el análisis, la transformación calcula una puntuación media según las puntuaciones de las columnas seleccionadas. De forma predeterminada, la transformación asigna la misma ponderación a las puntuaciones de cada par de columnas. La transformación no infiere en la importancia relativa de los datos de columnas del conjunto de datos.

Puede editar los valores ponderados que utiliza la transformación para calcular la puntuación de coincidencia. Edite los valores ponderados cuando desee asignar una prioridad mayor o menor a las columnas del conjunto de datos.

También puede establecer las puntuaciones que la transformación aplica cuando encuentra un valor nulo en una columna. De forma predeterminada, la transformación trata los valores nulos como errores de datos y asigna una puntuación de coincidencia baja a cualquier par de valores que contengan un valor nulo.

Nota: El algoritmo que seleccione determina la puntuación de coincidencia entre dos valores. El algoritmo genera una única puntuación para los dos valores. Las puntuaciones de coincidencia no dependen del tipo de salida de coincidencia o del tipo de método de puntuación que haya seleccionado.

Puntuaciones ponderadas

Cuando seleccione varias columnas para el análisis de coincidencias, la transformación calcula una puntuación media para cada registro según las puntuaciones de las columnas. La puntuación media incluye cualquier valor ponderado que aplique a los algoritmos de comparación para cada columna.

De forma predeterminada, todos los algoritmos utilizan un valor ponderado de 0,5. Puede aumentar este valor si es más probable que las columnas seleccionadas contengan información duplicada. Puede reducir el valor ponderado si es menos probable que los valores duplicados de las columnas seleccionadas indiquen información duplicada original entre registros. La transformación de coincidencia utiliza la puntuación media como la puntuación de coincidencia única para cada par de registros.

Puntuaciones de coincidencia nulas

Un algoritmo de coincidencia aplica una puntuación de coincidencia predefinida a un par de valores cuando uno o ambos valores son NULL. Puede editar la puntuación de coincidencia para que un algoritmo de coincidencia de campos se aplique a los valores nulos.

Puntuaciones de coincidencia nulas y algoritmos de coincidencia de campos

Cuando configura un algoritmo de coincidencia de campos, compruebe los valores de la puntuación de coincidencia que aplica el algoritmo a los datos nulos. Un algoritmo de coincidencia de campos aplica una puntuación predeterminada de 0,5 cuando compara dos valores y uno o ambos son nulos. Una puntuación de 0,5 indica un nivel bajo de similitud entre los valores de datos.

Tenga en cuenta las siguientes reglas y directrices cuando compruebe las puntuaciones de coincidencia nulas:

- Cuando el algoritmo analiza las columnas que contienen claves principales u otros datos críticos, no edite las puntuaciones predeterminadas. En este caso, un valor nulo representa un error de datos y las puntuaciones predeterminadas son adecuados para los datos.
- Cuando el algoritmo analiza columnas que pueden contener datos de forma opcional, actualice los valores de puntuación de coincidencia nulos con el mismo valor que el umbral de coincidencia. Puede cancelar el efecto de los valores nulos en el análisis de coincidencia cuando configure las puntuaciones de coincidencia nulas con el valor del umbral de coincidencia.

Puntuaciones de coincidencia nulas y algoritmos de coincidencia de identidades

Un algoritmo de coincidencia de identidades aplica una puntuación de coincidencia de 0 cuando compara dos valores y uno o ambos son nulos. El análisis de coincidencia de identidades asigna un registro con una puntuación de coincidencia nula a un único clúster de registros y registra un valor de tamaño de clúster de 1. No puede editar la puntuación que un algoritmo de coincidencia de identidades aplica a datos nulos.

Opciones de salida del clúster

Seleccione una opción de salida del clúster cuando quiera organizar registros similares o idénticos en los datos de salida.

Cuando se selecciona una opción de salida del clúster, la transformación añade un valor de ID de clúster a cada registro de salida. Puede ordenar los registros por valor de ID de clúster. La salida de la transformación incluye una fila por cada registro. Si un registro no coincide con otro con una puntuación que satisfaga el umbral de coincidencia, la transformación asigna un ID de clúster único al registro. Utilice la vista **Salida de coincidencia** para seleccionar o actualizar las opciones de salida del clúster.

Puede seleccionar las siguientes opciones de salida del clúster:

Clústeres

Seleccione la opción para asignar valores de ID de clúster a los registros de salida.

Clústeres: mejor coincidencia

Seleccione la opción para añadir el par de registros con la puntuación más alta a un clúster. Dado que un registro podría representar la mejor coincidencia con más de un registro, es posible que más de un par de registros compartan un mismo valor de ID de clúster.

Clústeres: coincidir con todos

La opción **Clústeres: coincidir con todos** funciona de la misma manera que la opción **Clústeres**.

La transformación utiliza **Clústeres: coincidir con todos** y **Clústeres: mejor coincidencia** como nombres de opción en el análisis de coincidencia de identidad.

Nota: Si un servicio de integración de datos ejecuta varias transformaciones de coincidencia simultáneamente, el servicio de integración de datos genera valores de ID de clústeres únicos para la salida de cada transformación. Por tanto, los valores de ID de clústeres para los registros que cada transformación genera pueden ser no consecutivos.

La opción Clústeres y la opción Clústeres: coincidir con todos

Seleccione la opción Clústeres en el análisis de coincidencia de campos. Seleccione la opción Clústeres: coincidir con todos en el análisis de coincidencia de identidad.

La transformación de coincidencia utiliza las siguientes reglas para crear los clústeres:

- Cuando dos registros tienen una puntuación de coincidencia que cumple el umbral de coincidencia, la transformación de coincidencia los añade a un clúster.
- Cuando un registro del conjunto de datos coincide con cualquier registro del clúster, la transformación añade el registro al clúster.
- Si un registro de un clúster coincide con un registro de otro clúster, el proceso fusiona los clústeres.
- La transformación realiza un barrido continuo de los resultados de coincidencias hasta que todos los registros pertenecen a un clúster.
- Si un registro no coincide con otro en el conjunto de datos, la transformación asigna un valor de ID de clúster único al registro.

La opción Clústeres: mejor coincidencia

Seleccione la opción Clústeres: mejor coincidencia en el análisis de coincidencia de identidad.

La transformación utiliza las siguientes reglas para crear los clústeres:

- La transformación identifica el registro que tiene la mayor puntuación de coincidencia con el registro actual. Si la puntuación de coincidencia cumple con el umbral, la transformación añade el par de registros a un clúster.
- Si uno de los registros coincidentes está en un clúster, la transformación añade el otro registro al clúster actual.
- La transformación realiza un barrido continuo de los resultados de la puntuación de coincidencia hasta que todos los registros pertenecen a un clúster.
- Un clúster puede contener un único registro si el registro no coincide con ningún otro registro de los datos.

Nota: Puede usar la propiedad **Coincidencia** en la vista **Salida de coincidencia** para especificar el modo en que la transformación coteja un único origen de datos con un almacén de datos persistentes. La propiedad **Coincidencia** determina si la transformación busca duplicados en los datos de origen o en el almacén de datos persistentes.

TEMAS RELACIONADOS

- [“Pares de coincidencia y clústeres” en la página 481](#)

Puntuaciones de controlador y puntuaciones de enlace en el análisis de clústeres

Cuando selecciona una opción de salida del clúster en la transformación de coincidencia, puede añadir datos de puntuaciones de enlace y de puntuaciones del controlador a la salida.

La puntuación de enlace es la puntuación entre dos registros que los identifica como miembros del mismo clúster. Los enlaces entre registros determinan la composición del clúster. Cualquier registro puede enlazarse a cualquier otro registro del mismo clúster.

La puntuación del controlador es la puntuación entre el registro con el valor más alto de ID de secuencia en un clúster y otro registro del mismo clúster. Las puntuaciones del controlador proporcionan un medio para evaluar todos los registros de un clúster frente a un solo registro. Cuando añade puntuaciones del controlador a la salida de coincidencia, se ejecuta la asignación más lentamente, dado que la transformación de coincidencias no puede calcular las puntuaciones del controlador hasta que se hayan completado todos los clústeres.

Nota: El análisis de coincidencias genera un solo conjunto de puntuaciones para cada estrategia que defina. La puntuación del controlador y la puntuación de enlace indican las puntuaciones de coincidencia para diferentes pares de registros en cada clúster. Las puntuaciones del controlador y las puntuaciones de enlace pueden depender el orden en que los registros introduzca la transformación. La puntuación de controlador podría ser menor que el umbral de coincidencia.

Ejemplo de análisis del clúster

Puede configurar una estrategia de coincidencia de campos para analizar una columna de datos de apellidos. Puede establecer un umbral de coincidencia de 0,825 en la estrategia. Seleccione un formato de salida de clúster y ejecute el visor de datos en la transformación.

La siguiente tabla muestra los datos que el visor de datos muestra:

Apellido	ID de secuencia	Cluster_ID	Tamaño de clúster	ID de controlador	Puntuación de controlador	ID de enlace	Puntuación de enlace
SMITH	1	1	2	1 - 6	1	1 - 1	1
SMYTH	2	2	2	1 - 3	0.83333	1 - 2	1
SMYTHE	3	2	2	1 - 3	1	1 - 2	0.83333
SMITT	4	3	1	1 - 4	1	1 - 4	1
SMITS	5	4	1	1 - 5	1	1 - 5	1
SMITH	6	1	2	1 - 6	1	1 - 1	1

El visor de datos contiene la siguiente información sobre los datos de apellidos:

- SMITT y SMITS no coinciden con ningún registro con una puntuación que cumpla con el umbral de coincidencia. La transformación de coincidencia determina que los registros son únicos en el conjunto de datos.
SMITT y SMITS tienen un tamaño de clúster de 1. Para buscar registros únicos en la salida del clúster, busque clústeres que contengan un único registro.
- SMITH y SMITH tiene una puntuación de enlace de 1. La transformación de coincidencia determina que los registros son idénticos. La transformación añade los registros a un único clúster.
- SMYTH y SMYTHE tienen una puntuación de enlace de 0.83333. La puntuación supera el umbral de coincidencia. Por tanto, la transformación añade los registros a un único clúster.

Análisis de datos principales

Quando se desea analizar dos orígenes de datos en la transformación de coincidencia, debe identificar un origen como el conjunto de datos principales. La transformación compara los valores de datos de cada

registro del conjunto de datos especificado con los valores correspondientes de cada registro del otro conjunto de datos.

En muchas organizaciones, un conjunto de datos principales constituye un almacén de datos permanente y de alta calidad. Antes de añadir registros a un conjunto de datos principales, utilice la transformación de coincidencia para comprobar que los registros no añaden información duplicada a los datos principales.

Ejemplo de datos principales

Un banco mantiene un conjunto de datos principales de registros de cuentas de clientes. El banco actualiza el conjunto de datos principales cada día con registros que identifican nuevas cuentas de clientes. El banco utiliza una asignación de análisis de duplicados para comprobar que los nuevos registros no duplican información de clientes en el conjunto de datos principales. El conjunto de datos principales y las tablas de cuentas nuevas tiene una estructura común, y las tablas utilizan el mismo tipo de base de datos. Por tanto, el banco puede reutilizar la asignación de análisis de duplicados cada vez que necesite actualizar el conjunto de datos principales.

Direccionalidad en el análisis del conjunto de datos principal

La transformación de coincidencia compara los registros de un conjunto de datos con los de otro en una única dirección. La transformación compara cada registro del conjunto de datos principal con todos los registros del segundo conjunto de datos. No compara cada registro del segundo conjunto de datos con todos los registros del conjunto de datos principal. En consecuencia, la selección del conjunto de datos principal puede influir en los resultados del análisis de coincidencia.

La siguiente tabla recoge dos conjuntos de datos que se pueden comparar en el análisis de coincidencia de identidad:

Conjunto de datos 1	Conjunto de datos 2
Alex Bell	Alexander Bell
Alexander Graham Bell	Thomas Edison
Alva Edison	Nicola Tesla
Marie Curie	Irene Joliot Curie
Dorothy Crowfoot	Dorothy Hodgkin

Si selecciona el conjunto de datos 1 como conjunto de datos principal y selecciona la opción de salida **Mejor coincidencia**, el resultado incluirá los siguientes registros:

- Alex Bell, Alexander Bell
- Alexander Graham Bell, Alexander Bell

Si selecciona el conjunto de datos 2 como conjunto de datos principal y selecciona la opción de salida **Mejor coincidencia**, el resultado incluirá los siguientes registros:

- Alexander Bell, Alex Bell

Cuando el conjunto de datos 2 es el conjunto de datos principal, la transformación no halla una coincidencia entre Alexander Bell y Alexander Graham Bell, ya que Alexander Bell ya coincide con Alex Bell en los datos de salida.

Reutilización de la asignación

Si añade datos a un conjunto de datos principales de forma periódica, configure una asignación de análisis de duplicados que pueda reutilizar. Puede reutilizar la asignación si las configuraciones de puerto del origen de datos que compara con el conjunto de datos principal no cambian.

Cuando ejecute la asignación, compruebe que la transformación especifica los datos principales y los datos más recientes. Puede ejecutar la asignación sin ninguna otra actualización de configuración.

Datos del análisis de coincidencia de identidad y del índice persistente

Cuando se ejecuta una asignación para analizar la información de identidad, la transformación de coincidencia genera un índice que almacena las versiones alternativas de las identidades en el conjunto de datos. De forma predeterminada, la transformación de coincidencia escribe los datos de índice en archivos temporales. Puede configurar la transformación para que guarde los datos de índice en tablas de base de datos.

Las tablas de índices que genera representan un almacén de datos que puede reutilizar en posteriores asignaciones. Puede comparar las tablas de índices con un origen de datos y, opcionalmente, puede actualizar las tablas de índices con datos de índice del origen de datos. Dado que la transformación no vuelve a generar las tablas de índices, las asignaciones posteriores se ejecutan más rápidamente. Además, las tablas de índices pueden representar un almacén de datos de confianza de datos de identidad.

TEMAS RELACIONADOS

- [“Rendimiento de coincidencia en el análisis de identidad” en la página 490](#)

Reglas y directrices para datos de índice persistentes

Tenga en cuenta las siguientes reglas y pautas cuando configure una transformación de coincidencia para analizar un conjunto de datos principales de información de identidad:

- Para generar un índice reutilizable para un conjunto de datos principal, configure la transformación para que escriba los datos de índice en tablas de base de datos. Las tablas de base de datos constituyen un almacén persistente de los datos de índice.
- Para comparar las identidades de otro conjunto de datos con un almacén de datos de índice, configure el conjunto de datos como el origen de datos de la asignación. Configure la transformación de coincidencia para leer los datos de origen y el almacén de datos de índice. Seleccione las tablas de índices en el esquema predeterminado de la conexión de base de datos que especifique.
- La transformación de coincidencia añade el valor del identificador de secuencia del registro de entrada a las filas de datos de índice que corresponden al registro. El puerto de entrada *SequenceID* contiene los identificadores de secuencia. La transformación utiliza los identificadores de secuencia para realizar un seguimiento de los datos de índice a través de los diferentes pasos del análisis de coincidencia. No desconecte el puerto de ID de secuencia.
- Cuando conecta una transformación de coincidencia a un almacén de índice, la transformación reutiliza los valores de propiedad de los campos de población, nivel de clave, tipo de clave y clave de la transformación que creó el almacén. La transformación también reutiliza la configuración de puertos de la transformación que creó el almacén.

Si las propiedades de la transformación no coinciden, el análisis de identidad no podrá comparar los datos de origen de la asignación y los datos de índice correctamente.

- La transformación de coincidencia utiliza los datos del puerto de entrada seleccionado como campo de clave para generar el índice de identidad. La transformación también puede escribir datos de otros puertos en el índice. Si desconecta de la transformación los puertos de datos de campos que no son de clave, borrará los datos que haya en las columnas de índice correspondientes al ejecutar la asignación. Para conservar los datos de puerto de entrada en las tablas de índice, no desconecte los puertos de datos de entrada.
- Puede desactivar el análisis de coincidencia en la transformación de coincidencia al generar los datos de tablas de índices para un conjunto de datos. Por ejemplo, puede desactivar el análisis de coincidencia al crear un almacén de índice para un conjunto de datos. Al desactivar el análisis de coincidencia, la asignación se ejecuta más rápidamente.

Si deshabilita el análisis de coincidencias, la transformación de coincidencia puede generar y mostrar los códigos de los estados de persistencia y las descripciones de los mismos. La transformación no genera ni muestra puntuaciones de coincidencia u otros datos asociados con los resultados del análisis de coincidencias. Por ejemplo, si configura la transformación para asignar registros a clústeres y deshabilita el análisis de coincidencias, la transformación no generará ni mostrará los valores de ID de clúster.

- Puede determinar si la transformación de coincidencia actualiza el almacén de índice con datos del origen de la asignación. La transformación de coincidencia utiliza identificadores de secuencia para determinar si las filas del almacén de índice y de los datos de la asignación representan los mismos registros.

Rendimiento de asignación de coincidencias

Puede previsualizar los factores de datos que determinan el rendimiento de la transformación de coincidencia antes de ejecutar la asignación que contiene la transformación. Puede comprobar que el sistema tiene recursos para ejecutar la asignación. También puede comprobar que ha configurado la transformación correctamente para medir los niveles de similitud en los datos de entrada.

Utilice la opción **Análisis de rendimiento de coincidencia** para comprobar que el sistema tiene los recursos requeridos. Utilice la opción **Análisis del clúster de coincidencias** para comprobar que la asignación puede medir exactamente los niveles de similitud en los datos de entrada.

Ejecute el análisis de rendimiento de coincidencia y el análisis del clúster de coincidencia en cualquier transformación de coincidencia que lea un único origen de datos. Ejecute el análisis de rendimiento de coincidencia en cualquier transformación de coincidencia que realice el análisis de coincidencia de campo de origen dual. No ejecute el análisis de rendimiento de coincidencia ni el análisis del clúster de coincidencia en una estrategia de coincidencia de identidades que se conecte a tablas de índices.

Obtener detalles sobre el análisis de rendimiento de coincidencia

Puede obtener detalles sobre los datos del análisis de coincidencias para ver los pares de registros que cumplen o superan el umbral de coincidencia. Haga doble clic en un registro en la vista **Detalles** y utilice el visor de datos para ver los registros que coinciden con el registro que seleccione. El visor de datos muestra los datos de cada par de registros en una única fila. La fila contiene el ID de fila de cada registro en el par.

Obtener detalles sobre el análisis del clúster de coincidencia

Puede obtener detalles sobre los datos del análisis del clúster para ver los registros de cada clúster. Haga doble clic en un clúster en la vista **Detalles** para leer los datos en el visor de datos. El visor de datos muestra un clúster cada vez. Los datos del clúster incluyen las opciones de puntuación seleccionadas, tales como la puntuación del controlador, la puntuación de enlace, el ID del controlador o el ID de enlace.

Registro de la transformación de coincidencia

Cuando ejecuta una asignación que utiliza una transformación de coincidencia, el registro de Developer tool realiza un seguimiento del número cálculos de comparación que realiza la asignación. Para ver los datos del registro, seleccione la opción **Mostrar registro** en el visor de datos.

La asignación actualiza el registro cada 100.000 cálculos.

Visualización de datos de análisis del clúster de coincidencias

Puede ver datos estadísticos en los clústeres que la transformación puede crear. Las estadísticas de clústeres resumen el nivel de duplicación de registros en el conjunto de datos en función de la configuración de asignación actual.

Para ver los datos, haga clic con el botón derecho en la transformación de coincidencia en el lienzo de la asignación y seleccione **Análisis del clúster de coincidencias**.

Antes de ejecutar el análisis, valide la asignación que contiene la transformación.

El análisis del clúster de coincidencias muestra datos de las propiedades siguientes:

Propiedad	Descripción
Origen	El número de filas de datos de entrada.
Última ejecución	La fecha y la hora del análisis.
Número total de clústeres detectados	El número de clústeres que el análisis de coincidencia genera cuando se ejecuta la asignación.
Tamaño de clúster mínimo	El número de registros del clúster o clústeres que contienen el menor número de registros. Si el tamaño de clúster mínimo es 1, el conjunto de datos contiene al menos un único registro.
Tamaño de clúster máximo	El número de registros del clúster o clústeres que contienen el mayor número de registros. Si este valor excede considerablemente el tamaño del clúster medio, el clúster de mayor tamaño podría contener falsos duplicados.
Número de registros únicos	El número de registros en el conjunto de datos que no coinciden con ningún otro registro con una puntuación que satisfaga el umbral de coincidencia.
Número de registros duplicados	El número de registros en el conjunto de datos que coinciden con otro registro con una puntuación que satisfaga el umbral de coincidencia.
Total de comparaciones	El número de operaciones de comparación que realiza la asignación.
Tamaño de clúster medio	El número medio de registros en un clúster.

Visualización de los datos del análisis de rendimiento de coincidencia

Puede ver datos estadísticos en los grupos de registros que la asignación lee como datos de entrada.

Para ver los datos, haga clic con el botón derecho en la transformación de coincidencia en el lienzo de la asignación y seleccione **Análisis de rendimiento de coincidencia**.

Antes de ejecutar el análisis, valide la asignación que contiene la transformación.

El análisis de rendimiento de coincidencia muestra datos de las propiedades siguientes:

Propiedad	Descripción
Origen	El número de filas de datos de entrada.
Última ejecución	La fecha y la hora del análisis.
Número total de grupos detectados	El número de grupos definidos para el conjunto de datos, en función del valor de clave de grupo seleccionada.
Productividad (registros por minuto)	Valor de variable que calcula la velocidad del análisis de coincidencia. Establezca este valor. Utilice el valor para calcular el tiempo necesario para ejecutar el análisis de coincidencia.
Tiempo estimado para buscar coincidencias en los registros	El tiempo necesario para analizar todos los registros en el conjunto de datos, en función de la configuración de la transformación de coincidencia.
Número total de pares generados	El número de comparaciones que la transformación debe realizar, en función del número de filas de datos de entrada y el número de grupos.
Tamaño de grupo mínimo	Valor de variable que indica el número mínimo de registros que un grupo puede contener. Establezca este valor. Utilice el valor para comprobar que la asignación creará grupos de un tamaño utilizable. Nota: El valor del tamaño de grupo mínimo no determina el tamaño de los grupos creados cuando se ejecuta la asignación.
Número de grupos por debajo del umbral mínimo	El número de grupos que contienen menos registros que el valor del tamaño de grupo mínimo. Si varios grupos se encuentran por debajo del tamaño de grupo mínimo, quizás deba editar la transformación y seleccionar una clave de grupo diferente.
Tamaño de grupo máximo	Valor de variable que indica el número máximo de registros que un grupo puede contener. Establezca este valor para el análisis de rendimiento. Utilice el valor para comprobar que la asignación creará grupos de un tamaño utilizable. Nota: El valor no determina el tamaño de los grupos creados cuando se ejecuta la asignación.
Número de grupos por encima del umbral máximo	El número de grupos que contienen más registros que el valor de tamaño máximo de grupo. Si hay varios grupos por encima del tamaño de grupo máximo, quizás deba editar la transformación y seleccionar una clave de grupo diferente.

Rendimiento de coincidencia en el análisis de identidad

Para mejorar el rendimiento de la asignación al realizar el análisis de identidad en dos conjuntos de datos, configure la transformación de coincidencia para leer datos de índice de identidad desde tablas de base de

datos. Ejecute una asignación para crear las tablas de índices para el conjunto de datos principal. Ejecute la asignación de nuevo para comparar los datos de índice con otro origen de datos.

Utilice las opciones de la vista **Tipo de coincidencia** para identificar las tablas de base de datos para almacenar los datos de índice. Utilice las mismas opciones para seleccionar las tablas de índices al configurar la transformación para comparar los datos de índice con los datos de otro origen.

Para escribir los datos de índice en tablas de base de datos, realice las tareas siguientes:

1. Cree una asignación que lea un origen de datos de información de identidad.
2. Configure una transformación de coincidencia en la asignación para escribir los datos de índice en una base de datos.
3. Ejecute la asignación para generar los datos de índice. Los datos de índice representan un almacén de datos que puede reutilizar.

Para leer los datos de índice desde las tablas de base de datos, realice las tareas siguientes:

1. Cree una asignación que lea otro origen de datos de identidad.
2. Configure una transformación de coincidencia en la asignación para leer los datos de índice desde la base de datos que ha especificado anteriormente.

Cuando el origen de datos de la asignación y los datos de índice comparten una estructura común, puede reutilizar la asignación que generó los datos de índice.

3. Ejecute la asignación para comparar el origen de datos con los datos de índice.

La asignación genera datos de índice para el origen de datos. La asignación no necesita generar datos de índice para el conjunto de datos más grande. Por tanto, la asignación se ejecuta más rápidamente que una asignación de origen dual que genera datos de índice para ambos conjuntos de datos.

TEMAS RELACIONADOS

- [“Datos del análisis de coincidencia de identidad y del índice persistente” en la página 487](#)

Creación de un almacén de datos para datos de índice de identidad

Configure una asignación que lea un origen de datos que contiene información de identidad. Utilice una transformación de coincidencia para escribir los datos de índice en una base de datos.

1. Cree una asignación y añada el origen de datos al lienzo de la asignación.
2. Añada una transformación de coincidencia al lienzo de la asignación.
3. En el origen de datos, seleccione los puertos que contienen la información de identidad.
 - Conecte los puertos de información de identidad a la transformación de coincidencia.
 - Conecte el puerto que contiene los valores de identificador de secuencia a la transformación de coincidencia.
4. En la transformación de coincidencia, seleccione la vista **Tipo de coincidencia**.
5. Establezca el tipo de coincidencia en **Coincidencia de identidad con ID de registro persistente**.
6. Configure las opciones siguientes para crear el almacén de datos de índice.
 - Establezca el método de persistencia en **Actualizar la base de datos con ID nuevos**.
 - Compruebe el valor del proceso de coincidencia. El valor predeterminado es **Habilitar**. Si desactiva el proceso de coincidencia, la asignación crea las tablas de índice de identidad, pero no realiza el análisis de coincidencia en los datos.

- En el menú de conexión de base de datos, seleccione una base de datos para las tablas de índice.
 - En el menú Almacén persistente, seleccione **Crear nuevo**. Especifique un nombre para el índice en el cuadro de diálogo **Crear tablas de almacenamiento**.
7. Realice cualquier otro paso de configuración que la transformación de coincidencia requiere. Por ejemplo, configure una estrategia de transformación.
 8. Añada un objeto de datos de destino a la asignación.
 9. Conecte los puertos de salida de la transformación de coincidencia con el objeto de datos de destino.
 10. Ejecute la asignación.

La asignación escribe los datos de índice para el origen de datos en las tablas de base de datos que especifique.

Utilización del almacén de datos de índice en el análisis de un solo origen

Configure una asignación que lea un origen de datos de información de identidad. Utilice una transformación de coincidencia para comparar el origen de datos con el almacén de datos de índice del conjunto de datos principales.

Antes de configurar la asignación, compruebe que el origen de datos contiene una columna de valores de identificador de secuencia.

Para ahorrar tiempo, puede copiar o reutilizar la asignación que creó el almacén de datos de índice.

1. Abra la asignación que generó el almacén de datos de índice.
Alternativamente, abra una copia de la asignación.
2. Compruebe el origen de datos en la asignación.
Si es necesario, reemplace el origen de datos por un origen de datos que contenga los datos actuales.
Nota: Si elimina el origen de datos, también puede eliminar las conexiones del puerto en la transformación de coincidencia.
3. Identifique los puertos de origen de datos que contienen información de identidad.
 - Conecte los puertos de información de identidad a la transformación de coincidencia.
 - Conecte el puerto que contiene los valores de identificador de secuencia a la transformación de coincidencia.
Los puertos de entrada y el orden de puertos de entrada deben corresponderse con los puertos de entrada de la transformación que creó las tablas de índices.
4. En la transformación de coincidencia, seleccione la vista **Tipo de coincidencia**.
5. Establezca el tipo de coincidencia en **Coincidencia de identidad con ID de registro persistente**.
6. Compruebe la población, el nivel de clave, el tipo de clave y los valores de campo de clave.
7. Configure las opciones para identificar el almacén de datos de índice:
 - Defina el método de persistencia. Por ejemplo, seleccione **No actualizar la base de datos** para conservar los datos actuales en las tablas de índices.
 - Establezca el proceso de coincidencia en **Habilitar**.
 - En el menú Conexión de base de datos, seleccione la base de datos que contiene las tablas de índices.
 - En el menú Almacén persistente, vaya a las tablas que contienen los datos de índice.
8. Configure las siguientes propiedades en la vista Salida de coincidencia:

- Coincidir. Identifica los registros que analizar cuando la transformación lee los datos de índice de las tablas de base de datos.
 - Salida. Filtra los registros que la transformación escribe como salida.
9. Realice cualquier otro paso de configuración que la transformación de coincidencia requiere.
Por ejemplo, configure una estrategia de transformación.
 10. Compruebe el destino de datos en la asignación.
Si es necesario, reemplace el destino de datos por otro objeto de destino.
 11. Conecte los puertos de salida de la transformación de coincidencia con el objeto de datos de destino.
 12. Ejecute la asignación.

La asignación compara los registros de origen de datos con el almacén de datos de índice. La transformación escribe los datos de índice para el origen de datos en el almacén de datos.

Vistas de transformación de coincidencia

La transformación de coincidencia organiza las opciones que puede configurar en una serie de vistas. En una transformación reutilizable, las vistas aparecen como tabuladores en el lienzo de Developer tool. Para abrir una vista, haga clic en una ficha. En una transformación no reutilizable, las vistas aparecen como una lista en el lienzo. Para abrir una vista, haga clic en un elemento en la lista.

Cuando se configura una operación de análisis de coincidencia, puede configurar las siguientes vistas:

General

Utilice la vista General para actualizar el nombre y la descripción de una transformación de coincidencia no reutilizable. La descripción es opcional.

Puertos

Utilice la vista Puertos para comprobar los puertos de entrada y los puertos de salida en una transformación de coincidencia no reutilizable.

Resumen

Utilice el resumen para actualizar el nombre y la descripción de una transformación reutilizable. La descripción es opcional. También puede utilizar el resumen para crear los puertos de entrada y los puertos de salida en una transformación reutilizable.

Tipo de coincidencia

Utilice la vista Tipo de coincidencia para seleccionar el tipo de análisis de duplicados que realiza la transformación. Puede seleccionar un análisis de coincidencia de campos o un análisis de coincidencia de identidades. Puede especificar un único origen de datos o dos orígenes de datos.

Estrategias

Utilice la vista Estrategias para definir una o más estrategias para analizar los datos de entrada. Seleccione dos columnas de datos en cada estrategia y asigne un algoritmo de análisis de coincidencia para las columnas.

Salida de coincidencia

Utilice la vista Salida de coincidencia para especificar la estructura y el formato de los datos de salida.

Parámetros

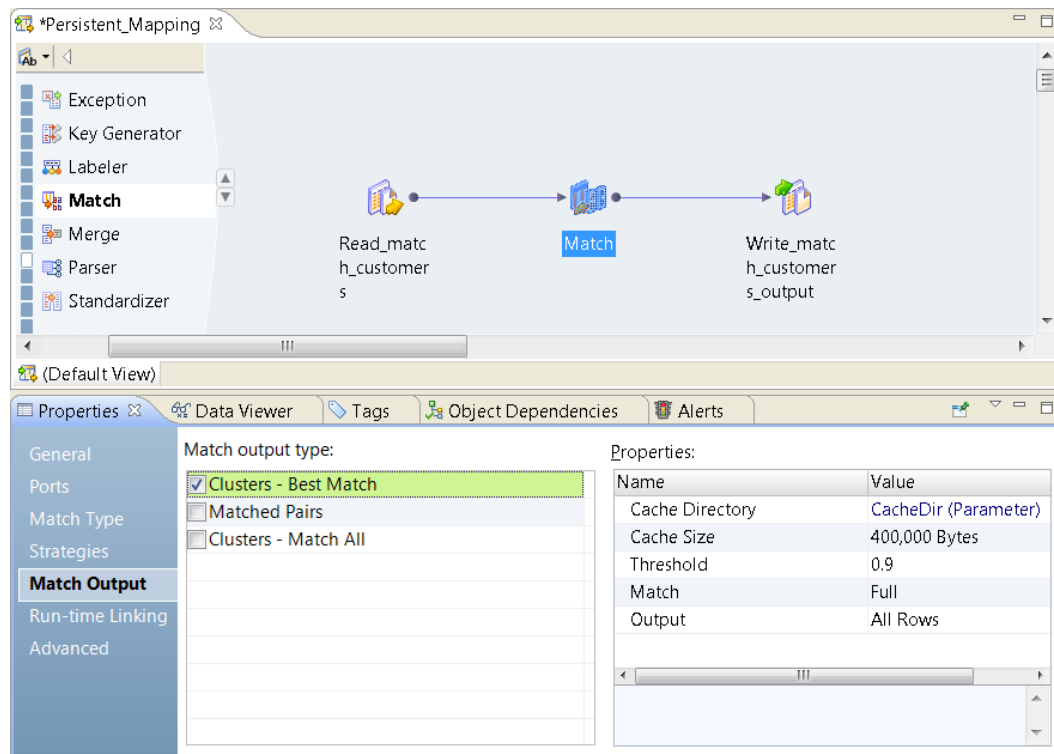
Utilice la vista Parámetros para definir un parámetro que el Servicio de integración de datos pueda aplicar a la transformación cuando ejecute una asignación que contiene la transformación.

Avanzadas

Utilice la vista Avanzada para especificar las propiedades siguientes:

- El nivel de detalle en todos los mensajes de registro que la asignación escribe para la transformación.
- El número de procesos que utilizan las operaciones de coincidencia de identidades al ejecutar una asignación que contiene la transformación.
- Si la transformación pasa datos de registros idénticos directamente a los puertos de salida. Puede filtrar registros idénticos cuando configure la transformación para escribir una salida agrupada.

La siguiente imagen muestra las vistas en una transformación de coincidencia que ha configurado para el análisis de identidades con un almacenamiento de índice persistente:



Puertos de transformación de coincidencia

La transformación de coincidencia incluye un conjunto de puertos de entrada y de salida predefinidos que contienen metadatos para las operaciones de análisis de coincidencias que defina. La transformación selecciona o elimina los puertos cuando configure el tipo de coincidencia y las opciones de salida de coincidencia.

Cuando configure la transformación, revise los puertos de metadatos. Cuando añada la transformación a una asignación, compruebe que conecta los puertos de metadatos a los puertos correctos en los objetos de asignación del nivel superior e inferior.

Puertos de entrada de la transformación de coincidencia

Los puertos de entrada predefinidos contienen los metadatos que la transformación necesita para el análisis de coincidencias.

Después de crear una transformación de coincidencia, puede configurar los siguientes puertos de entrada:

SequenceId

Identificador único para cada registro en el conjunto de datos del origen de la asignación. Cada registro de un conjunto de datos de entrada debe incluir un identificador único de secuencia. Si un conjunto de datos contiene identificadores de secuencia duplicados, la transformación de coincidencia no puede identificar los registros duplicados correctamente. Utilice la transformación de generador de claves para crear identificadores únicos en el caso de que no exista ninguno en los datos.

Cuando cree un almacén de datos de índice para los datos de identidad, la transformación de coincidencia añade el identificador de secuencia para cada registro en el almacén de datos. Cuando configure la transformación para comparar un origen de datos con el almacén de datos de índice, la transformación podría encontrar un identificador de secuencia común en ambos conjuntos de datos. La transformación puede analizar los identificadores de secuencia si son únicos en los respectivos conjuntos de datos.

GroupKey

Valor de clave que identifica el grupo al que pertenece el registro.

Nota: Para aumentar el rendimiento de la asignación, configure el puerto de entrada GroupKey y el puerto de salida al que está conectado con el mismo valor de precisión.

Puertos de salida de la transformación de coincidencia

Los puertos de salida predefinidos contienen metadatos sobre el análisis que realiza la transformación.

Después de crear una transformación de coincidencia, puede configurar los siguientes puertos de salida:

GroupKey

Valor de clave que identifica el grupo al que pertenece el registro.

Las transformaciones de nivel inferior como la transformación de asociación pueden leer el valor de clave de grupo.

ClusterId

El identificador del clúster al que pertenece el registro. Se utiliza en la salida del clúster.

ClusterSize

Número de registros en el clúster al que pertenece un registro. Cuando un clúster contiene un único registro, el tamaño de clúster es 1. Se utiliza en la salida del clúster.

RowId y RowId1

Un identificador de fila único para el registro. La transformación de coincidencia utiliza el identificador de fila para identificar la fila durante las operaciones de análisis de coincidencia. El identificador podría no coincidir con el número de fila en los datos de entrada.

DriverId

El identificador de fila del registro de controlador en un clúster. Se utiliza en la salida del clúster. El registro de controlador es el registro en el clúster con el valor más alto en el puerto de entrada SequenceID.

DriverScore

La transformación asigna una puntuación de controlador de salida de pares de coincidencia y de salida agrupada. En un par de coincidencia, la puntuación del controlador es la puntuación de coincidencia entre el par de registros. En un clúster, la puntuación del controlador es la puntuación de coincidencia entre el registro actual y el registro de controlador en el clúster.

LinkId

El identificador de fila del registro que coincidió con el registro actual y que se enlazó al clúster. Se utiliza en la salida del clúster.

LinkScore

La puntuación de coincidencia entre dos registros que da como resultado la creación de un clúster o la adición de un registro a un clúster. El puerto LinkID identifica el registro con el que el registro actual comparte la puntuación de vínculo. Se utiliza en la salida del clúster.

PersistenceStatus

Un código de ocho caracteres que representa los resultados del análisis de coincidencia en un registro de entrada. Se usa en el análisis de identidad de un solo origen cuando la transformación compara el origen de datos con un almacén de datos de índice.

La transformación rellena los tres primeros caracteres del código. La transformación puede devolver diferentes caracteres en cada posición. La transformación devuelve 0 para las posiciones de cuatro a ocho.

Cuando configure la transformación para generar la salida de pares de coincidencia, la transformación crea un puerto PersistenceStatus y un puerto PersistenceStatus1.

PersistenceStatusDesc

Una descripción de texto de los valores de códigos de estado de persistencia. Se usa en el análisis de identidad de un solo origen cuando la transformación compara el origen de datos con un almacén de datos de índice.

Cuando configure la transformación para generar la salida de pares de coincidencia, la transformación crea un puerto PersistenceStatusDesc y un puerto PersistenceStatusDesc1.

Códigos de estado de persistencia y descripciones de estado de persistencia

Los códigos de estado de persistencia y las descripciones de estado de persistencia describen la relación entre los diferentes tipos de datos de índice que analiza la transformación de coincidencia. La transformación genera los códigos de estado y las descripciones de estado al configurar la transformación para leer un almacén de datos de identidad persistente.

La transformación escribe el código de estado de persistencia en el puerto PersistenceStatus. El código contiene ocho caracteres. La transformación rellena las tres primeras posiciones en la cadena con valores de código. La transformación devuelve 0 para las posiciones de cuatro a ocho.

La transformación escribe la descripción de estado de persistencia en el puerto PersistenceStatusDesc. La descripción contiene tres cadenas de texto separadas por comas que describen los valores de las tres primeras posiciones del código de estado de persistencia.

La transformación utiliza los valores de identificador de secuencia de los registros de datos de origen para comparar los datos de índice de los dos conjuntos de datos.

La tabla siguiente describe los tipos de información que la transformación escribe en cada posición de la descripción de estado y el código de estado:

Posición	Descripción
1	Identifica el conjunto de datos que contiene el registro.
2	Indica el estado duplicado del registro. La transformación busca identificadores de secuencia comunes entre los datos de entrada de la transformación y el almacén de datos de índice.
3	Describe las acciones que la transformación realiza en los datos.
4-8	El código de estado contiene 0 en cada posición. La descripción de estado no contiene texto para las posiciones.

Valores de código de estado y valores de descripción de estado

Los códigos de estado de persistencia y las descripciones de estado de persistencia describen la relación entre los registros de entrada de la transformación y los registros que el almacén de datos representa. La transformación utiliza valores de identificador de secuencia para identificar los registros y determinar la relación entre los registros de los conjuntos de datos.

El código de estado de persistencia y las descripciones de estado de persistencia tienen una estructura común. Los códigos de estado y las descripciones de estado contienen la misma información en cada posición de la cadena de datos de salida.

Estado de conjunto de datos

El primer valor del código de estado y de la descripción de estado identifica el conjunto de datos que contiene el registro.

La tabla siguiente describe los códigos de estado y las descripciones de estado que la transformación puede devolver en la primera posición:

Código de estado	Descripción de estado
S	Almacén. El registro actual se origina en el almacén de datos de índice.
I	Entrada. El registro actual se origina en los datos de entrada de la transformación.

Estado de registros duplicados

El segundo valor del código de estado y de la descripción de estado describe la relación entre los datos de índice de la transformación y el almacén de datos persistente.

La tabla siguiente describe los códigos de estado y las descripciones de estado que la transformación puede devolver en la segunda posición:

Código de estado	Descripción de estado
A	Ausente. El almacén de datos de índice no contiene datos para el registro actual.
E	Existe. El registro actual está presente en el almacén de datos de índice y en los datos de entrada de la transformación.
I	No válido. La transformación no puede analizar el registro actual. Por ejemplo, la transformación no puede generar datos de índice para el registro porque el campo clave en la ficha Tipo de coincidencia no es compatible con los datos del registro.
N	Nuevo. El registro está presente en el origen de datos.
0	[Dash] El registro está presente en el almacén de datos de índice.

Estado del almacén de datos

El tercer valor del código de estado y de la descripción de estado describe cualquier acción que la transformación realiza en las tablas de datos de índice.

La tabla siguiente describe los códigos de estado y las descripciones de estado que la transformación puede devolver en la tercera posición:

Código de estado	Descripción de estado
A	Añadido. La transformación agrega los datos de índice para el registro de entrada actual al almacén de datos persistente. Los datos de entrada de la transformación y los datos de índices persistentes tienen identificadores de secuencia diferentes.
I	Se ha omitido. La transformación no agrega ningún dato de índice para el registro de entrada actual al almacén de datos persistente.
N	La transformación devuelve una de las siguientes descripciones: <ul style="list-style-type: none"> - No hay cambios. El registro actual se origina en el almacén de datos persistente y la transformación no realiza ninguna acción. - No añadidos. La transformación no actualiza el almacén de datos persistente con datos para el registro de entrada actual debido a la política de coincidencia que ha definido.

Código de estado	Descripción de estado
R	Quitada. La transformación elimina los datos de índice para el registro del almacén de datos de índice.
U	Actualizado. La transformación actualiza las filas en el almacén de datos persistentes con datos de índice del registro de entrada de la transformación. Los datos de entrada de la transformación y los datos de índice persistentes tienen identificadores de secuencia comunes.

Ejemplo de descripción de estado de persistencia

El código de estado de persistencia INA00000 tiene la siguiente descripción de estado de persistencia:

Entrada, Nuevo, Añadido

El código de estado y la descripción de estado contienen la siguiente información sobre el registro:

- El registro se origina en los datos de entrada de la transformación.
- El almacén de datos persistente no contiene ninguna copia del registro.
- La transformación añade datos de índice para el registro al almacén de datos persistente.

Selección de puertos de salida y de salida de coincidencia

Las opciones de salida de coincidencia que seleccione determinarán los puertos de salida en la transformación. Por ejemplo, la transformación crea un puerto ClusterId y un puerto ClusterSize cuando se selecciona un tipo de salida de clúster.

Seleccione el tipo de salida de transformación que necesita y revise los puertos en la transformación.

Si actualiza el tipo de salida de coincidencia, compruebe la configuración del puerto de salida en la transformación después de hacerlo. Si utiliza la transformación en una asignación, puede que sea necesario volver a conectar los puertos de salida a objetos de un nivel inferior en la asignación.

Mapplets de coincidencia

Un mapplet de coincidencia es un tipo de mapplet que se puede crear e incrustar en transformaciones de coincidencia.

Para crear mapplets de coincidencia, guarde la configuración de una transformación de coincidencia como mapplet de coincidencia. Cuando se crea un mapplet de coincidencia se convierten los ajustes de la transformación de coincidencia en transformaciones de comparación y de media ponderada.

Después de crear un mapplet de coincidencia, puede añadir transformaciones para personalizar el análisis de coincidencia. Por ejemplo, puede añadir una transformación de expresión para evaluar las puntuaciones de enlace de dos estrategias y elegir la puntuación más alta.

A diferencia de las transformaciones de coincidencia, los mapplets de coincidencia son pasivos, lo que significa que puede usarlos como reglas dentro de Analyst Tool. Utilice mapplets de coincidencia en Analyst Tool para hacer coincidir registros como parte de procesos de creación de perfiles de datos.

La transformación de coincidencia solamente puede leer mapplets creados dentro de una transformación de coincidencia.

Creación de un mapplet de coincidencia

Cree un mapplet de coincidencia para definir una operación de análisis de coincidencia que utiliza varias transformaciones.

1. Abra una transformación de coincidencia en el editor y seleccione la vista **Estrategias**.
2. Seleccione **Usar regla de coincidencia**.
3. En el campo **Nombre**, seleccione **Crear nuevo**.
Se abre la ventana **Nuevo mapplet**.
4. En la ventana **Nuevo mapplet**, escriba un nombre para el mapplet y elija una ubicación donde guardarlo.
5. Opcionalmente, seleccione **Reutilizar las estrategias de la transformación de coincidencia** para copiar las entradas, las estrategias y ponderaciones desde la transformación de coincidencia actual al mapplet de coincidencia.
Nota: Informatica recomienda usar esta configuración para crear mapplets de coincidencia rápidamente y replicar las funciones de coincidencia definidas actualmente en las transformaciones de coincidencia.
6. Haga clic en **Finalizar**.
El mapplet de coincidencia se abre en el editor.
7. Si lo desea, cree operaciones de coincidencia añadiendo y configurando transformaciones de comparación y transformaciones de media ponderada en el mapplet de coincidencia.
8. Haga clic en **Archivo > Guardar** para guardar el mapplet.
9. Cierre el mapplet y seleccione el editor que contiene la transformación de coincidencia. Compruebe que el mapplet que ha creado aparece en el campo **Nombre**.
10. Si lo desea, configure los campos de coincidencia en el mapplet haciendo clic en el botón **Campos coincidentes**.
Se abre la ventana **Configurar regla de coincidencia**.
11. Haga doble clic en los campos de las columnas **Campos de entrada** y **Entradas disponibles** para asignar puertos de entrada a entradas coincidentes.
12. Haga clic en **Archivo > Guardar** para guardar la transformación.

Cómo usar un mapplet de coincidencia

Puede seleccionar y configurar un mapplet de coincidencia definido previamente en la transformación de coincidencia.

1. Abra una transformación de coincidencia en el editor y seleccione la vista **Estrategias**.
2. Seleccione **Usar regla de coincidencia**.
3. En el campo **Nombre**, seleccione **Usar existente**.
Se abre la ventana **Configurar regla de coincidencia**.
4. Haga clic en **Explorar** para localizar un mapplet de coincidencia en el repositorio.
Importante: Solamente se pueden seleccionar mapplets creados por la transformación de coincidencia.
Se abre la ventana **Seleccionar mapplet coincidente**.
5. Seleccione un mapplet coincidente y haga clic en **Aceptar**.

6. Haga doble clic en los campos de las columnas **Campos de entrada** y **Entradas disponibles** para asignar puertos de entrada a entradas coincidentes.
7. Haga clic en **Aceptar**.
Se cierra la ventana **Configurar regla de coincidencia**.
8. Haga clic en **Archivo > Guardar** para guardar la transformación de coincidencia.

Configuración de una operación de análisis de coincidencia

Para configurar una operación de coincidencia, conecte los datos de origen a la transformación de coincidencia y edite las propiedades en las vistas de la transformación.

1. Cree una transformación de coincidencia y conecte los datos de origen a la transformación.
2. Seleccione la vista **Tipo de coincidencia** y elija un tipo de coincidencia.
3. Configure las propiedades para el tipo de operación de coincidencia que haya seleccionado.
Si ha seleccionado un tipo de coincidencia de origen dual, configure la propiedad **Conjunto de datos principal**.
4. Seleccione la vista **Estrategias** y elija **Definir estrategias de coincidencia**.
5. Haga clic en **Nueva**.
Se abre el asistente **Nueva estrategia de coincidencia**.
6. Elija una estrategia de coincidencia y haga clic en **Siguiente**.
7. Opcionalmente, edite la configuración de ponderación y de coincidencia nula. Haga clic en **Siguiente**.
8. Haga doble clic en las celdas en la columna Disponibles para seleccionar los puertos de entrada que analizar.
Haga clic en **Siguiente** para configurar otra estrategia o haga clic en **Finalizar** para salir del asistente.
Nota: Para editar la configuración de la estrategia, haga clic en la flecha de la celda de esa estrategia en la vista **Estrategias**.
9. Seleccione la vista **Salida de coincidencia**.
Elija un tipo de salida de coincidencia y configure las propiedades.

Nota: También puede configurar estrategias de coincidencia seleccionando o creando un mapplet de coincidencia en la vista **Estrategias**. Un mapplet de coincidencia es un tipo de mapplet que se puede incrustar en una transformación de coincidencia.

Transformación de coincidencia en un entorno no nativo

El procesamiento de transformaciones de coincidencia en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de coincidencia en el motor de Blaze

La transformación de coincidencia es compatible con las siguientes restricciones:

- Una transformación de coincidencia genera valores de ID de clúster de modo diferente en entornos nativos y no nativos. En un entorno no nativo, la transformación anexa un valor de ID de grupo al ID de clúster.
- La validación de asignaciones no se realiza si la transformación de coincidencia está configurada para escribir datos del índice de identidades en tablas de bases de datos.

Transformación de coincidencia en el motor de Spark

La transformación de coincidencia es compatible con las siguientes restricciones:

- Una transformación de coincidencia genera valores de ID de clúster de modo diferente en entornos nativos y no nativos. En un entorno no nativo, la transformación anexa un valor de ID de grupo al ID de clúster.
- La validación de asignaciones no se realiza si la transformación de coincidencia está configurada para escribir datos del índice de identidades en tablas de bases de datos.

Transformación de coincidencia en el motor de Databricks Spark

La transformación de coincidencia es compatible con las siguientes restricciones:

- Una transformación de coincidencia genera valores de ID de clúster de modo diferente en entornos nativos y no nativos. En un entorno no nativo, la transformación anexa un valor de ID de grupo al ID de clúster.
- La validación de asignaciones no se realiza si la transformación de coincidencia está configurada para escribir datos del índice de identidades en tablas de bases de datos.

CAPÍTULO 30

Transformaciones de coincidencia en el análisis de campos

Este capítulo incluye los siguientes temas:

- [Análisis de coincidencia de campos, 503](#)
- [Flujo de proceso para el análisis de coincidencia de campos, 504](#)
- [Opciones del tipo de coincidencia de campos, 504](#)
- [Estrategias para la coincidencia de campos, 505](#)
- [Opciones de salida de coincidencia de campos, 508](#)
- [Propiedades avanzadas de la coincidencia de campos, 510](#)
- [Ejemplo de análisis de coincidencia de campos, 511](#)

Análisis de coincidencia de campos

Realizar el análisis de coincidencia de campos para buscar registros similares o duplicados en un conjunto de datos o entre dos conjuntos de datos.

Cuando configure la transformación de coincidencia para el análisis de coincidencia de campos, establezca las opciones en las siguientes vistas:

- Tipo de coincidencia
- Estrategias
- Salida de coincidencia

Opcionalmente, establezca las opciones en las vistas Parámetros y Avanzadas.

Flujo de proceso para el análisis de coincidencia de campos

El siguiente flujo de procesos resume los pasos necesarios para configurar una transformación de coincidencia para el análisis de coincidencia de campos. Puede definir un proceso que utiliza la transformación de coincidencia solamente o que utiliza la transformación de coincidencia y otras transformaciones.

Nota: Cuando se añade una transformación de coincidencia a una asignación en el análisis de coincidencia de campos, añada una transformación de generador de claves de nivel superior a la asignación.

Para preparar los datos para la transformación de coincidencia, realice los pasos siguientes:

1. Organice los registros de datos de origen en grupos.
Utilice una transformación de generador de claves para asignar un valor de clave de grupo a cada registro. Las asignaciones de grupos reducen el número de cálculos que debe realizar la transformación de coincidencia.
2. Compruebe que los registros del origen de datos contienen valores únicos de identificadores de secuencia. Puede utilizar una transformación de generador de claves para crear los valores.

Realice los pasos siguientes en la transformación de coincidencia:

1. Especifique el análisis de campos como el tipo de coincidencia y especifique el número de orígenes de datos.
Si configura la transformación para analizar dos conjuntos de datos, seleccione un conjunto de datos principales.
Utilice la vista **Tipo de coincidencia** para establecer el tipo y el número de orígenes de datos.
2. Defina una estrategia de análisis de coincidencia. Seleccione un algoritmo y asigne un par de columnas al algoritmo.
Utilice la vista **Estrategias** para definir la estrategia.
3. Especifique el método que utiliza la transformación para generar los resultados del análisis de coincidencia.
4. Establezca el valor del umbral de coincidencia. El umbral de coincidencia es la puntuación mínima que puede identificar a dos registros como duplicados el uno del otro.
Utilice la vista **Salida de coincidencia** para seleccionar el método de salida y el umbral de coincidencia.
Nota: Puede establecer el umbral de coincidencia en una transformación de coincidencia o una transformación de media ponderada. Utilice la transformación de media ponderada si va a crear un maplet de coincidencia.

Opciones del tipo de coincidencia de campos

La vista Tipo de coincidencia contiene una única opción que se aplica a un análisis de coincidencia de origen dual. La opción identifica el conjunto de datos principales. La vista Tipo de coincidencia no contiene una opción para el análisis de coincidencia de un solo origen.

Cuando desee analizar dos conjuntos de datos, debe seleccionar uno como el conjunto de datos principales. Si ningún conjunto de datos representa un conjunto de datos principales en el proyecto o en la organización, seleccione el conjunto de datos de mayor tamaño como el conjunto de datos principales.

Utilice la opción **Conjunto de datos principal** para especificar el conjunto de datos principales.

Estrategias para la coincidencia de campos

La vista Estrategias enumera las estrategias que define para los datos de entrada.

Las estrategias determinan cómo mide la transformación las similitudes y diferencias entre los registros del origen de datos.

Algoritmos de coincidencia de campos

La transformación de coincidencia incluye algoritmos que comparan los valores de datos entre dos columnas. Cada algoritmo calcula el grado de diferencia entre los valores de datos de una manera diferente.

Seleccione un algoritmo que puede medir los tipos de diferencia de datos que espera encontrar en las columnas que seleccione.

Bigram

Utilice el algoritmo Bigram para comparar cadenas de texto largas, como las direcciones postales introducidas en un único campo.

El algoritmo Bigram calcula una puntuación de coincidencia para dos cadenas de datos según la ocurrencia de caracteres consecutivos en ambas cadenas. El algoritmo busca pares de caracteres consecutivos que sean comunes en ambas cadenas. Divide el número de pares que coincide en ambas cadenas por el número total de pares de caracteres.

Ejemplo de Bigram

Considere las siguientes cadenas:

- larder
- lerder

Estas cadenas producen los siguientes grupos Bigram:

```
l a, a r, r d, d e, e r  
l e, e r, r d, d e, e r
```

Observe que la segunda ocurrencia de la cadena "e r" dentro de la cadena "lerder" no tiene coincidencia, puesto que no hay una segunda ocurrencia correspondiente de "e r" en la cadena "larder".

Para calcular la puntuación de coincidencia Bigram, la transformación divide el número de pares coincidentes (6) por el número total de pares en ambas cadenas (10). En este ejemplo, las cadenas son similares en un 60% y la puntuación de coincidencia es de 0,60.

Distancia de Hamming

Utilice el algoritmo Distancia de Hamming cuando la posición de los caracteres de datos sea un factor crítico; por ejemplo, en campos numéricos o de códigos como números de teléfono, códigos postales o códigos de producto.

El algoritmo de la distancia de Hamming calcula una puntuación de coincidencia para dos cadenas de datos calculando el número de posiciones en las que los caracteres difieren de una cadena de datos a otra. En el caso de que las cadenas tengan una longitud diferente, cada carácter adicional de la cadena más larga se cuenta como una diferencia de una cadena a otra.

Ejemplo de distancia de Hamming

Considere las siguientes cadenas:

- Morlow
- Marlowes

Los caracteres resaltados indican las posiciones que el algoritmo de Hamming identifica como diferentes.

Para calcular la puntuación de coincidencia Hamming, la transformación divide el número de caracteres coincidentes (5) por la longitud de la cadena más larga (8). En este ejemplo, las cadenas son similares en un 62,5% y la puntuación de coincidencia es de 0,625.

Distancia de edición

Utilice el algoritmo Distancia de edición para comparar palabras o cadenas de texto cortas, como un nombre.

El algoritmo Distancia de edición calcula el "coste" mínimo de transformar una cadena en otra por medio de la inserción, eliminación o reemplazo de caracteres.

Ejemplo de Distancia de edición

Considere las siguientes cadenas:

- Levenston
- Levenshtein

Los caracteres resaltados indican las operaciones necesarias para transformar una cadena en otra.

El algoritmo Distancia de edición divide el número de caracteres no modificados (8) entre la longitud de la cadena más larga (11). En este ejemplo, las cadenas son similares en un 72,7% y la puntuación de coincidencia es de 0,727.

Distancia de Jaro

Utilice el algoritmo Distancia de Jaro para comparar dos cadenas cuando la similitud entre los caracteres iniciales de las cadenas tenga prioridad.

La puntuación de coincidencia de la Distancia de Jaro refleja el grado de similitud entre los primeros cuatro caracteres de ambas cadenas y el número de transposiciones de carácter identificadas. La transformación pondera la importancia de la coincidencia entre los primeros cuatro caracteres utilizando el valor que se introduce en la propiedad Penalización.

Propiedades de la distancia de Jaro

Cuando se configura un algoritmo Distancia de Jaro, puede configurar las siguientes propiedades:

Penalización

Determina la penalización para la puntuación de coincidencia si los primeros cuatro caracteres en dos cadenas comparadas no son idénticos. La transformación resta el valor completo de la penalización cuando no hay ninguna coincidencia con el primer carácter. La transformación resta fracciones de la penalización según la posición que ocupan los otros caracteres sin coincidencia. El valor predeterminado de la penalización es 0,20.

Distinguir mayúsculas de minúsculas

Determina si el algoritmo Distancia de Jaro tiene en cuenta el formato de mayúsculas y minúsculas de los caracteres cuando los compara.

Ejemplo de la distancia de Jaro

Considere las siguientes cadenas:

- 391859
- 813995

Si utiliza el valor de Penalización predeterminado de 0,20 para analizar estas cadenas, el algoritmo Distancia de Jaro devuelve una puntuación de coincidencia de 0,513. Esta puntuación de coincidencia indica que las cadenas tienen una similitud del 51,3%.

Invertir la distancia de Hamming

Utilice el algoritmo Invertir la distancia de Hamming para calcular el porcentaje de posiciones de carácter que difiere entre dos cadenas, leídas de derecha a izquierda.

El algoritmo de la distancia de Hamming calcula una puntuación de coincidencia para dos cadenas de datos calculando el número de posiciones en las que los caracteres difieren de una cadena de datos a otra. En el caso de que las cadenas tengan una longitud diferente, el algoritmo cuenta cada carácter adicional de la cadena más larga se cuenta como una diferencia de una cadena a otra.

Ejemplo de Invertir la distancia de Hamming

Considere las siguientes cadenas, que utilizan la alineación de derecha a izquierda para imitar el algoritmo Invertir la distancia de Hamming:

- 1-999-9999
- **011-01**-999-9991

Los caracteres resaltados indican las posiciones que el algoritmo Invertir la distancia de Hamming identifica como diferentes.

Para calcular la puntuación de coincidencia de Invertir la distancia de Hamming, la transformación divide el número de caracteres coincidentes (9) entre la longitud de la cadena más larga (15). En este ejemplo, la puntuación de coincidencia es 0,6, lo que indica que las cadenas tienen una similitud del 60%.

Propiedades de la estrategia para la coincidencia de campos

Abra el asistente de **estrategia** en la vista **Estrategias** y configure las propiedades para cada estrategia de coincidencia de campos.

Cuando se configura una estrategia de coincidencia de campos, se pueden configurar las siguientes propiedades:

Nombre

Identifica la estrategia por nombre.

Ponderación

Determina la prioridad de relación asignada a la puntuación de coincidencia cuando se calcula la puntuación general del registro. El valor predeterminado es 0,5.

Campo único nulo

Define la puntuación de coincidencia que el algoritmo aplica a un par de valores de datos cuando un valor es nulo. El valor predeterminado es 0,5.

Ambos campos nulos

Define la puntuación de coincidencia que el algoritmo aplica a un par de valores de datos cuando ambos valores son nulos. El valor predeterminado es 0,5.

Nota: Un algoritmo de coincidencia no calcula una puntuación de coincidencia cuando uno o ambos valores de columna de coincidencia son nulos. El algoritmo aplica las puntuaciones definidas en las propiedades de coincidencia nula. No puede borrar las propiedades de coincidencia nula.

Opciones de salida de coincidencia de campos

Configure las opciones de la **Salida de coincidencia** para definir el formato de salida del análisis de coincidencia de campos.

Puede configurar las opciones en el área **Tipo de salida de coincidencia** y el área **Propiedades**.

Tipos de salida de coincidencia

La vista Salida de coincidencia incluye opciones que especifican el formato de datos de salida. Puede configurar la transformación para escribir los registros en clústeres o en pares de coincidencia.

Elija uno de los siguientes tipos de salida de coincidencia:

Mejor coincidencia

Escribe cada registro en el conjunto de datos principales con el registro que representa la mejor coincidencia en el segundo conjunto de datos. La operación de coincidencia selecciona el registro en el segundo conjunto de datos que tenga la mayor puntuación de coincidencia para el registro de datos principal. Si dos o más registros devuelven la mayor puntuación, la operación de coincidencia selecciona el primer registro del segundo conjunto de datos. La mejor coincidencia escribe cada par de registros en una única fila.

Puede seleccionar **Mejor coincidencia** cuando configure la transformación para el análisis de origen dual.

Clústeres

Escribe clústeres que contienen conjuntos de registros que se vinculan entre sí con puntuaciones de coincidencia que se encuentran dentro del umbral de coincidencia. Cada registro debe coincidir con al menos otro registro del clúster con una puntuación que cumpla con el umbral.

Puede seleccionar **Clústeres** al configurar la transformación para análisis de un solo origen o de origen dual.

Pares de coincidencia

Escribe todos los pares de registros que coincidan con otros con una puntuación que cumple el umbral de coincidencia. La transformación escribe cada par en una única fila y añade la puntuación de coincidencia de cada par en cada fila. Si un registro coincide con más de un registro, la transformación escribe una fila por cada par de registros.

Puede seleccionar **Pares de coincidencia** al configurar la transformación para análisis de un solo origen o de origen dual.

Propiedades de salida de coincidencia

La vista Salida de coincidencia incluye las propiedades que especifican el comportamiento de la memoria caché, el umbral de puntuación de coincidencia y las puntuaciones de coincidencia que aparecen en la salida de transformación.

También puede utilizar las propiedades de la salida de coincidencia para especificar cómo la transformación añade los valores de puntuación de coincidencia a los registros de salida.

Después de seleccionar un tipo de salida de coincidencia, configure las propiedades siguientes:

Directorio de la memoria caché

Especifica el directorio en el que el Servicio de integración de datos escribe los datos temporales durante el análisis de coincidencia de campos. El Servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos que genera el análisis de coincidencia es mayor que la memoria disponible del sistema. El Servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del Servicio de integración de datos. El Servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de la memoria caché

Determina la cantidad de memoria del sistema que el Servicio de integración de datos asigna al análisis de coincidencia de campos. El valor predeterminado es 400.000 bytes.

Antes de ordenar los datos, el Servicio de integración de datos asigna la cantidad de memoria que se haya especificado. Si el análisis de coincidencia genera una cantidad de datos mayor, el Servicio de integración de datos escribirá los datos en exceso en el directorio de memoria caché. Si el análisis de coincidencia necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

Nota: Si especifica un valor de 65536 o superior, la transformación leerá el valor en bytes. Si especifica un valor inferior, la transformación leerá el valor en megabytes.

Umbral

Establece la puntuación de coincidencia mínima que identifica dos registros como duplicados potenciales entre sí.

Puede asignar un parámetro al valor de umbral. Establezca un valor decimal entre 0 y 1.

Método de puntuación

Determina los valores de puntuación de coincidencia que aparecen en la salida de transformación. Seleccione un método de puntuación para las salidas de los clústeres.

La siguiente tabla describe las opciones de métodos de puntuación:

Opción de método de puntuación	Descripción
Ambos	Añade la puntuación de enlace y la puntuación de controlador a cada registro del clúster.
Puntuación de enlace	Añade la puntuación de enlace a cada registro del clúster. Opción predeterminada.

Opción de método de puntuación	Descripción
Puntuación de controlador	Añade la puntuación de controlador a cada registro del clúster.
Ninguna	No añade ninguna puntuación de coincidencia a los registros del clúster.

Nota: Si añade la puntuación de controlador a los registros, puede aumentar el tiempo de ejecución de la asignación. La asignación espera hasta que todos los clústeres estén completos antes de añadir los valores de puntuación del controlador a los registros.

Propiedades avanzadas de la coincidencia de campos

La transformación incluye las propiedades avanzadas que determinan el número de instancias de ejecución, el modo en que la transformación analiza filas idénticas y el nivel de seguimiento de los datos de registro.

Puede configurar las siguientes propiedades avanzadas:

Instancias de ejecución

Determina el número de subprocesos que usa la transformación en tiempo de ejecución.

La transformación de coincidencia utiliza una única instancia de ejecución en el análisis de coincidencia de campos. Puede editar el número de instancias de ejecución cuando configura la transformación para el análisis de coincidencia de identidades.

Filtrar coincidencia exacta

Determina si la transformación aplica el algoritmo de comparación de una estrategia de coincidencia a pares de registros idénticos en los datos de entrada.

Cuando la transformación encuentra un par de registros idénticos, el algoritmo no necesita analizar los niveles de similitud entre los registros. La transformación puede transferir los registros directamente a la fase de salida sin análisis adicionales. Para configurar la transformación para que transfiera los registros idénticos directamente a la fase de salida, seleccione Filtrar coincidencia exacta. Cuando los datos de entrada contienen muchas filas idénticas, el algoritmo de comparación realiza menos cálculos y la asignación se ejecuta más rápido.

Seleccione la opción cuando los datos de entrada contengan muchas filas idénticas. No seleccione esta opción si los datos de entrada no contienen muchas filas idénticas, porque la transformación se ejecutará con más lentitud.

Nota: La salida de transformación contiene los mismos datos de registro cuando se selecciona o se desactiva la opción. Puede que la transformación asigne diferentes puntuaciones de enlace y de controlador a los registros de salida al seleccionar y desactivar la opción.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Ejemplo de análisis de coincidencia de campos

Usted trabaja como gestor de datos en un banco comercial. Recibe un conjunto de registros de cuentas de clientes que abrieron cuentas bancarias en los últimos siete días. Desea comprobar que el conjunto de datos no contiene registros duplicados. Diseña una asignación para buscar datos duplicados en los registros.

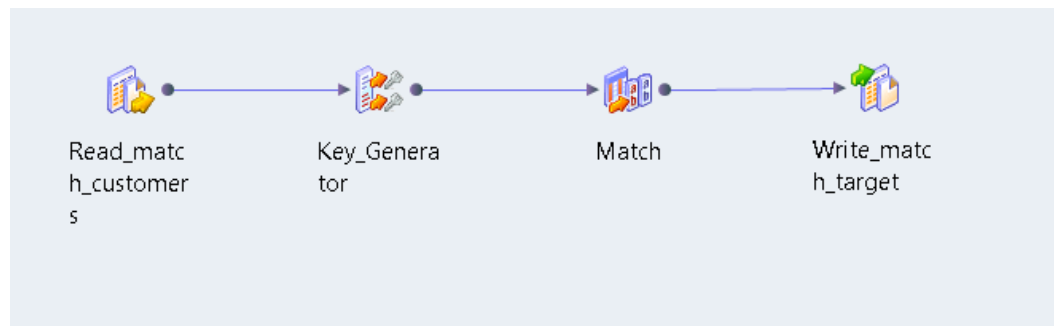
Crear la asignación

Cree una asignación que busque datos duplicados en varios campos.

La asignación realiza las siguientes tareas:

- Lee un origen de datos.
- Añade un valor de clave de grupo y un valor de identificador de secuencia a los registros de origen.
- Analiza los datos de los campos en los registros.
- Escribe los resultados en un destino de datos.

La siguiente imagen muestra la asignación en la herramienta del desarrollador:



La asignación que cree contiene los siguientes objetos:

Nombre de objeto	Descripción
Read_Match_Customers	Origen de datos. Contiene los detalles de los titulares de cuentas, e incluye los nombres, las direcciones y los números de cuentas.
Key_Generator	Transformación de generador de claves. Añade valores de clave de grupo y valores de identificador de secuencia a los datos de origen.
Coincidencia	Transformación de coincidencia. Analiza los niveles de duplicación en los datos de cuentas.
Write_match_target	Destino de datos. Contiene los resultados del análisis de campos.

Muestra de datos de entrada

El conjunto de datos contiene el número de cuenta, el nombre, la dirección y el empleador para cada cliente. Cree un origen de datos desde el conjunto de datos en el repositorio de modelos. Puede añadir el origen de datos a la asignación.

El siguiente fragmento de datos muestra un ejemplo de los datos de cuentas de clientes:

CustomerID	Lastname	City	State	ZIP
15954467	JONES	SCARSDALE	NY	10583
10110907	JONES	MINNEAPOLIS	MN	55437
19131127	JONES	INDIANAPOLIS	IN	46240
10112097	JONES	HOUSTON	TX	77036
19133807	JONES	PLANTATION	FL	33324
10112447	JONES	SCARSDALE	NY	10583
15952487	JONES	HOUSTON	TX	77002
10112027	JONES	OAKLAND	CA	94623

Configuración de la transformación de generador de claves

Cuando configure la transformación de generador de claves, conecte los puertos de origen de datos que desea analizar. Especifique el puerto que contiene los datos de la clave de grupo. Si los registros no incluyen un identificador único, utilice el puerto de ID de secuencia para añadir identificadores únicos al registro.

Cuando se especifica el puerto de claves de grupo, considere las siguientes directrices:

- Seleccione un puerto que contenga valores que se repiten con regularidad en los datos del puerto. Preferiblemente, seleccione un puerto que crea grupos que son similares en tamaño.
- Seleccione un puerto que no sea relevante para el análisis de duplicados.

En el ejemplo actual, seleccione el puerto City como la clave de grupo. Si un nombre de cuenta aparece más de una vez en una ciudad, es posible que las cuentas contengan datos duplicados. Si un nombre de cuenta aparece más de una vez en diferentes ciudades, es poco probable que las cuentas sean duplicadas.

Consejo: Ejecute un perfil de columna en el origen de datos antes de seleccionar el puerto de claves de grupo. Los resultados del perfil pueden indicar el número de veces que cada valor aparece en un puerto.

Configuración de la transformación de coincidencia

Añada una transformación de coincidencia no reutilizable a la asignación para realizar el análisis de campos.

Realice las tareas siguientes para configurar la transformación de coincidencia:

1. Seleccione el tipo de análisis de coincidencia que desea realizar.
2. Conecte los puertos de entrada a la transformación.
3. Configure estrategias para comparar los datos de registro.
4. Seleccione el tipo de datos de salida de coincidencia que crea la transformación.

5. Conecte los puertos de salida a un destino de datos.

Seleccionar el tipo de operación de coincidencia

Utilice las opciones en la vista **Tipo de coincidencia** para seleccionar la operación de coincidencia. Para comparar los datos de cuentas, configure la transformación para realizar el análisis de campos de un solo origen.

Conectar los puertos de entrada

Conecte los puertos de datos de las cuentas de clientes a la transformación de coincidencia.

La transformación de coincidencia utiliza puertos de entrada preestablecidos para determinar el orden en que procesa los registros. La transformación utiliza un identificador de secuencia para hacer un seguimiento de los registros desde los puertos de entrada hasta los clústeres o pares de coincidencia que escribe como salida. La transformación utiliza una clave de grupo para ordenar los registros que procesa.

Conecte los siguientes puertos preestablecidos en la transformación de coincidencia a los puertos de salida preestablecidos en la transformación de generador de claves:

- SequenceID
- GroupKey

Configurar las estrategias para el análisis de campos

Utilice las opciones en la vista **Estrategias** para configurar las estrategias. Las estrategias determinan los tipos de análisis que la transformación realiza en los datos de registros.

Cree las siguientes estrategias:

- Cree una estrategia que utilice el algoritmo Distancia de edición para analizar los números de identificación de los clientes. Seleccione los puertos CustomerID_1 y CustomerID_2.
- Cree una estrategia que utilice el algoritmo Distancia de Jaro para analizar los datos de los apellidos. Seleccione los puertos LastName_1 y LastName_2.

Nota: El algoritmo Distancia de Jaro aplica una penalización adicional a las cadenas similares que comienzan con caracteres diferentes. Por tanto, el algoritmo Distancia de Jaro puede aplicar una puntuación de coincidencia alta a PATTON y PATTEN, pero aplicar una puntuación de coincidencia inferior a BAYLOR y TAYLOR.

- Cree una estrategia que utilice el algoritmo Invertir la distancia de Hamming para analizar los datos de los códigos postales. Seleccione los puertos Zip_1 y Zip_2.

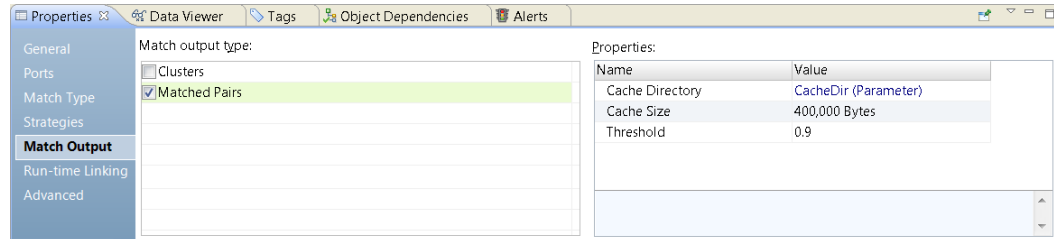
TEMAS RELACIONADOS

- [“Algoritmos de coincidencia de campos” en la página 505](#)

Seleccionar el tipo de salida de coincidencia

Utilice las opciones en la vista **Salida de coincidencia** para definir el formato de salida para los resultados del análisis de coincidencia.

La siguiente imagen muestra la vista Salida de coincidencia para el análisis de campos de un solo origen:



Configure la transformación para organizar los registros de salida en pares de coincidencia. Como la transformación devuelve los registros en pares, los resultados del análisis no incluyen registros únicos.

Conectar los puertos de salida

Conecte los puertos de salida de la transformación de coincidencia con el destino de datos en la asignación. Seleccione los puertos que contienen los datos de registro que desea escribir en el destino de datos.

La transformación incluye puertos preestablecidos que identifican los registros en cada par de coincidencia. Los nombres de los puertos preestablecidos son **RowId** y **RowId1**. Cada valor de identificador de fila identifica un registro en los datos de salida.

Los valores de identificador de fila corresponden los puertos que seleccione en una estrategia de coincidencia. Cuando configure una estrategia, seleccione los nombres de puertos con el sufijo **_1** o **_2**. El valor de RowId identifica el registro que contiene un puerto con el sufijo **_1**. El valor de RowId1 identifica el registro que contiene un puerto con el sufijo **_2**.

Puede utilizar otros puertos de salida para revisar las relaciones entre los registros. Los valores de puerto de enlace y los valores de puerto de controlador indican el alcance de la similitud entre los registros en cada clúster.

En el ejemplo, puede conectar todos los puertos con el destino de datos. Para ver los datos de salida en los puertos, ejecute el visor de datos.

Ejecutar el visor de datos

Ejecute el visor de datos para revisar los resultados del análisis de coincidencia. De forma predeterminada, el visor de datos muestra todos los puertos de salida en la transformación de coincidencia. Cuando se ejecuta la asignación, puede actualizar el destino de datos con los datos procedentes de los puertos de salida.

La siguiente imagen muestra los datos de salida en el visor de datos:

The screenshot shows the Field Match tool interface. At the top, a data flow diagram illustrates the process: Read_matc_h_customer_s_13 -> Key_Genera_tor -> Match -> Write_matc_h_target. Below the diagram, the Data Viewer window is open, displaying the output of the Match transformation. The output is a table with 12 columns: Lastname_1, Firstname_1, Company_1, Address_1_1, City_1, State_1, Zip_1, Customer1..., Lastname_2, Firstname_2, Company_2, Address_1_2, and City_2. The table contains 4 rows of data, with the first row being a header and the subsequent three rows representing individual records. The status bar at the bottom indicates 'Row 1 to 4'.

	Lastname_1	Firstname_1	Company_1	Address_1_1	City_1	State_1	Zip_1	Customer1...	Lastname_2	Firstname_2	Company_2	Address_1_2	City_2
1	Chan	David	SANDS BR...	515 FOLS...	SAN FRAN...	CA	94105	15954522	Chan	Roy	PG&E (PA...	77 BEALE...	SAN FRAN...
2	Andersen	Keith	CHARLES...	1100 N.E....	SAN ANT...	TX	78209	10110428	Andersen	Keith	CHARLES...	1100 N.E....	SAN ANT...
3	Abrary	David	AIRFINAN...	488 MADL...	NEW YORK	NY	10022	15954929	Abrary	Julianne	TRANSOC...	39 EAST 5...	NEW YORK
4	Chan	Agusta	J. STREICH...	86 TRINIT...	NEW YORK	NY	10006	19132518	Chan	Carmen	AMERICA...	86 TRINIT...	NEW YORK

El visor de datos comprueba que los datos de cuentas de clientes incluyen uno o más registros duplicados.

Tenga en cuenta los siguientes datos en el visor de datos:

- La transformación determina que los registros para Augusta Chan y Carmen Chan podrían contener la misma información porque comparten apellido y datos de dirección. Cuando revise los registros, puede decidir que los registros son únicos en el conjunto de datos. Sin embargo, observa que los registros comparten un valor de identificador de cliente común. Como la columna del identificador de cliente es una clave principal en el conjunto de datos, se pone en contacto con la sucursal de Nueva York. La sucursal de Nueva York resuelve el error.
- La transformación determina que los registros para Keith Anderson podrían contener la misma información. Cuando revisa los registros, comprueba que los dos registros representan la misma cuenta. Sin embargo, observa que los registros tienen diferentes valores de identificador de cliente. Como una cuenta de cliente debe tener un único valor de identificador, se pone en contacto con la sucursal de San Antonio. La sucursal de San Antonio resuelve el error.

Conclusión

Los resultados del análisis de coincidencia indican que el conjunto de datos de las cuentas de clientes incluyen al menos un par de registros duplicados. Usted se pone en contacto con la sucursal local para las cuentas que necesite comprobar. Comprueba que los demás registros en el conjunto de datos identifican de forma exclusiva a una cuenta de cliente.

CAPÍTULO 31

Transformaciones de coincidencia en el análisis de identidad

Este capítulo incluye los siguientes temas:

- [Análisis de coincidencia de identidades, 516](#)
- [Flujo de proceso para el análisis de coincidencia de identidades, 517](#)
- [Propiedades de tipo de coincidencia de identidades, 517](#)
- [Estrategias de coincidencia de identidades, 521](#)
- [Opciones de salida de coincidencia de identidades, 524](#)
- [Propiedades avanzadas de coincidencia de identidades, 527](#)
- [Caso práctico de índice persistente, 528](#)
- [Ejemplo de análisis de coincidencia de identidades, 530](#)

Análisis de coincidencia de identidades

Realice el análisis de coincidencia de identidades para encontrar identidades similares o duplicadas en un conjunto de datos o entre dos conjuntos de datos.

Las identidades similares en dos o más registros podrían indicar que hay registros duplicados. O bien, las identidades similares podrían indicar una conexión entre los registros, como una identidad familiar compartida o una identidad de empresario compartida.

Cuando configure la transformación de coincidencia para el análisis de coincidencia de identidades, establezca las opciones en las siguientes vistas:

- Tipo de coincidencia
- Estrategias
- Salida de coincidencia

Opcionalmente, establezca las opciones en las vistas Parámetros y Avanzadas.

Cuando configure una transformación de coincidencia para un análisis de coincidencia de identidades, conecte los puertos de entrada a todos los campos primarios requeridos de una estrategia de identidad. Muchas estrategias de identidad contienen campos primarios requeridos. Algunas de las estrategias

también contienen campos secundarios requeridos. Conecte un puerto de entrada a un campo secundario obligatorio como mínimo.

Flujo de proceso para el análisis de coincidencia de identidades

El siguiente flujo de proceso resume los pasos necesarios para configurar una transformación de coincidencia para el análisis de coincidencia de identidades. Puede definir un proceso que utiliza la transformación de coincidencia solamente o que utiliza la transformación de coincidencia y otras transformaciones.

Antes de conectar la transformación de coincidencia a objetos de datos de nivel superior, compruebe que los registros contienen valores únicos de identificador de secuencia. Puede utilizar una transformación de generador de claves para crear los valores. Cuando se realiza el análisis de coincidencia de identidades, también puede organizar los datos de entrada en grupos.

Realice los pasos siguientes en la transformación de coincidencia:

1. Especifique el análisis de identidad como el tipo de coincidencia y especifique el número de orígenes de datos.
Si configura la transformación para analizar dos conjuntos de datos, seleccione un conjunto de datos principales.
Utilice la vista **Tipo de coincidencia** para establecer el tipo y el número de orígenes de datos.
2. Identifique la ubicación para guardar los datos de índice. La transformación puede escribir los datos de índice en archivos temporales o guardarlos en tablas de base de datos.
Utilice la vista **Tipo de coincidencia** para especificar el almacén de datos de índice.
3. Defina una estrategia de análisis de coincidencia. Seleccione un algoritmo de relleno y de comparación y asigne un par de columnas al algoritmo.
La población indica los pares de columnas que seleccionar.
Utilice la vista **Estrategias** para definir la estrategia.
4. Especifique el método que utiliza la transformación para generar los resultados del análisis de coincidencia.
5. Establezca el valor del umbral de coincidencia. El umbral de coincidencia es la puntuación mínima que puede identificar a dos registros como duplicados el uno del otro.
Utilice la vista **Salida de coincidencia** para seleccionar el método de salida y el umbral de coincidencia.
Nota: Puede establecer el umbral de coincidencia en una transformación de coincidencia o una transformación de media ponderada. Utilice la transformación de media ponderada si va a crear un mapplet de coincidencia.

Propiedades de tipo de coincidencia de identidades

Utilice la vista Tipo de coincidencia para especificar el tipo de análisis que realiza la transformación de coincidencia y para establecer las propiedades que definen el análisis. Puede especificar un análisis de un

solo origen o un análisis de origen dual. También puede especificar un almacén de datos persistente para los datos de índice de identidad.

Las propiedades que configure dependen del tipo de análisis que seleccione. Muchas de las opciones son comunes a todos los tipos de análisis.

Propiedades comunes

Las siguientes propiedades son comunes a todos los tipos de análisis de identidad:

Población

Identifica el archivo de población que utiliza la transformación. El archivo de población contiene los algoritmos de generación de claves que crean las claves de índice.

Nivel de clave

Determina el número de claves que generan los algoritmos de identidad. La configuración predeterminada es Estándar. El ajuste Limitada produce un número menor de claves y una mayor precisión, y conlleva tiempos de procesamiento más largos. El ajuste Ampliada produce un número mayor de claves y una menor precisión y conlleva tiempos de procesamiento más reducidos.

Tipo de clave

Describe el tipo de información que contiene el campo de clave. El análisis de identidad puede generar claves para nombres de personas, organizaciones y direcciones. Seleccione el tipo de clave que describa mejor la columna que ha especificado en la propiedad **Campo de clave**.

Nivel de búsqueda

Indica el equilibrio entre la profundidad y la velocidad de búsqueda que la transformación aplica al análisis de coincidencia. La profundidad de búsqueda es inversamente proporcional al número de coincidencias devueltas. Por ejemplo, la opción **Exhaustiva** devuelve menos coincidencias.

Campo de clave

Especifica la columna que la transformación de coincidencia utiliza para generar los datos de clave de índice. Compruebe que la columna que ha seleccionado contiene el tipo de información que ha especificado en la propiedad **Tipo de clave**.

Directorio del índice

Identifica el directorio en el que el Servicio de integración de datos escribe los datos de clave de índice de la transformación actual. De forma predeterminada, la propiedad está en blanco. Si no especifica un directorio de índice, el Servicio de integración de datos utilizará la ubicación que haya establecido en el servicio de administración de contenido.

Puede especificar una ruta de acceso al directorio, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del Servicio de integración de datos. El Servicio de integración de datos debe poder escribir en el directorio.

Directorio de la memoria caché

Identifica el directorio en el que el Servicio de integración de datos escribe los datos temporales durante la fase de creación de índices del análisis de coincidencia de identidades. Actualice la propiedad para especificar una ubicación para los datos de la transformación actual. De forma predeterminada, la propiedad está en blanco. Si no especifica un directorio de la memoria caché, el Servicio de integración de datos utilizará la ubicación que haya establecido en el servicio de administración de contenido.

Puede especificar una ruta de acceso al directorio, o puede utilizar un parámetro para identificar el directorio. Especifique una ruta de acceso local en el equipo del Servicio de integración de datos. El Servicio de integración de datos debe poder escribir en el directorio.

Tamaño de la memoria caché

Determina la cantidad de memoria del sistema que el Servicio de integración de datos asigna a la creación de índices de identidad. El valor predeterminado es 400.000 bytes.

Si la operación de creación de índices genera una cantidad de datos mayor, el Servicio de integración de datos escribirá los datos en exceso en el directorio de la memoria caché. Si la operación necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

Nota: Si especifica un valor de 65536 o superior, la transformación leerá el valor en bytes. Si especifica un valor inferior, la transformación leerá el valor en megabytes.

Propiedades de origen dual

Además de las propiedades comunes, defina la siguiente propiedad al configurar la transformación para el análisis de origen dual:

Conjunto de datos principales

Identifica el origen de datos que contiene los datos principales. Especifique un conjunto de datos principal en el análisis de origen dual.

Propiedades de almacenamiento de datos persistente

Además de las propiedades comunes, defina las siguientes propiedades al configurar la transformación para usar el almacén de datos de índice persistente:

Método de persistencia

Especifica si la transformación actualiza las tablas de índices con datos de índice del origen de datos de la asignación. Seleccione una de las siguientes opciones:

- Actualizar la base de datos con ID nuevos.
La transformación agrega todas las filas a los datos de índice que no duplican un identificador de secuencia en los datos de índice. La transformación no actualiza las filas actuales del índice.
De forma predeterminada, la transformación realiza el análisis de coincidencia al seleccionar la opción. Puede utilizar la opción de proceso de coincidencia para activar o desactivar el análisis de coincidencia.
- No actualizar la base de datos.
La transformación no actualiza las tablas de índice con datos de índice del origen de datos de la asignación.
La transformación realiza el análisis de coincidencia al seleccionar la opción.
- Quite los ID de la base de datos.
La transformación elimina filas de las tablas de índice si las filas comparten identificadores de secuencia con los datos de origen de la asignación.
La transformación no realiza análisis de coincidencia cuando se selecciona la opción.
- Actualice los ID actuales en la base de datos.
La transformación reemplaza filas en las tablas de índice por filas de los datos de origen de la asignación si las filas comparten identificadores de secuencia. La transformación no añade filas al índice.
De forma predeterminada, la transformación realiza el análisis de coincidencia al seleccionar la opción. Puede utilizar la opción de proceso de coincidencia para activar o desactivar el análisis de coincidencia.

El método de persistencia predeterminado es **Actualizar la base de datos con ID nuevos**.

Proceso de coincidencia

Determina si la transformación actual realiza el análisis de identidad.

La opción que seleccione en la propiedad Método de persistencia determina las opciones en la propiedad Proceso de coincidencia.

Conexión de base de datos

Identifica la base de datos que contiene las tablas de índices.

Almacenamiento persistente

Identifica las tablas de índices en la base de datos que especifique.

TEMAS RELACIONADOS

- [“Caso práctico de índice persistente” en la página 528](#)
- [“Parámetros del método de persistencia” en la página 521](#)

Propiedades del directorio de índice y del directorio de la memoria caché

Los directorios de índice y de memoria caché almacenan los datos temporales que la transformación de coincidencia genera durante el análisis de identidad.

El servicio de integración de datos escribe los datos en el directorio de índice si no configura la transformación para que escriba los datos de índice en las tablas de base de datos. El servicio de integración de datos escribe los datos en el directorio de memoria caché si el análisis de identidad requiere más memoria de la que especifica el tamaño de la memoria caché. De forma predeterminada, las propiedades están en blanco en la vista Tipo de coincidencia. Si no especifica un directorio de índice ni un directorio de memoria caché, el servicio de integración de datos leerá las rutas de acceso de directorio desde el servicio de administración de contenido.

Cuando especifique un directorio de índice o de memoria caché en la vista Tipo de coincidencia, introduzca la ruta de acceso local del equipo del servicio de integración de datos. Puede especificar una ruta de acceso completa o una ruta de acceso relativa. Si especifica una ruta de acceso relativa, inicie la ruta de acceso con un punto. La ruta de acceso es relativa al directorio `tomcat/bin` del equipo del servicio de integración de datos.

La siguiente tabla muestra una ruta de acceso relativa en un directorio de memoria caché o una propiedad de directorio de índice e identifica la ruta de acceso completa que representa la propiedad:

Ruta de acceso relativa	Ruta de acceso completa
<code>./ch</code>	<code>[Informatica_installation_directory]/tomcat/bin/ch</code>

Puede configurar las propiedades de directorio de índice y de memoria caché para identificar el mismo directorio. Si especifica el mismo directorio en cada propiedad, el servicio de integración de datos creará directorios en el directorio que especifique. El servicio de integración de datos crea un directorio de `índice` para los datos del índice y un directorio de `memoria caché` para los datos de la memoria caché. Si especifica directorios diferentes en cada propiedad, el servicio de integración de datos escribirá los datos en los directorios que especifique.

El servicio de integración de datos elimina los archivos de índice y de memoria caché de los directorios después de ejecutar la asignación. El servicio de integración de datos no elimina los directorios. El servicio de integración de datos puede volver a utilizar los directorios en otras asignaciones si la transformación de coincidencia identifica los directorios.

Propiedades del servicio de administración de contenido

El servicio de administración de contenido especifica las siguientes ubicaciones predeterminadas para los directorios de índice y de memoria caché:

- `./identityIndex`. El directorio predeterminado de los datos de índice de identidad.
- `./identityCache`. El directorio predeterminado de los datos de la memoria caché de identidad.

Si no establece las propiedades en la transformación de coincidencia, el servicio de integración de datos creará los directorios predeterminados cuando se ejecute una asignación de coincidencia de identidad. El servicio de integración de datos crea los directorios en el directorio `tomcat/bin`.

Parámetros del método de persistencia

Cuando se selecciona un índice de datos de persistencia en el análisis de coincidencia de identidad, se puede utilizar un parámetro para identificar el método de persistencia. Utilice una cadena para definir el valor del parámetro.

En la siguiente tabla se muestran los métodos de persistencia que se pueden seleccionar en la ficha **Tipo de coincidencia**, así como los valores de parámetro que se pueden definir en cada uno de esos métodos:

Método de persistencia	Parámetro
Actualizar la base de datos con ID nuevos	ignore
No actualizar la base de datos	addNone
Quitar los ID de la base de datos	remove
Actualizar los identificadores en la base de datos	update

En los valores de parámetro se distingue entre mayúsculas y minúsculas.

TEMAS RELACIONADOS

- [“Propiedades de tipo de coincidencia de identidades” en la página 517](#)

Estrategias de coincidencia de identidades

La vista Estrategias enumera las estrategias que defina para los datos de identidad. Las estrategias determinan cómo mide la transformación las similitudes y diferencias entre los datos en el índice de identidad.

Algoritmos de coincidencia de identidades

La transformación de coincidencia incluye algoritmos de identidades predefinidos que comparan los valores de datos en el índice de identidad. Seleccione el algoritmo que mejor representa el tipo de datos de identidad en el conjunto de datos.

La siguiente tabla describe los algoritmos e identifica las entradas que selecciona para cada uno:

Algoritmo de identidad	Descripción
Dirección	Identifica los registros que comparten una dirección. El algoritmo requiere la siguiente entrada principal: - Dirección El algoritmo no requiere una entrada secundaria.
Contacto	Identifica los registros que comparten un contacto en una ubicación de una organización. El algoritmo requiere las siguientes entradas principales: - Person_Name - Organization_Name - Address_Part1 El algoritmo no requiere una entrada secundaria.
Entidad corporativa	Identifica los registros que comparten datos de identificación corporativa. Opcionalmente, seleccione este algoritmo para analizar los datos de direcciones y teléfonos. El algoritmo requiere la siguiente entrada principal: - Organization_Name El algoritmo no requiere una entrada secundaria.
División	Identifica los registros que comparten una misma oficina dentro de una organización. El algoritmo requiere las siguientes entradas principales: - Organization_Name - Address_Part1 El algoritmo no requiere una entrada secundaria.
Familia	Identifica individuos que pertenecen a una familia. Analiza datos de nombres, direcciones y números de teléfono. El algoritmo requiere la siguiente entrada principal: - Person_Name El algoritmo requiere una de las siguientes entradas secundarias: - Address_Part1 - Telephone_Number
Campos	Identifica los registros que comparten datos en los puertos que seleccione. El algoritmo no especifica ninguna entrada requerida. Seleccione el puerto o puertos que pueden contener datos de identidad duplicados.
Doméstico	Identifica individuos que pertenecen a un mismo hogar. Analiza datos de nombres y direcciones. El algoritmo requiere las siguientes entradas principales: - Person_Name - Address_Part1

Algoritmo de identidad	Descripción
Individuo	Identifica individuos duplicados. Analiza los nombres, las fechas de nacimiento y los datos de identificación personal, tales como números de seguridad social, números de cuenta y números de identificación de vehículos. El algoritmo requiere la siguiente entrada principal: - Person_Name El algoritmo requiere una de las siguientes entradas secundarias: - Fecha - ID
Organización	Identifica los registros que comparten datos de una organización. El algoritmo requiere la siguiente entrada principal: - Organization_Name El algoritmo no requiere una entrada secundaria.
Nombre de persona	Identifica los registros que comparten la información sobre los individuos. El algoritmo requiere la siguiente entrada principal: - Person_Name El algoritmo no requiere una entrada secundaria.
Residente	Identifica individuos duplicados en una dirección. Opcionalmente, configure esta estrategia para analizar datos de identificación personal. El algoritmo requiere la siguiente entrada principal: - Person_Name - Address_Part1 El algoritmo no requiere una entrada secundaria.
Contacto en general	Identifica los registros que comparten un contacto en una organización. El algoritmo requiere las siguientes entradas principales: - Person_Name - Organization_Name El algoritmo no requiere una entrada secundaria.
Doméstico en general	Identifica los individuos que pertenecen a un mismo hogar. El algoritmo requiere la siguiente entrada principal: - Address_Part1 El algoritmo no requiere una entrada secundaria.

Propiedades de la estrategia de coincidencia de identidades

Configure las propiedades de cada estrategia de coincidencia de identidades.

Cuando configure una estrategia de identidad, puede configurar las siguientes propiedades de estrategia:

Población

Determina la población a la que aplicar un análisis de identidad. Las poblaciones pueden contener algoritmos de creación de claves para configuraciones locales e idiomas específicos.

Nivel de coincidencia

Indica el equilibrio entre la calidad y la velocidad de búsqueda. La velocidad de búsqueda es inversamente proporcional al número de coincidencias devueltas. Las búsquedas que utilizan el ajuste

Aproximada devuelven menos coincidencias, mientras que las búsquedas que usan el ajuste Conservadora devuelven más coincidencias.

Opciones de salida de coincidencia de identidades

La vista Salida de coincidencia incluye opciones que especifican el formato de datos de salida. Puede configurar la transformación para escribir los registros en clústeres o en pares de coincidencia. También puede configurar la transformación para incluir o excluir diferentes categorías de identidad cuando realiza el análisis de identidad en un almacén de datos persistente de índice.

Puede configurar las opciones en el área **Tipo de salida de coincidencia** y en el área **Propiedades**.

Tipos de salida de coincidencia

La vista Salida de coincidencia incluye opciones que especifican el formato de datos de salida. Puede configurar la transformación para escribir los registros en clústeres o en pares de coincidencia.

Elija uno de los siguientes tipos de salida de coincidencia:

Mejor coincidencia

Escribe cada registro en el conjunto de datos principales con el registro que representa la mejor coincidencia en el segundo conjunto de datos. La operación de coincidencia selecciona el registro en el segundo conjunto de datos que tenga la mayor puntuación de coincidencia para el registro de datos principal. Si dos o más registros devuelven la mayor puntuación, la operación de coincidencia selecciona el primer registro del segundo conjunto de datos. La mejor coincidencia escribe cada par de registros en una única fila.

Puede seleccionar **Mejor coincidencia** cuando configure la transformación para el análisis de origen dual.

Clústeres: mejor coincidencia

Escribe clústeres que representan la mejor coincidencia entre un registro y otro en el mismo conjunto de datos o entre dos conjuntos de datos. La puntuación de coincidencia entre ambos registros debe cumplir el umbral de coincidencia. Los clústeres de mejor coincidencia pueden contener más de dos registros si un registro representa la mejor coincidencia con más de un registro.

Puede seleccionar **Clústeres: mejor coincidencia** en cualquier tipo de análisis de identidad.

Nota: El método de almacenamiento de datos de índice que seleccione puede afectar al contenido de la salida del clúster en el modo **Clústeres: mejor coincidencia**. Una transformación que se conecta con las tablas de índices puede crear clústeres diferentes a los de una transformación que almacena datos de índice para los mismos registros en los archivos temporales. El método de almacenamiento de datos de índice no afecta a las puntuaciones de coincidencia que genera la transformación para los pares de registros.

Clústeres: coincidir con todos

Escribe clústeres de registros que coinciden con una puntuación que cumple el umbral de coincidencia. Cada registro debe coincidir con al menos un registro del clúster.

Puede seleccionar **Clústeres: coincidir con todos** en cualquier tipo de análisis de identidad.

Pares de coincidencia

Escribe todos los pares de registros que coincidan con otros con una puntuación que cumple el umbral de coincidencia. La transformación escribe cada par en una única fila y añade la puntuación de

coincidencia de cada par en cada fila. Si un registro coincide con más de un registro, la transformación escribe una fila por cada par de registros.

Puede seleccionar **Pares de coincidencia** en cualquier tipo de análisis de identidad.

Propiedades de salida de coincidencia

La vista Salida de coincidencia contiene las propiedades que especifican el comportamiento de la memoria caché y el umbral de puntuación de coincidencia. También puede utilizar las propiedades para determinar el modo en el que la transformación selecciona registros del almacén de datos para el análisis y escribe registros del almacén de datos como salida.

Después de seleccionar un tipo de salida de coincidencia, configure las propiedades siguientes:

Directorio de memoria caché

Especifica el directorio en el que el servicio de integración de datos escribe los datos temporales durante el análisis de coincidencia de identidad. El servicio de integración de datos escribe archivos temporales en el directorio cuando el volumen de los datos que genera el análisis de coincidencia es mayor que la memoria disponible del sistema. El servicio de integración de datos elimina los archivos temporales después de la ejecución de asignación.

Puede especificar una ruta de acceso de directorio en la propiedad, o puede utilizar un parámetro del sistema para identificar el directorio. Especifique una ruta de acceso local en el equipo del servicio de integración de datos. El servicio de integración de datos debe poder escribir en el directorio. El valor predeterminado es el parámetro del sistema CacheDir.

Tamaño de la memoria caché

Determina la cantidad de memoria del sistema que el servicio de integración de datos asigna al análisis de coincidencia de identidades. El valor predeterminado es 400.000 bytes.

Si el análisis de coincidencia genera una cantidad de datos mayor, el servicio de integración de datos escribirá los datos en exceso en el directorio de memoria caché. Si el análisis de coincidencia necesita más memoria de la que pueden proporcionar la memoria del sistema y el almacenamiento de archivos, la asignación generará un error.

Nota: Si especifica un valor de 65536 o superior, la transformación leerá el valor en bytes. Si especifica un valor inferior, la transformación leerá el valor en megabytes.

Coincidencia

Identifica los registros que analizar cuando la transformación lee los datos de índice de las tablas de base de datos. Utilice las opciones en la vista **Tipo de coincidencia** para identificar las tablas de índices.

De forma predeterminada, la transformación analiza todos los registros del origen de datos y las tablas de base de datos de índices. Configure la propiedad de coincidencia para especificar un subconjunto de registros para el análisis de duplicados.

Salida

Filtra los registros que la transformación escribe como salida cuando se configura la transformación para leer las tablas de la base de datos de índices. Utilice las opciones en la vista **Tipo de coincidencia** para identificar las tablas de índices.

De forma predeterminada, la transformación de coincidencia escribe todos los registros desde el origen de datos y de las tablas de la base de datos de índices como salida. Configure la propiedad **Salida** cuando no necesite revisar todos los registros de los datos de entrada.

Umbral

Establece la puntuación de coincidencia mínima que identifica dos registros como duplicados potenciales entre sí.

Puede asignar un parámetro al valor de umbral. Establezca un valor decimal entre 0 y 1.

Configuración de la propiedad de coincidencia

Utilice la propiedad **Coincidencia** en la vista **Salida de coincidencia** para especificar cómo la transformación selecciona los datos de entrada para el análisis. Configure la propiedad al configurar la transformación de coincidencia para leer un almacén persistente de datos de índice. La propiedad de coincidencia ajusta las opciones que establece en la vista **Tipo de coincidencia**.

Puede configurar la propiedad de coincidencia para realizar los siguientes tipos de análisis:

Comparar los registros del origen de datos con los registros de datos de índice

Para buscar registros duplicados entre el origen de datos y las tablas de datos de índice, seleccione **Exclusivo**.

Cuando se selecciona la opción **Exclusivo**, la transformación de coincidencia compara los registros del origen de datos con el almacén de datos de índice. La transformación no analiza los registros en el origen de datos ni en el almacén de datos.

Seleccione la opción **Exclusivo** si sabe que el almacén de datos de índice no contiene registros duplicados y sabe que el origen de datos no contiene registros duplicados.

Comparar los registros del origen de datos con los registros de datos de índice y comparar los registros del origen de datos entre sí

Para buscar duplicados en el origen de datos y también entre el origen de datos y las tablas de índices, seleccione **Parcial**.

La transformación compara los registros del origen de datos con el almacén de datos de índice. La transformación también compara los registros en el origen de datos entre sí.

Seleccione la opción **Parcial** si sabe que el almacén de datos de índice no contiene registros duplicados, pero no ha realizado ningún análisis de duplicados en el origen de datos.

Comparar todos los registros en el origen de datos y en las tablas de índices como un único conjunto de datos

Para buscar duplicados entre el origen de datos y las tablas de índices y buscar duplicados dentro del origen de datos y dentro de las tablas de índices, seleccione **Completo**. La opción predeterminada es completo.

La transformación analiza el origen de datos y el almacén de datos como un solo conjunto de datos y compara todos los datos de los registros en el conjunto de datos.

Seleccione la opción **Completo** cuando no pueda comprobar si un conjunto de datos está libre de registros duplicados.

Configuración de la propiedad de salida

Utilice la propiedad **Salida** en la vista **Salida de coincidencia** para filtrar los registros que la transformación escribe como salida. Configure la propiedad cuando especifique tablas de datos de índice y seleccione un formato de salida agrupada. Filtre los registros para limitar la salida a clústeres que contienen uno o más registros del origen de datos.

Puede filtrar los datos de salida de las siguientes maneras:

Escribir todos los clústeres que incluyan un registro del origen de datos o de las tablas de índices

Seleccione **Todas las filas**. La transformación escribe todos los clústeres que contengan al menos un registro desde el origen de datos o del almacén de datos de índice. El valor predeterminado es Todas las filas.

Como un clúster puede contener un único registro, la salida contiene todos los registros.

Escribir todos los clústeres que incluyan un registro del origen de datos

Seleccione **Filas nuevas y asociadas**. La transformación escribe todos los clústeres que contengan al menos un registro del origen de datos.

Como un clúster puede contener un único registro, la salida contiene todos los registros en el origen de datos. Los clústeres también pueden incluir registros de las tablas de índices.

Escribir todos los clústeres desde el origen de datos

Seleccione **Solo filas nuevas**. La transformación escribe los clústeres que contienen registros del origen de datos. La salida no contiene ningún registro de las tablas de índices.

Propiedades avanzadas de coincidencia de identidades

La transformación incluye las propiedades avanzadas que determinan el número de instancias de ejecución, si la transformación analiza filas idénticas y el nivel de seguimiento de los datos de registro.

Puede configurar las siguientes propiedades avanzadas:

Instancias de ejecución

Especifica el número de subprocesos que el Servicio de integración de datos intenta crear para la transformación actual en tiempo de ejecución. El Servicio de integración de datos tiene en cuenta el valor de Instancias de ejecución si se anula la propiedad en tiempo de ejecución Número máximo de paralelismos en la asignación que contiene la transformación. El valor predeterminado de Instancias de ejecución es Automático.

El Servicio de integración de datos tiene en cuenta varios factores a la hora de determinar el número de subprocesos que debe asignar a la transformación. Los principales factores son el valor de Instancias de ejecución y los valores en la asignación y en los servicios de aplicación asociados del dominio.

El Servicio de integración de datos que ejecuta la asignación lee los siguientes valores para saber el número de subprocesos que debe usar para la transformación:

- El valor de *Número máximo de paralelismos* del Servicio de integración de datos. El valor predeterminado es 1.
- Cualquier valor de *Número máximo de paralelismos* que se haya establecido en el nivel de asignación. El valor predeterminado es Automático.
- El valor de *Instancias de ejecución* en la transformación. El valor predeterminado es Automático.

Si reemplaza el valor de Número máximo de paralelismos en el nivel de asignación, el Servicio de integración de datos intentará usar el valor más bajo de las propiedades para determinar el número de subprocesos.

Si utiliza el valor predeterminado de Número máximo de paralelismos en el nivel de asignación, el Servicio de integración de datos omitirá el valor de Instancias de ejecución.

Tenga en cuenta las siguientes reglas y directrices cuando establezca el número de instancias de ejecución:

- Puede que varios usuarios ejecuten asignaciones simultáneas en un Servicio de integración de datos. Para calcular el número correcto de subprocesos, divida el número de unidades de procesamiento central a las que puede acceder el Servicio de integración de datos entre el número de asignaciones simultáneas.

- Cuando se utiliza el valor predeterminado de Instancias de ejecución y los valores predeterminados de Número máximo de paralelismos, las operaciones de transformación no admitirán particiones.

Filtrar coincidencia exacta

Determina si la transformación aplica el algoritmo de comparación de una estrategia de coincidencia a pares de registros idénticos en los datos de entrada.

Cuando la transformación encuentra un par de registros idénticos, el algoritmo no necesita analizar los niveles de similitud entre los registros. La transformación puede transferir los registros directamente a la fase de salida sin análisis adicionales. Para configurar la transformación para que transfiera los registros idénticos directamente a la fase de salida, seleccione Filtrar coincidencia exacta. Cuando los datos de entrada contienen muchas filas idénticas, el algoritmo de comparación realiza menos cálculos y la asignación se ejecuta más rápido.

Seleccione la opción cuando los datos de entrada contengan muchas filas idénticas. No seleccione esta opción si los datos de entrada no contienen muchas filas idénticas, porque la transformación se ejecutará con más lentitud.

Nota: La salida de transformación contiene los mismos datos de registro cuando se selecciona o se desactiva la opción. Puede que la transformación asigne diferentes puntuaciones de enlace y de controlador a los registros de salida al seleccionar y desactivar la opción.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Caso práctico de índice persistente

Usted es gestor de datos en una entidad bancaria con diversas sucursales. Administra un conjunto principal de registros de cuentas de clientes de todas las sucursales. Utiliza un conjunto de tablas de base de datos de índice para comprobar que la base de datos de cuentas de clientes no contiene registros redundantes ni duplicados.

Para crear y administrar el almacén de datos de índice, puede realizar las operaciones siguientes:

- Crea el almacén de datos.
- Actualiza el almacén de datos con los datos más recientes de las sucursales bancarias.

Puede agregar datos de cuentas al almacén de datos o puede actualizar los datos actuales del almacén de datos.

- Quita registros obsoletos del almacén de datos.

Es consciente de que cada operación podría crear registros duplicados en el almacén de datos. Decide desarrollar una política para analizar los datos de sucursales antes de agregarlos a los datos del almacén de datos principal. Utiliza el análisis de coincidencia de identidades para analizar los datos de sucursales y para comprobar que los datos no crean identidades duplicadas en el almacén de datos. Configura las opciones de índice persistente en la transformación de coincidencia para analizar los datos de sucursales y el almacén de datos.

Desarrollar una política para la administración de datos de índice persistentes

Como gestor de datos, define una regla empresarial que indica que el almacén de datos de cuentas de clientes no puede contener identidades duplicadas. Diseña asignación de coincidencia de identidad para

analizar los datos de sucursales en una base de datos de ensayo antes de agregar los datos al almacén de datos.

Las operaciones para agregar los datos de sucursales al almacén de datos pueden crear identidades duplicadas en los casos siguientes:

- Los datos de sucursales contienen identidades duplicadas.
- Los datos de sucursales contienen una identidad que el índice también contiene.
- Los datos de sucursales contienen una versión más reciente de una identidad en el almacén de datos y la versión más reciente coincide con otra identidad en el índice.

Al comparar la base de datos de ensayo con el almacén de datos, seleccione las opciones de índice persistente que reflejan el estado de registro duplicado de los datos de sucursales. Antes de actualizar el almacén de datos, puede decidir comparar los datos de sucursales con los datos de índice.

Nota: Puede activar y desactivar el análisis de coincidencia en algunas de las opciones. Habilite el análisis de coincidencia para analizar los datos de la asignación o para comparar el almacén de datos de índice con los datos de la asignación. Desactive el análisis de coincidencia cuando no necesite comparar los datos. También puede utilizar las propiedades de coincidencia de la ficha Salida de coincidencia para incluir o excluir los datos del análisis de coincidencia.

Comparar un origen de datos de asignación con el almacén de datos de índice

Para comparar los datos de entrada de la asignación con el almacén de datos de índice sin realizar cambios en el almacén de datos, seleccione la opción siguiente:

- No actualizar la base de datos

La asignación compara los datos de entrada con el almacén de datos de índice. La asignación no añade, quita ni actualiza datos del almacén de datos de índice.

No puede desactivar el análisis de coincidencia de identidad si selecciona la opción.

Dado que no actualiza los datos de índice, no puede crear filas duplicadas en el almacén. Seleccione la opción de las propiedades de coincidencia de la ficha Salida de coincidencia que cumpla las necesidades actuales del proyecto de datos. Por ejemplo, seleccione la opción **Completo**. La opción **Completo** comprueba que los datos de asignación no contienen duplicados y verifica que los datos de asignación no agregan duplicados al almacén de datos.

Nota: Utilice la opción para comparar los datos de asignación y el almacén de datos antes de actualizar el almacén de datos. Si la salida de asignación indica que los datos de asignación no agregan duplicados al almacén de datos, ejecute la asignación de nuevo. Seleccione la opción para actualizar la base de datos al ejecutar la asignación de nuevo.

Crear el almacén de datos y agregar filas al almacén de datos

Para crear un almacén de datos o agregar filas de los datos de asignación a un almacén de datos, seleccione la opción siguiente:

- Actualizar la base de datos con ID nuevos

La asignación agrega una fila al almacén de datos si la fila no comparte un identificador de secuencia con una fila del almacén de datos. La asignación no sobrescribe ninguna fila en las tablas de índices. Al especificar tablas de base de datos vacías, la asignación escribe todos los datos de índice de la asignación en las tablas.

Puede activar o desactivar el análisis de coincidencia de identidad si selecciona la opción. La opción habilita el análisis de coincidencia de forma predeterminada.

Dado que no actualiza las filas de índice, seleccione la opción **Exclusivo** o la opción **Parcial** en las propiedades de coincidencia en la ficha Salida de coincidencia. Utilice la opción **Exclusivo** si ha comprobado la exclusividad de las filas de datos de asignación en un proceso anterior.

Actualizar las filas en el almacén de datos

Para actualizar una fila actual en el almacén de datos con los datos de asignación, seleccione la opción siguiente:

- Actualizar los identificadores en la base de datos

La asignación actualiza un registro actual en el almacén de datos si el registro comparte un identificador de secuencia con un registro de los datos de la asignación. La asignación no añade ninguna fila a las tablas de índices.

Puede activar o desactivar el análisis de coincidencia de identidad si selecciona la opción. La opción desactiva el análisis de coincidencia de forma predeterminada.

Dado que no añade filas de índice a las tablas de índices, seleccione la opción **Completo** de las propiedades de coincidencia en la ficha Salida de coincidencia.

Nota: Al actualizar las filas en el almacén de datos, espera encontrar duplicados entre los datos de origen de la asignación y el almacén de datos. Seleccione la opción **Completo** para comprobar que los datos de identidad que añade al almacén no coinciden con los datos actuales del almacén.

Eliminar filas en el almacén de datos

Para eliminar las filas del almacén de datos, seleccione la opción siguiente:

- Quitar los ID de la base de datos

La asignación elimina una fila del almacén de datos si la fila comparte un identificador de secuencia con un registro de los datos de la asignación.

Puede activar o desactivar el análisis de coincidencia de identidad si selecciona la opción. La opción desactiva el análisis de coincidencia de forma predeterminada.

Nota: Al quitar datos de un almacén de datos, cambie las relaciones entre las filas en el almacén. Si el almacén contiene identidades duplicadas, puede quitar datos de un registro del controlador o un registro vinculado en un clúster. También puede eliminar datos de la mejor coincidencia de un par de coincidencia. Cuando ejecute la asignación de nuevo, la asignación podría generar diferentes pares duplicados o clústeres. Si quita filas de un almacén de datos que no contiene registros duplicados, no puede cambiar el estado duplicado de los registros. Cuando ejecute la asignación después de eliminar las filas, la asignación genera las mismas puntuaciones de coincidencia para las identidades que permanecen en el conjunto de datos.

Ejemplo de análisis de coincidencia de identidades

Usted es un oficial de recursos humanos en una empresa de software con centros de desarrollo en diferentes ciudades. La organización almacena registros del personal en una base de datos en su sede central. Los centros de desarrollo contratan personal a intervalos regulares y le envían a usted los datos del personal que contratan.

Usted añade los registros del personal a un archivo de hoja de cálculo y usa los datos del archivo para actualizar la base de datos de empleados. Le preocupa que el archivo actual pueda contener identidades duplicadas.

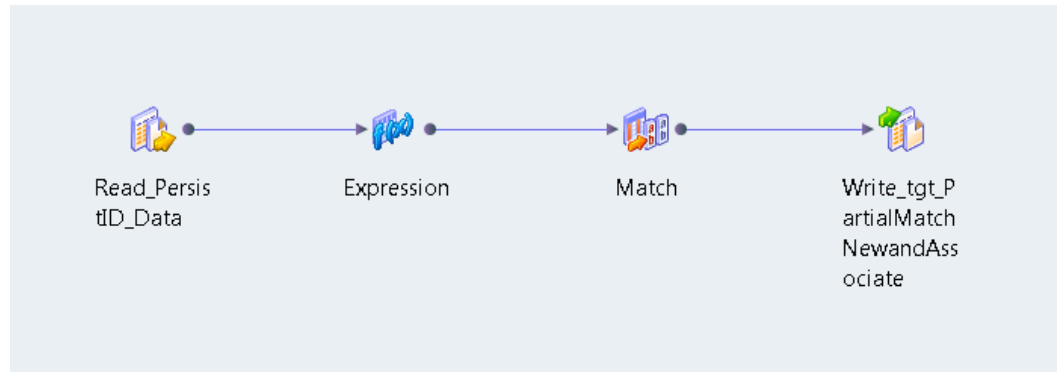
Diseña una asignación para realizar análisis de identidad en los registros de empleados. Configura una transformación de coincidencia para buscar identidades duplicadas en el archivo de hoja de cálculo. También debe comprobar que los datos del archivo no duplican los datos de empleados de la base de datos principal. Configura la transformación de coincidencia para comparar los datos del archivo con los datos principales que almacena para los empleados de la organización.

Dado que la base de datos es un conjunto de datos principal, almacena los datos de índice de los registros del personal en un almacén de datos persistente.

Crear la asignación

Cree una asignación que busque identidades duplicadas. La asignación lee un origen de datos, añade un identificador de secuencia a los registros de origen, realiza el análisis de identidad y escribe los resultados en un destino de datos.

La siguiente imagen muestra la asignación en la herramienta Developer:



La asignación que cree contiene los siguientes objetos:

Nombre de objeto	Descripción
Read_PersistID_Data	Origen de datos. Contiene los nombres y los detalles de los empleados.
Expresión	Expresión transformación. Añade valores de identificador de secuencia a los datos de origen.
Coincidencia	Transformación de coincidencia. Analiza los niveles de duplicación en las identidades de datos de origen.
Write_tgt_PartialMatchNewandAssociate	Destino de datos. Contiene los resultados del análisis de identidad.

Nota: La asignación no utiliza una transformación de generador de claves. En el análisis de coincidencia de identidades, la transformación de generador de claves es opcional.

Muestra de datos de entrada

El archivo de personal contiene el nombre del empleado, la ciudad en la que el empleado trabaja y la función designada del empleado. Cree un origen de datos desde el archivo de personal en el repositorio de modelos. Puede añadir el origen de datos a la asignación.

El siguiente fragmento de datos muestra un ejemplo de los datos de empleados en el archivo de personal:

Nombre	City	Designación
Chaithra	Bangalore	SE
Ramanan	Chennai	SSE
Ramesh	Chennai	SSE
Ramesh	Chennai	Director
Sunil	Bangalore	Principal
Venu	Hyderabad	Principal
Harish	Bangalore	SE
Sachin	Bangalore	SSE

Configuración de la transformación de expresión

Cuando configure la transformación de expresión, conecte todos los puertos de origen de datos que desea incluir en la salida de asignación. Conecte los puertos como puertos de transferencia. Cree una expresión para añadir un valor de identificación de secuencias a los puertos.

La siguiente expresión crea la variable *Init3* y añade el valor entero 1267 a cada identificador de secuencia:

Init3+1267

La siguiente tabla describe los puertos en la transformación de expresión que lee el origen de datos de empleados:

Nombre	Tipo de puerto	Grupo de puertos
SEQID	bigint	Solo salida
Nombre	string	Transferencia
City	string	Transferencia
Designación	string	Transferencia
Init3	integer	Variable

Configuración de la transformación de coincidencia

Añada una transformación de coincidencia no reutilizable a la asignación para realizar el análisis de identidad.

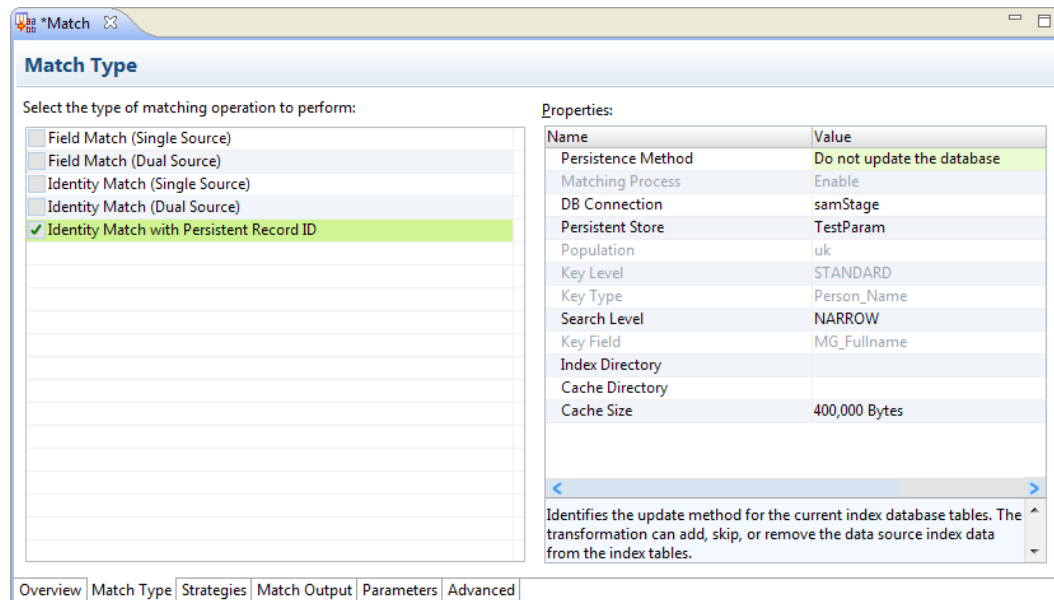
Realice las tareas siguientes para configurar la transformación:

1. Seleccione el tipo de análisis de coincidencia que desea realizar.
2. Conecte los puertos de entrada a la transformación.
3. Configure una estrategia para comparar los datos de registros.
4. Seleccione el tipo de datos de salida de coincidencia que crea la transformación.
5. Conecte los puertos de salida a un destino de datos.

Seleccionar el tipo de operación de coincidencia

Utilice las opciones en la vista Tipo de coincidencia para seleccionar la operación de coincidencia.

La siguiente imagen muestra la vista Tipo de coincidencia:



Para comparar los datos de índice del origen de datos con los datos de índice del conjunto de datos principal, seleccione **Coincidencia de identidad con ID de registro persistente**. Actualice el método de persistencia para que el análisis de coincidencia no añada datos a las tablas de índices. Puede determinar si actualizar las tablas de índices después de revisar los resultados de la asignación.

Utilice la opción **Conexión de base de datos** para identificar la base de datos que contiene los datos de índices. Utilice la opción **Almacenamiento persistente** para seleccionar las tablas de índices.

Nota: La transformación de coincidencia lee los valores de propiedad de los campos de población de identidad, nivel de clave, tipo de clave y clave de los metadatos de las tablas de base de datos de índices. Los valores coinciden con las propiedades correspondientes de la transformación que creó el almacén de datos de índice.

Conectar los puertos de entrada

Conecte los puertos de datos de entrada a la transformación. Compruebe que los nombres de los puertos de entrada, el orden de los puertos, los tipos de datos y las precisiones coincidan con las configuraciones de puerto de la transformación que ha creado el almacén de datos.

La transformación de coincidencia utiliza puertos de entrada preestablecidos para determinar el orden en que procesa los registros. La transformación utiliza un identificador de secuencia para hacer un seguimiento de los registros desde los puertos de entrada hasta los clústeres o pares de coincidencia que escribe como salida. La transformación utiliza una clave de grupo para ordenar los registros que procesa.

Conecte los puertos preestablecidos a los siguientes puertos en la transformación de expresión:

- SequenceID. Conecte con el puerto SEQID desde la transformación de expresión.
- GroupKey. Conecte con el puerto Ciudad desde la transformación de expresión.

Configurar una estrategia para el análisis de identidad

Utilice las opciones en la vista Estrategias para configurar una estrategia. La estrategia determina el tipo de análisis que la transformación realiza en los datos de registros.

Seleccione el algoritmo Person_Name para los datos del registro. Seleccione el puerto de entrada Nombre para el análisis. Como la transformación crea una copia de los datos de puerto, seleccione el puerto Name_1 y el puerto Name_2.

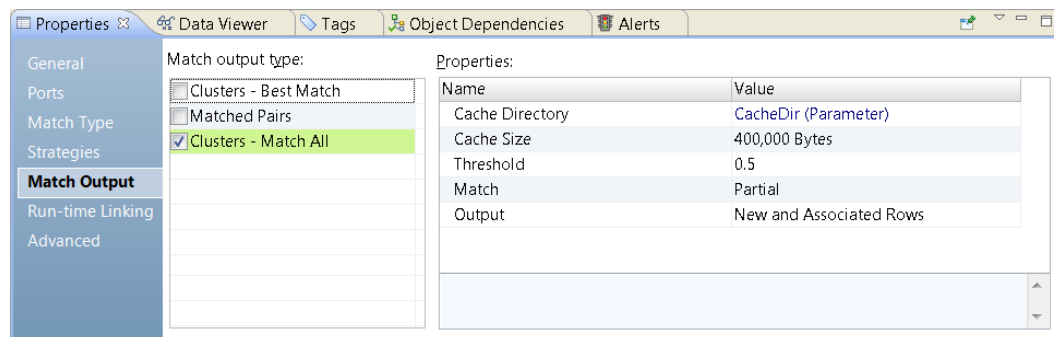
TEMAS RELACIONADOS

- [“Algoritmos de coincidencia de identidades” en la página 522](#)

Seleccionar el tipo de salida de coincidencia

Usar las opciones en la vista Salida de coincidencia para definir el formato de salida para los resultados del análisis de coincidencia

La siguiente imagen muestra la vista Salida de coincidencia para un análisis de identidad de origen único:



Configure la transformación para organizar los registros de salida en clústeres. Cada clúster contiene todos los registros que coinciden con al menos otro registro, en función de las propiedades de coincidencia que especifique. Las propiedades de coincidencia determinan cómo compara la transformación los registros del origen de datos con los registros de índice.

La siguiente tabla describe las opciones de la propiedad de coincidencia que especifique para analizar los datos de registros del personal:

Propiedad de coincidencia	Opción	Descripción de la opción
Coincidencia	Parcial	La transformación compara los registros del origen de datos con el almacén de datos de índice. La transformación también compara los registros en el origen de datos entre sí.
Salida	Filas nuevas y asociadas	La transformación escribe todos los clústeres que contengan al menos un registro del origen de datos. Los clústeres pueden incluir registros del almacén de datos de índice. Como un clúster puede contener un único registro, la salida contiene todos los registros en el origen de datos.

TEMAS RELACIONADOS

- [“Configuración de la propiedad de coincidencia” en la página 526](#)
- [“Configuración de la propiedad de salida” en la página 526](#)

Conectar los puertos de salida

Conecte los puertos de salida de la transformación de coincidencia con el destino de datos en la asignación. Seleccione los puertos que contienen los datos de registro que desea escribir en el destino de datos.

La transformación incluye una serie de puertos preestablecidos para datos en clúster. Seleccione los puertos preestablecidos que indican estados duplicados de los registros e identifique el origen de datos que almacena cada registro.

Los siguientes puertos contienen datos que puede usar para buscar registros duplicados y determinar el origen o los registros:

- El puerto **ClusterSize** indica el número de registros en un clúster. Si un registro pertenece a un clúster con un tamaño de clúster mayor que 1, la transformación considera que el registro es un duplicado de otro registro.
- El puerto **ClusterID** identifica el clúster al que pertenece un registro. Utilice los datos de ClusterID para buscar los registros que son duplicados del registro actual.
- El puerto **PersistenceStatus** utiliza un valor de código para describir la relación entre los datos de índice del origen de asignación y los datos de índice del almacén de datos.
- El puerto **PersistenceStatusDesc** devuelve una descripción de texto de los valores del código de puerto PersistenceStatus.

Puede utilizar otros puertos para revisar las relaciones entre los registros del clúster. Los valores de puerto de enlace y los valores de puerto de controlador indican el alcance de la similitud entre los registros en cada clúster.

En el ejemplo, puede conectar todos los puertos con el destino de datos. Para ver los datos de salida en los puertos, ejecute el visor de datos.

TEMAS RELACIONADOS

- [“Códigos de estado de persistencia y descripciones de estado de persistencia” en la página 496](#)
- [“Valores de código de estado y valores de descripción de estado” en la página 497](#)

Ejecutar el visor de datos

Ejecute el visor de datos para revisar los resultados del análisis de coincidencia. De forma predeterminada, el visor de datos muestra todos los puertos de salida en la transformación de coincidencia. Cuando se ejecuta la asignación, puede actualizar el destino de datos con los datos procedentes de los puertos de salida.

La siguiente imagen muestra los datos de salida en el visor de datos:

The screenshot shows a data flow diagram with four components: 'Read_IncreP', 'Expression', 'Match', and 'Write_tgtP'. Below the diagram is a 'Data Viewer' window showing a table of output data. The table has 16 rows and 10 columns. The columns are: Name, ClusterSize, PersistenceStatusDesc, Designation, GroupKey, LinkId, ClusterId, LinkScore, and PersistenceSt.. The data is as follows:

Name	ClusterSize	PersistenceStatusDesc	Designation	GroupKey	LinkId	ClusterId	LinkScore	PersistenceSt..
1 Chaith	2	Input, Exists, Ignored	SE	Bangalore	S - 1	000000001	1	IEI00000
2 Chaith	2	Store, -, No change	SE	Bangalore	1 - 1267	000000001	1	SON00000
3 Ramana	2	Input, Exists, Ignored	SSE	Chennai	S - 2	000000002	1	IEI00000
4 Ramana	2	Store, -, No change	SSE	Chennai	1 - 2534	000000002	1	SON00000
5 Ramesh	4	Input, Exists, Ignored	SSE	Chennai	S - 3	000000003	1	IEI00000
6 Ramesh	4	Input, Exists, Ignored	Lead	Chennai	S - 3	000000003	1	IEI00000
7 Ramesh	4	Store, -, No change	SSE	Chennai	1 - 3801	000000003	1	SON00000
8 Ramesh	4	Store, -, No change	Lead	Chennai	1 - 3801	000000003	1	SON00000
9 Sunil	2	Input, Exists, Ignored	Principal	Bangalore	S - 5	000000004	1	IEI00000
10 Sunil	2	Store, -, No change	Principal	Bangalore	1 - 6335	000000004	1	SON00000
11 Venu	2	Input, Exists, Ignored	Principal	Hydrabad	S - 6	000000005	1	IEI00000
12 Venu	2	Store, -, No change	Principal	Hydrabad	1 - 7602	000000005	1	SON00000
13 Harish	2	Input, Exists, Ignored	SE	Bangalore	S - 7	000000006	1	IEI00000
14 Harish	2	Store, -, No change	SE	Bangalore	1 - 8869	000000006	1	SON00000
15 Sachin	2	Input, Exists, Ignored	SSE	Bangalore	S - 8	000000007	1	IEI00000
16 Sachin	2	Store, -, No change	SSE	Bangalore	1 - 10136	000000007	1	SON00000

El visor de datos comprueba que el archivo contiene registros que duplican los datos en el conjunto de datos principales.

Tenga en cuenta los siguientes valores de datos en el visor de datos:

- Cada registro en el conjunto de datos pertenece a un clúster que contiene dos o más registros. Por tanto, cada registro es un duplicado de al menos otro registro. La transformación asigna un tamaño de clúster de 1 a los registros únicos. El origen de datos no contiene ningún registro que no contenga el conjunto de datos principales.
- Los datos de **PersistenceStatusDesc** identifica el origen del registro e indica si la transformación de coincidencia añade el registro a las tablas de índices. La columna indica que cada registro de entrada existe en el conjunto de datos principales. La transformación no añade datos al índice de datos principales.

Conclusión

Los resultados del análisis de coincidencia indican que el archivo de registro de personal no contiene ningún registro que el conjunto de datos principales no contenga. La persistencia de las descripciones de estado indica que la asignación no actualiza las tablas de índices con los datos desde el origen de datos. Descarte el archivo de registro de personal.

Cuando reciba otra actualización desde las oficinas regionales, puede crear otro archivo y compararlo con el conjunto de datos principales. Se puede reutilizar la asignación y las tablas de índices. Como se almacenan los datos de índice para el conjunto de datos principales en las tablas de base de datos, no necesita volver a generar los datos de índice.

CAPÍTULO 32

Transformación de normalizador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de normalizador, 538](#)
- [Campos de ocurrencia múltiple, 539](#)
- [Registros de ocurrencia múltiple, 540](#)
- [Definición de la jerarquía de entrada, 541](#)
- [Grupos y puertos de salida de la transformación de normalizador, 549](#)
- [Generación de claves para grupos de salida, 553](#)
- [Propiedades avanzadas de transformación de normalizador, 554](#)
- [Creación de una transformación de normalizador, 554](#)
- [Creación de una transformación de normalizador desde un origen de nivel superior, 555](#)
- [Ejemplo de asignación de normalizador, 556](#)
- [Transformación de normalizador en un entorno no nativo, 559](#)

Resumen de la transformación de normalizador

La transformación de normalizador es una transformación activa que transforma una fila de origen en varias filas de destino. Cuando la transformación de normalizador recibe una fila que contiene datos de ocurrencia múltiple, devuelve una fila para cada instancia de los datos de ocurrencia múltiple.

La transformación de normalizador analiza datos de ocurrencia múltiple en filas de salida individuales. Por ejemplo, una fila de origen relacional podría contener cuatro trimestres de ventas. La transformación de normalizador genera una fila de salida individual para cada ocurrencia de ventas.

Cuando defina la transformación de normalizador, configure una jerarquía de filas de entrada que describa la estructura de los datos de origen. Opcionalmente, puede definir los registros en la jerarquía de entrada de transformación. Un registro es un contenedor de un grupo de campos. Defina un registro cuando un grupo de campos ocurra varias veces en los datos de origen. La jerarquía de entrada determina cómo se pueden configurar los grupos de salida de transformación.

La transformación de normalizador transforma los datos de tablas relacionales o de orígenes de archivo sin formato.

Campos de ocurrencia múltiple

Cuando un campo se repite varias veces en los datos de origen, puede definirlo como un campo de ocurrencia múltiple en la jerarquía de filas de entrada. La transformación de normalizador puede devolver una fila diferente para cada ocurrencia de un campo de ocurrencia múltiple o de un grupo de campos en un origen.

Una fila de origen puede contener cuatro trimestres de ventas por cada tienda:

Tienda	Ventas (1)	Ventas (2)	Ventas (3)	Ventas (4)
Tienda1	100	300	500	700
Tienda2	250	450	650	850

Al definir la jerarquía de entrada del normalizador, puede combinar los cuatro campos Ventas en un único campo de ocurrencia múltiple. Defina un nombre de campo (por ejemplo, Qtr_Sales) y configúrelo de manera que tenga cuatro iteraciones en el origen.

Si el grupo de salida contiene los datos de la tienda y los datos de ventas, la transformación de normalizador devuelve una fila para cada combinación de Tienda y Qtr_Sales. La fila de salida contiene un índice que identifica la instancia de Qtr_Sales que está en la fila de salida.

La transformación devuelve las siguientes filas:

Tienda	Qtr_Sales	Qtr (GCID)
Tienda1	100	1
Tienda1	300	2
Tienda1	500	3
Tienda1	700	4
Tienda2	250	1
Tienda2	450	2
Tienda2	650	3
Tienda2	850	4

Cuando un grupo de salida contiene columnas de ocurrencia individual y una columna de ocurrencia múltiple, el normalizador devuelve datos duplicados para las columnas de ocurrencia individual de cada fila de salida. Por ejemplo, Tienda1 y Tienda2 se repiten en cada instancia de Qtr_Sales.

Una fila de origen podría contener más de un nivel de datos de ocurrencia múltiple. Puede configurar la transformación de normalizador para devolver filas independientes en cada nivel según la definición de la jerarquía de entrada.

ID de columna generados

La transformación de normalizador devuelve un puerto de salida de ID de columna generado (GCID) para cada instancia de un campo de ocurrencia múltiple.

El puerto de ID de columna generado es un índice para la instancia de los datos de ocurrencia múltiple. Por ejemplo, si un campo ocurre cuatro veces en un registro de origen, la herramienta del desarrollador devuelve un valor de 1, 2, 3 o 4 en el puerto de ID de columna generado según sea la instancia de los datos de ocurrencia múltiple que aparece en la fila.

Registros de ocurrencia múltiple

Puede definir registros de ocurrencia múltiple en los datos de origen de la transformación de normalizador. Los registros son grupos de campos. Defina los registros de la transformación de normalizador cuando tenga que definir grupos de campos de origen que sean de ocurrencia múltiple.

Ejemplo de registros de ocurrencia múltiple

La siguiente fila Customer (cliente) contiene información de clientes con información de dirección de domicilio e información de dirección comercial:

```
CustomerID
FirstName
LastName
Home_Street
Home_City
Home_State
Home_Country
Business_Street
Business_City
Business_State
Business_Country
```

Cuando configure la transformación de normalizador, podrá definir una estructura de entrada que contenga los campos de clientes y un registro de direcciones de ocurrencia múltiple. El registro de direcciones se produce dos veces. Cuando configure los grupos de salida de la transformación de normalizador, podrá devolver el registro Address (dirección) a un destino diferente que los campos CustomerID, FirstName y LastName.

El siguiente ejemplo muestra una estructura de entrada con un registro de direcciones de ocurrencia múltiple:

```
CustomerID
FirstName
LastName
Address (occurs twice)
  Street
  City
  State
  Country
```

Los subregistros son registros dentro de registros. Cuando defina registros y subregistros, podrá definir una jerarquía de entrada de campos en la fila de origen. Cada registro es un nodo de una jerarquía a la que puede hacer referencia cuando defina la salida de transformación.

Por ejemplo, la fila de origen podría contener varios números de teléfono para cada tipo de dirección:

```
CustomerID
FirstName
LastName
Home_Street
Home_City
Home_State
Home_Country
Telephone_No
Cell_Phone_No
Alternate_Phone_No
Business_Street
Business_City
Business_State
Business_Country
Business_Telephone_No
Business_Cell_Phone_No
Business-Alternate_Phone1
```

Define una jerarquía de entrada en la que Address es el elemento principal de Phone. Cuando defina la salida de la transformación de normalizador, podrá devolver las direcciones y los números de teléfono para separar los destinos de la información de clientes.

Defina una jerarquía de entrada similar a la del siguiente ejemplo:

```
CustomerID
FirstName
LastName
Address (occurs twice)
  Street
  City
  State
  Country
Phone
  Telephone_No (occurs three times)
```

Definición de la jerarquía de entrada

Cuando cree una transformación de normalizador, defina una jerarquía de entrada que describa los registros y los campos en el origen. Defina la jerarquía de entrada en la vista **Normalizador** de la transformación.

Developer tool crea los puertos de entrada de transformación en función de los campos definidos en la jerarquía de entrada. Defina la estructura de grupo de entrada antes de definir los grupos de salida de la transformación.

Cuando defina una jerarquía de entrada, deberá definir una estructura de entrada que corresponda a la estructura de los datos de origen. Los datos de origen podrían contener más de un grupo de campos de ocurrencia múltiple. Para definir la estructura, puede configurar un registro que se produzca en el mismo nivel que otro registro del origen. O bien, puede definir los registros que se producen dentro de otros registros.

Ejemplo de jerarquía de entrada

La siguiente fila de origen contiene campos de cliente y un registro de direcciones que se produce dos veces:

```
CustomerID
  FirstName
  LastName
  Address
    Street
    City
    State
    Country
  Address1
    Street1
    City1
    State1
    Country1
```

Cuando defina la estructura de entrada en la vista **Normalizador**, podrá añadir CustomerID, FirstName y LastName como campos. Defina un registro Address e incluya los campos Street, City, State y Country en la dirección. Cambie el valor de Ocurrencias de Address a 2.

La siguiente imagen muestra la jerarquía de entrada en la vista **Normalizador**:

Normalizer						
Name	Level	Occurs	Type	Precision	Scale	
CustomerID	1	1	string	10	0	
FirstName	1	1	string	10	0	
LastName	1	1	string	10	0	
Address	1	2				
Street	2	1	string	10	0	
City	2	1	string	10	0	
State	2	1	string	10	0	
Country	2	1	string	10	0	

La columna **Ocurrencias** de la vista **Normalizador** identifica el número de instancias de un campo o registro de una fila de origen. Cambie el valor de la columna **Ocurrencias** para los campos o registros de ocurrencia múltiple. En este ejemplo, los campos de cliente aparecen una vez y el registro Address aparece dos veces.

La columna **Nivel** de la vista **Normalizador** indica el lugar donde aparece un campo o un registro en la jerarquía de entrada. Los campos de cliente están en el nivel 1 de la jerarquía. El registro Address también es de nivel 1.

Puertos de entrada de la transformación de normalizador

La herramienta del desarrollador crea los puertos de entrada de la transformación de normalizador cuando se define la jerarquía de entrada en la vista **Normalizador**. Cuando cambie campos en la jerarquía de entrada, la herramienta del desarrollador cambiará los puertos de entrada.

Vea los puertos de entrada de la transformación de normalizador en la vista **Resumen**. Puede cambiar el orden de los puertos de entrada en la vista **Resumen**. Para cambiar los puertos de entrada, actualice la jerarquía de entrada en la vista **Normalizador**.

Cuando defina un campo como de ocurrencia múltiple en la jerarquía de entrada, la herramienta del desarrollador crea un puerto de entrada para cada instancia del campo de ocurrencia múltiple. Cuando un registro es de ocurrencia múltiple, la herramienta del desarrollador crea un puerto de entrada para cada instancia de los campos del registro.

Ejemplo de puertos de entrada

La siguiente imagen muestra los puertos de entrada que la herramienta del desarrollador crea para los datos de clientes y los datos de direcciones de ocurrencia múltiple:

Ports					
	Name	Type	Precision	Scale	Location
[-] Input					
1	CustomerID	string	10	0	CustomerID
2	FirstName	string	10	0	FirstName
3	LastName	string	10	0	LastName
4	Street	string	10	0	Address.Street
5	Street1	string	10	0	Address.Street
6	City	string	10	0	Address.City
7	City1	string	10	0	Address.City
8	State	string	10	0	Address.State
9	State1	string	10	0	Address.State
10	Country	string	10	0	Address.Country
11	Country1	string	10	0	Address.Country

Fusionar campos

Puede combinar campos de datos similares en un único campo de ocurrencia múltiple en la vista **Normalizador**. Es posible que necesite fusionar campos cuando arrastre puertos desde otro objeto para crear una transformación de normalizador en una asignación.

Una fila de origen podría contener varios campos con distintos tipos de datos de sueldo, como Base_Salary, Bonus_Pay y Sales_Commissions. Puede combinar los campos para crear un campo Salary que se produce tres veces.

La siguiente imagen muestra una fila de empleado con tres tipos de sueldos seleccionados en la vista Normalizador:

Name	Level	Occurs	Type	Preci...	Scale
EmployeeID	1	1	string	10	0
Base_Salary	1	1	decimal	10	0
Bonus_Pay	1	1	decimal	10	0
Sales Commissions	1	1	decimal	10	0

Overview Normalizer Advanced

Puede combinar los tres tipos de datos de sueldo en un campo Salary que se produce tres veces.

La siguiente imagen muestra el campo Salary:

Normalizer						
Name	Level	Occurs	Type	Preci...	Scale	
EmployeeID	1	1	string	10	0	
Salary	1	3	decimal	10	0	

Fusionar los campos

Fusione los campos del mismo tipo en un campo de ocurrencia múltiple en la vista Normalizador.

1. Haga clic en la vista **Normalizador**.
2. Seleccione los campos que desea fusionar.
3. Haga clic en el botón **Fusionar**.
Aparece el cuadro de diálogo **Fusionar campos**.
4. Especifique un nombre para el campo de fusión, el tipo, la precisión, la escala y la ocurrencia del campo de ocurrencia múltiple.
5. Haga clic en **Aceptar**.

Aplanar campos

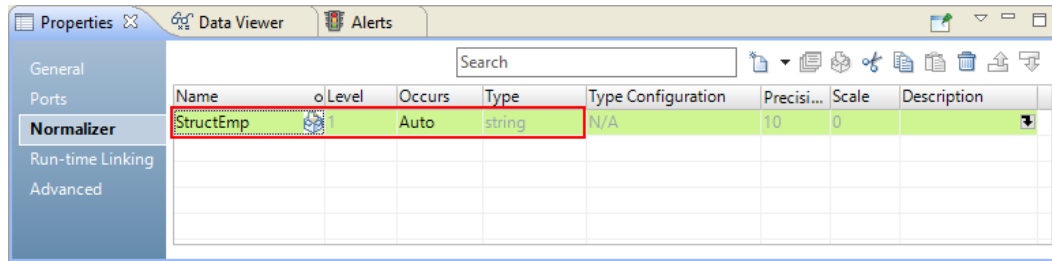
Puede aplanar los campos de un tipo de datos complejo de las asignaciones que se ejecutan en el motor de Spark. Los campos se aplanan en la vista **Normalizador** para modificar los datos jerárquicos que pasan por un puerto complejo.

La salida de la acción de aplanado depende del tipo de datos complejo. Cuando se aplanan un tipo de datos array o struct, la transformación de normalizador crea una fila por cada elemento del tipo de datos complejo. Cuando se aplanan un tipo de datos map, la transformación de normalizador crea dos columnas, una para la clave de asignación y otra para los elementos de valor de asignación.

La acción de aplanado de un tipo de datos anidado extrae elementos solo del primer nivel. Para aplanar todos los niveles de un tipo de datos anidado, use el asistente de conversiones jerárquicas **Aplanar puerto complejo** de Developer tool. La opción **Aplanar todo** extrae elementos de cada nivel y devuelve datos relacionales de tipo de datos primitivo. Para obtener más información acerca de los asistentes de conversiones jerárquicas, consulte la *Guía del usuario de Data Engineering Integration*.

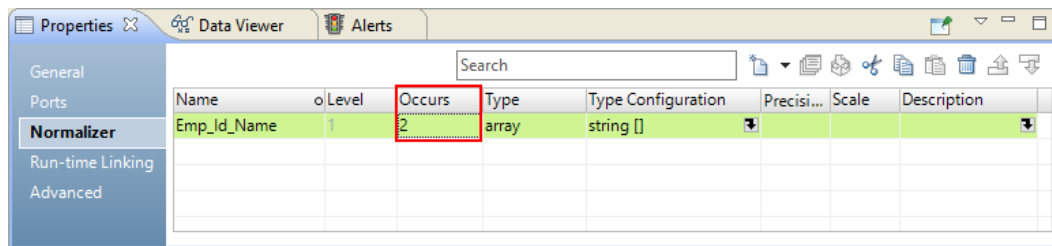
La acción de aplanado cambia el valor de la columna Ocurrencias de la vista Normalizador a Automático y añade un icono de aplanado al lado del campo aplanado. El valor Automático indica que la transformación aplanan todos los elementos del tipo datos complejo.

En la siguiente imagen se muestra una estructura aplanada a un campo de cadena con un icono de aplanado al lado y el valor de Ocurrencias establecido en Automático:



Un campo con varias ocurrencias no se puede aplanar. Por ejemplo, no se puede aplanar un campo de matriz con un valor de Ocurrencias de 2.

En la siguiente imagen se muestra un campo con varias ocurrencias de un tipo de datos array que no se puede aplanar:



Aplanar una matriz

La transformación de normalizador aplanar una matriz unidimensional a un tipo de datos primitivo y una matriz de n dimensiones a una matriz de (n-1) dimensiones. El número de filas que la transformación crea es el mismo que el tamaño de la matriz.

Por ejemplo, si se aplanar un puerto de matriz con 10 elementos de cadena, la salida devuelve 10 puertos de cadena. Si se aplanar una matriz de 3 dimensiones, la salida devuelve una matriz de 2 dimensiones.

Una tabla contiene un puerto de cadena Nombre y un puerto de matriz Teléfonos. Quiere aplanar el puerto de matriz. La tabla contiene los siguientes valores:

Name	Phones
Adams	[205-128-6478, 722-515-2889, 650-213-4020]
Jane	[650-321-4506]

Cuando el puerto de matriz se aplanar, la salida es la siguiente:

Name	Phones	GCID_Phones
Adams	205-128-6478	1
Adams	722-515-2889	2

Name	Phones	GCID_Phones
Adams	650-213-4020	3
Jane	650-321-4506	1

Puede editar el valor de la opción Ocurrencias de un campo aplanado para extraer un número específico de elementos de la matriz. El valor debe ser un entero positivo mayor que 1. El valor determina el número de elementos que se deben extraer. Por ejemplo, puede cambiar el valor de Ocurrencias a 2 para extraer los dos primeros elementos de la matriz. La salida es la siguiente:

Name	Phones	GCID_Phones
Adams	205-128-6478	1
Adams	722-515-2889	2
Jane	650-321-4506	1

Aplanar una estructura

La transformación aplanar una estructura en un campo del tipo de datos de los elementos de la estructura. Para aplanar una estructura, todos los elementos de esa estructura deben tener el mismo tipo de datos. La transformación crea una fila por cada elemento en el tipo de datos struct.

Por ejemplo, suponga que desea aplanar el siguiente campo de estructura:

```
customer_address{
  city : string
  state : string
  zip : string
}
```

La tabla contiene los siguientes valores:

Name	customer_address
Clara	{ New York NY 10032 }

Cuando el puerto de estructura se aplanar, la salida es la siguiente:

Name	customer_address	GCID_customer_address
Clara	New York	1
Clara	NY	2
Clara	10032	3

Cuando los elementos de la estructura tienen tipos de datos diferentes y al menos los dos primeros elementos tienen el mismo tipo de datos, podrá aplanar los datos de estructura de los elementos consecutivos del mismo tipo de datos. Para extraer elementos de estructura consecutivos del mismo tipo de

datos, edite el valor de Ocurrencias. El valor debe ser un entero positivo mayor que 1. Por ejemplo, una estructura emp_address contiene los siguientes elementos:

```
emp_address{
  city : string
  state : string
  zip : int
  country : string
}
```

Puede establecer el valor de Ocurrencias en 2 para extraer los elementos de estructura "city" y "state". Si establece el valor en 3 o 4, se produce un error de validación de la asignación.

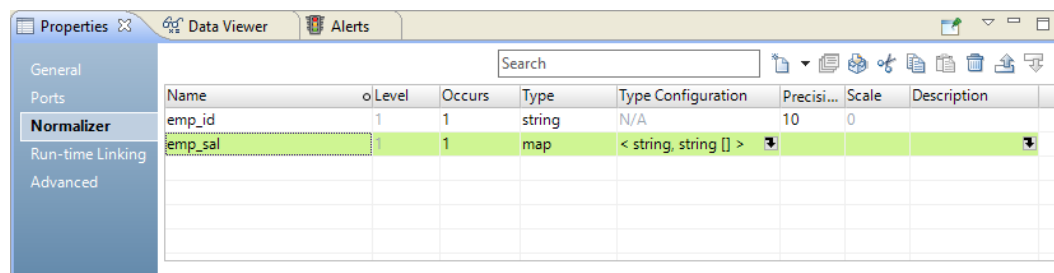
Aplanar una asignación

La transformación aplanada una asignación en dos campos para los elementos de clave y valor de la asignación. En un campo de asignación que se ha aplanado no se puede cambiar el valor de Ocurrencias de Automático a un valor de entero.

Por ejemplo, desea aplanar el siguiente campo de asignación emp_sal con una clave de cadena y una matriz de valores enteros:

```
<emp_name -> [base_sal, bonus, commision]>
```

En la siguiente imagen se muestra el campo de asignación que desea aplanar en la vista Normalizador:



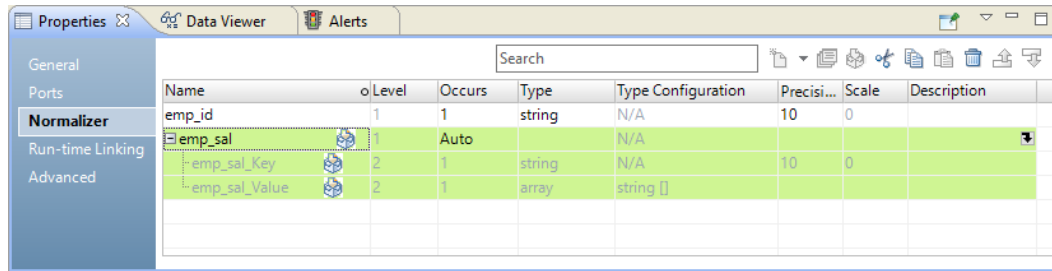
La tabla contiene los siguientes valores:

emp_id	emp_sal
12200	<Greg -> [4000, 1000, 500]>
12201	<Patricia -> [3800, 1500, 1000]>

Cuando el puerto de asignación se aplanar, la salida devuelve un campo de cadena para la clave de asignación y un campo de matriz para el valor de asignación, como se indica a continuación:

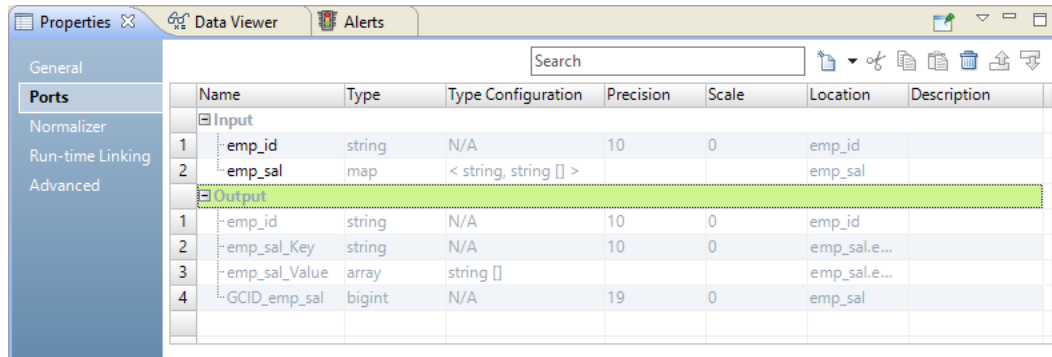
emp_id	emp_sal_Key	emp_sal_Value	GCID_emp_salary
12200	Greg	[4000, 1000, 500]	1
12201	Patricia	[3800, 1500, 1000]	1

En la siguiente imagen se muestra el campo de asignación aplanado a un campo de clave de cadena y un campo de valor de matriz en la vista Normalizador:



Name	Level	Occurs	Type	Type Configuration	Precisi...	Scale	Description
emp_id	1	1	string	N/A	10	0	
emp_sal	1	Auto		N/A			
emp_sal_Key	2	1	string	N/A	10	0	
emp_sal_Value	2	1	array	string []			

En la siguiente imagen se muestra el grupo Salida en la vista Puertos:



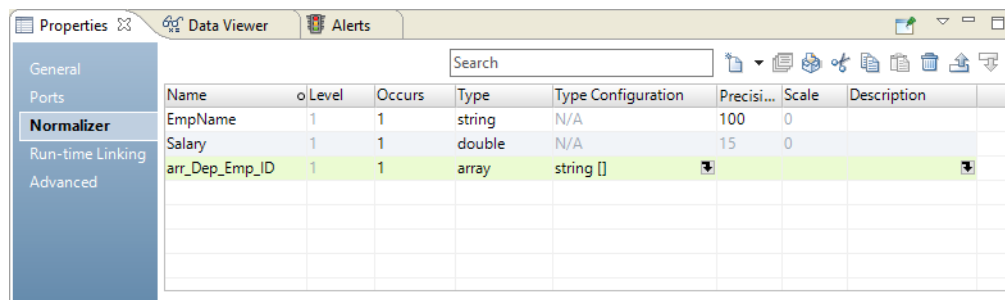
Name	Type	Type Configuration	Precision	Scale	Location	Description
Input						
1 emp_id	string	N/A	10	0	emp_id	
2 emp_sal	map	< string, string [] >			emp_sal	
Output						
1 emp_id	string	N/A	10	0	emp_id	
2 emp_sal_Key	string	N/A	10	0	emp_sal.e...	
3 emp_sal_Value	array	string []			emp_sal.e...	
4 GCID_emp_sal	bigint	N/A	19	0	emp_sal	

Aplanar campos

Aplane los campos de un tipo de datos complejo para modificar los datos jerárquicos o para convertirlos en datos relacionales.

1. Haga clic en la vista **Normalizador**.
2. Seleccione el campo de un tipo de datos complejo.

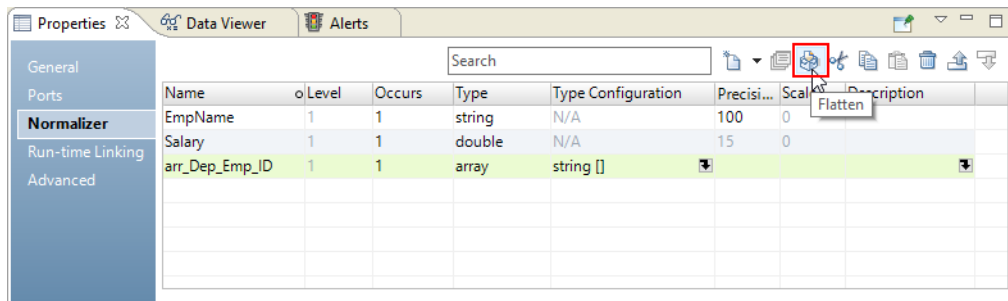
En la siguiente imagen se muestra un campo de matriz con elementos de tipo cadena:



Name	Level	Occurs	Type	Type Configuration	Precisi...	Scale	Description
EmpName	1	1	string	N/A	100	0	
Salary	1	1	double	N/A	15	0	
arr_Dep_Emp_ID	1	1	array	string []			

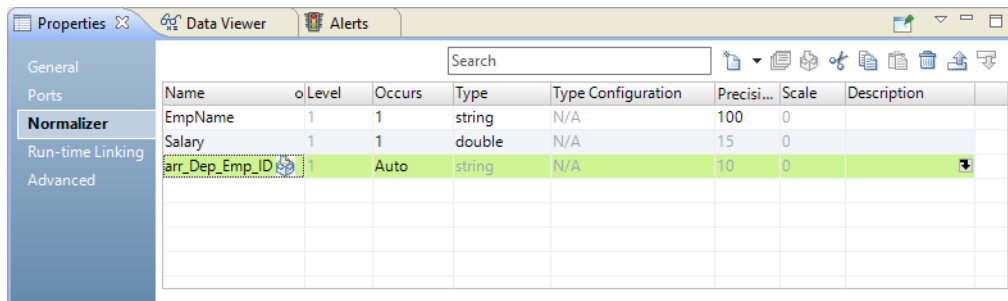
3. Haga clic en el botón **Aplanar**.

En la siguiente imagen se muestra el botón Aplanar:

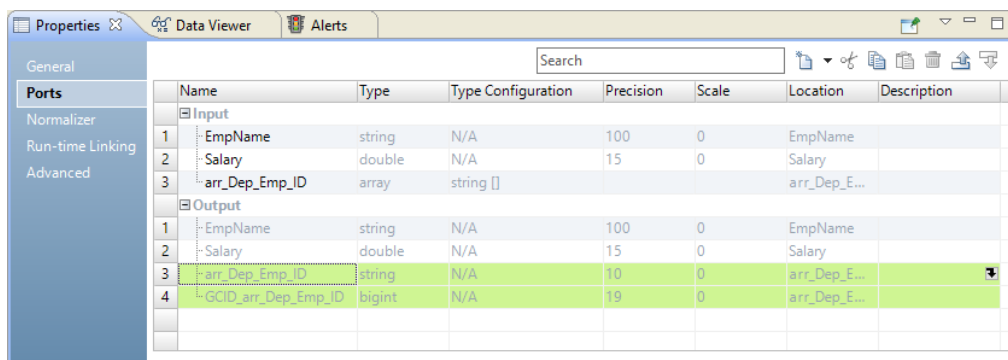


La acción Aplanar reemplaza el campo de un tipo de datos complejo por un campo aplanado y cambia el valor de **Ocurrencias** a Automático. El tipo de datos de los campos aplanados depende del tipo de datos complejo que se aplane.

En la siguiente imagen se muestra el campo aplanado de tipo string:



En la siguiente imagen se muestra el puerto de salida de cadena aplanado y el puerto de salida GCID en la vista **Puertos**:



Grupos y puertos de salida de la transformación de normalizador

Defina los grupos y puertos de salida en la vista Resumen de la transformación de normalizador. Puede definir los grupos de salida tras definir la jerarquía de entrada de transformación.

La herramienta del desarrollador genera al menos un grupo de salida de forma predeterminada. El grupo de salida contiene todos los campos de nivel 1 de los puertos de entrada y el primer campo de ocurrencia

múltiple. Al definir más de un campo de ocurrencia múltiple en la vista Normalizador, la herramienta del desarrollador crea un grupo para salida para cada campo de ocurrencia múltiple adicional.

La siguiente fila de origen contiene los datos de clientes, el campo de ventas de ocurrencia múltiple y el campo de teléfono de ocurrencia múltiple:

```
CustomerID
LastName
FirstName
Sales (occurs 4 times)
Phone (occurs 3 times)
```

La herramienta del desarrollador crea dos grupos de salida a partir de la estructura de entrada:

```
Output
CustomerID
LastName
FirstName
Sales

Output1
Phone
GCID_Phone
```

La herramienta del desarrollador crea un grupo Salida1 debido a que los datos de origen contienen más de un campo de ocurrencia múltiple. La herramienta del desarrollador no crea grupos para los campos que se definen en los registros. Al definir registros, debe definir los grupos de salida que contienen campos en los registros.

La siguiente fila de origen contiene campos de cliente y un registro de direcciones que se produce dos veces:

```
CustomerID
  FirstName
  LastName
  Address
    Street
    City
    State
    Country
  Address1
    Street1
    City1
    State1
    Country1
```

La herramienta del desarrollador crea un grupo de salida que contiene los siguientes campos.

```
CustomerID
FirstName
LastName
```

La herramienta del desarrollador crea un grupo de salida predeterminado para los campos de cliente de nivel 1. El grupo de salida predeterminado no incluye el registro de direcciones. Debe configurar la forma en que desea devolver los datos de direcciones en la salida.

Cree los grupos de salida según cómo necesite estructurar las filas de salida. Si la fila de origen contiene datos de clientes y de direcciones, puede crear un grupo de salida para los campos de cliente. Puede crear otro grupo de salida para los campos de dirección. O bien, puede actualizar el grupo de salida predeterminado y añadirle los campos de dirección. Los siguientes ejemplos muestran resultados de salida distintos según la configuración del grupo de salida.

Crear un grupo de salida

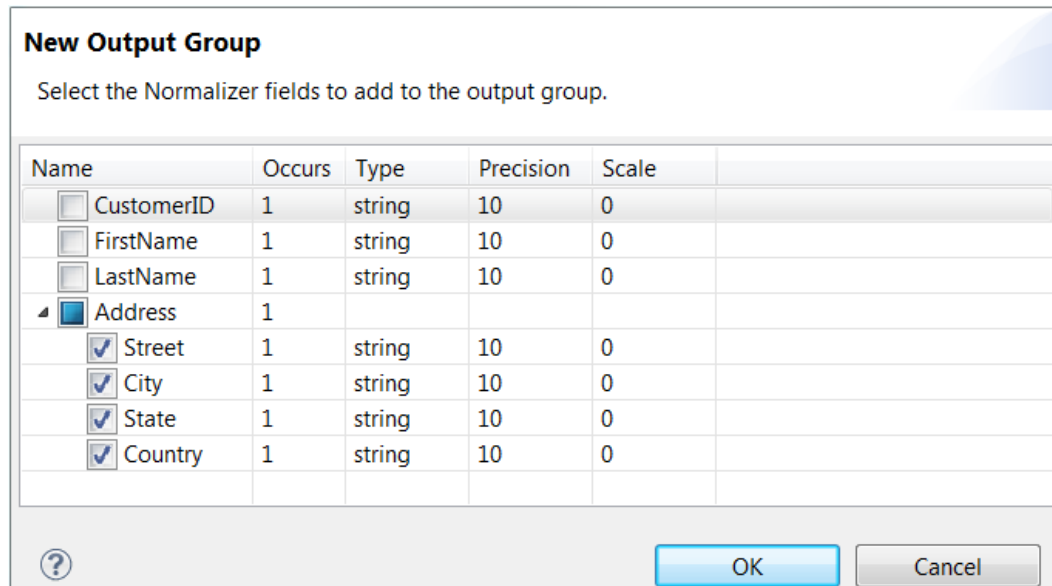
Cree un grupo de salida en la vista **Resumen** de la transformación de normalizador.

Cuando se abre la vista **Resumen** de la transformación de normalizador, la herramienta del desarrollador muestra los grupos predeterminados que crea desde la jerarquía de entrada.

Cuando se crea un nuevo grupo de salida, un cuadro de diálogo muestra una lista de los campos y los registros de la jerarquía de entrada. Seleccione qué campos o registros desea incluir en el grupo.

Ejemplo de grupo de salida

La siguiente imagen muestra el cuadro de diálogo **Nuevo grupo de salida**:



Cuando se selecciona el registro Address, la herramienta del desarrollador crea un grupo de puertos de salida que corresponden a los campos del registro Address. El grupo de Output1 contiene los puertos Street, City, State y Country (calle, ciudad, estado, país). Puede cambiar los puertos en el grupo de salida.

La siguiente imagen muestra el grupo Output y el grupo Output1 en la vista **Resumen**:

	Output				
1	CustomerID	string	10	0	CustomerID
2	FirstName	string	10	0	FirstName
3	LastName	string	10	0	LastName
	Output1				
1	Street	string	10	0	Address.Street
2	State	string	10	0	Address.State
3	Country	string	10	0	Address.Country
4	GCID_Address	bigint	19	0	Address

Puede configurar la transformación de normalizador para que devuelva las filas del grupo Output a una tabla de clientes.

La tabla de clientes recibe datos similares a las siguientes filas:

```
100, Robert, Bold
200, James, Cowan
```

Puede devolver las filas del grupo Output1 a una tabla Address. La tabla Address recibe Street, City, State, Country y GCID_Address (calle, ciudad, estado, país, dirección GCID).

La tabla Address recibe datos similares a las siguientes filas:

```
100 Summit Dr, Redwood City, CA, United States,1
41 Industrial Way, San Carlos, CA, United States,2
85 McNulty Way, Los Angeles, CA, United States,1
55 Factory Street, Los Vegas, NV, United States,2
```

El GCID identifica qué instancia de la dirección de cliente se encuentra en la fila de salida. En este ejemplo, la transformación de normalizador devuelve dos instancias del registro Address. Cada fila de salida contiene un valor de GCID de 1 o 2.

Actualizar un grupo de salida

Puede actualizar un grupo de salida de transformación de normalizador. Puede añadir o quitar los campos del grupo.

De manera predeterminada, la herramienta del desarrollador crea grupos de salida de nivel 1 cuando se define una jerarquía de entrada. La herramienta del desarrollador no incluye registros en los grupos. Puede actualizar los grupos de salida predeterminados y añadirles registros.

Para actualizar un grupo de salida, resalte el nombre del grupo y haga clic en **Nuevo > Actualizar grupo**. El cuadro de diálogo **Editar grupo de salida** muestra los campos en la jerarquía de entrada. Elija qué campos quiere incluir en el grupo.

Ejemplo de actualización de grupo de salida

En el ejemplo anterior, la herramienta del desarrollador creaba un grupo de salida predeterminado con los campos CustomerID, FirstName y LastName.

La siguiente imagen muestra el grupo de salida predeterminado:

Output					
1	CustomerID	string	10	0	CustomerID
2	FirstName	string	10	0	FirstName
3	LastName	string	10	0	LastName

Puede actualizar el grupo de salida predeterminado y añadirle el registro Address.

La siguiente imagen muestra el cuadro de diálogo **Editar grupo de salida**:

Edit Output Group

Select the Normalizer fields to include in the output group.

Name	Occurs	Type	Precision	Scale
<input checked="" type="checkbox"/> CustomerID	1	string	10	0
<input checked="" type="checkbox"/> FirstName	1	string	10	0
<input checked="" type="checkbox"/> LastName	1	string	10	0
<input checked="" type="checkbox"/> Address	2			
<input checked="" type="checkbox"/> Street	1	string	10	0
<input checked="" type="checkbox"/> State	1	string	10	0
<input checked="" type="checkbox"/> Country	1	string	10	0

En este ejemplo, CustomerID, FirstName y LastName son nodos de nivel 1. El registro Address también es un nodo de nivel 1. La transformación de normalizador puede devolver Address en la misma fila que los datos de clientes. Como Address es de ocurrencia múltiple, la herramienta del desarrollador añade el índice GCID_Address al grupo de salida.

La siguiente imagen muestra los puertos del grupo de salida:

Output					
1	CustomerID	string	10	0	CustomerID
2	FirstName	string	10	0	FirstName
3	LastName	string	10	0	LastName
4	Street	string	10	0	Address.Street
5	State	string	10	0	Address.State
6	Country	string	10	0	Address.Country
7	GCID_Address	bigint	19	0	Address

Cuando los campos de clientes de ocurrencia múltiple y los campos de direcciones están en el grupo de salida, la transformación de normalizador devuelve los mismos campos de clientes por cada instancia de los datos de direcciones.

El siguiente ejemplo muestra las filas que genera la transformación de normalizador a partir del grupo de salida:

```
100, Robert, Bold, 100 Summit Dr, Redwood City, CA, United States,1
100, Robert, Bold, 41 Industrial Way, San Carlos, CA, United States,2
200, James, Cowan, 85 McNulty Way, Los Angeles, CA, United States,1
200, James, Cowan, 55 Factory Street, Los Vegas, NV, United States,2
```

El puerto GCID contiene el número de la instancia de Address. El valor de GCID es 1 o 2.

Generación de claves para grupos de salida

Puede configurar una transformación de generador de secuencia para generar claves que vinculan cada fila de salida devuelta por la transformación de normalizador desde la misma fila de origen.

Puede añadir una transformación de generador de secuencia antes de la transformación de normalizador en una asignación. La transformación de generador de secuencia añade un número de secuencia a cada fila de origen. Cuando la transformación de normalizador devuelve varios grupos o filas de salida desde la misma fila de origen, cada fila de salida recibe el mismo número de secuencia. Puede utilizar el número como clave en una relación de clave principal-clave externa entre las tablas de destino.

Por ejemplo, la transformación de normalizador devuelve información de clientes en un grupo de salida y devuelve información de pedidos en otro grupo de salida. Puede utilizar el número de secuencia para vincular la información de clientes de una tabla con la información de pedidos de otra tabla.

Propiedades avanzadas de transformación de normalizador

Configure las propiedades de la transformación de normalizador en la ficha **Avanzadas**.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Generar grupos de salida de primer nivel

De manera predeterminada, la herramienta del desarrollador genera el grupo de salida de primer nivel de la transformación de normalizador. La herramienta del desarrollador genera grupos de salida adicionales al definir más de un campo de ocurrencia múltiple en la vista Normalizador.

La herramienta del desarrollador crea un grupo de salida que contiene todos los campos de ocurrencia individual de nivel 1 y el primer campo de ocurrencia múltiple de nivel 1. La herramienta del desarrollador crea un grupo de salida adicional si la transformación contiene más de un campo de ocurrencia múltiple.

La herramienta del desarrollador no crea un grupo de salida para los registros de nivel 1. Al definir un registro de nivel 1, deberá configurar la ubicación en la que desea ponerlos en la salida.

Para evitar que la herramienta del desarrollador cree grupos de salida de campos de ocurrencia múltiple, deshabilite la propiedad avanzada **Generar automáticamente grupos de salida de primer nivel**.

Creación de una transformación de normalizador

Puede crear una transformación de normalizador reutilizable o no reutilizable. Las transformaciones reutilizables pueden existir en varias asignaciones. Las transformaciones no reutilizables existen en una única asignación.

1. Para crear una transformación, utilice uno de los siguientes métodos:

Opción	Descripción
Reutilizable	Seleccione un proyecto o una carpeta en la vista Object Explorer . Haga clic en Archivo > Nuevo > Transformación . Seleccione la transformación de normalizador y haga clic en Siguiente .
No reutilizable	En una asignación o en un mapplet, arrastre una transformación de normalizador desde la paleta Transformación al editor. Puede arrastrar puertos de objetos de datos de origen o de transformaciones de la asignación para definir la transformación.

Aparece el asistente para la **Nueva transformación de normalizador**.

2. Especifique un nombre para la transformación.
3. Haga clic en **Siguiente**.

Aparece la página **Definición de normalizador**.

4. Para añadir un registro, haga clic en el botón **Nuevo** y, a continuación, seleccione **Registro**.

5. Para añadir un campo, haga clic en el botón **Nuevo** y, a continuación, seleccione **Campo**.
Para añadir un campo a un registro, deberá seleccionar el registro antes de añadir el campo.
6. Opcionalmente, haga doble clic en el valor de la columna **Ocurre** para cambiar la ocurrencia de un campo o un registro.
7. Haga clic en **Siguiente**.
Aparece la página **Puertos de normalizador**.
8. Para añadir un grupo de salida, haga clic en el botón **Nuevo** y después seleccione **Nuevo grupo de salida**.
9. Para editar un grupo de salida, seleccione el grupo de salida que desee editar. Haga clic en el botón **Nuevo** y, a continuación, seleccione **Editar grupo de salida**.
10. Haga clic en **Finalizar**.
La transformación aparece en el editor.

Creación de una transformación de normalizador desde un origen de nivel superior

Puede crear una transformación de normalizador vacía y arrastrar los puertos desde un objeto de datos de origen o desde una transformación a la transformación de normalizador para crear entradas y puertos.

1. Cree una asignación que incluya el origen o la transformación para pasar los datos de origen a la transformación de normalizador.
2. Para crear la transformación de normalizador, seleccione la transformación de normalizador en la paleta de transformación y arrastre la transformación al editor.
Aparece el cuadro de diálogo Normalizador.
3. Haga clic en **Finalizar** para crear una transformación vacía.
4. Seleccione puertos desde un origen o una transformación en la asignación y arrástrelos a la transformación de normalizador.
Los puertos de entrada y salida aparecen en la transformación de normalizador. La herramienta del desarrollador crea un grupo de entrada y un grupo de salida.
5. Abra la vista **Normalizador** para actualizar el grupo predeterminado y organice los campos en los registros, según sea necesario.
6. Para fusionar varios campos en un único campo de ocurrencia múltiple, seleccione los campos en la vista **Normalizador** y haga clic en la opción **Fusionar**.
Elija un nombre para el campo de ocurrencia múltiple.

Ejemplo de asignación de normalizador

Una organización minorista recibe los totales de ventas de las tiendas de la organización. La organización recibe una fila de datos que contiene información de las tiendas y cuatro cantidades de ventas. Cada importe de ventas representa las ventas totales de un trimestre del año.

El siguiente ejemplo muestra cómo definir una transformación de normalizador para que devuelva los datos de ventas a un destino de tiendas y un destino de ventas. El destino Tienda recibe una fila por cada tienda. Las destino Ventas recibe cuatro filas de cada tienda. Cada fila contiene un trimestre de datos de ventas.

Una transformación de generador de secuencia genera un ID único para cada tienda. La transformación de normalizador devuelve el StoreID con cada fila de salida.

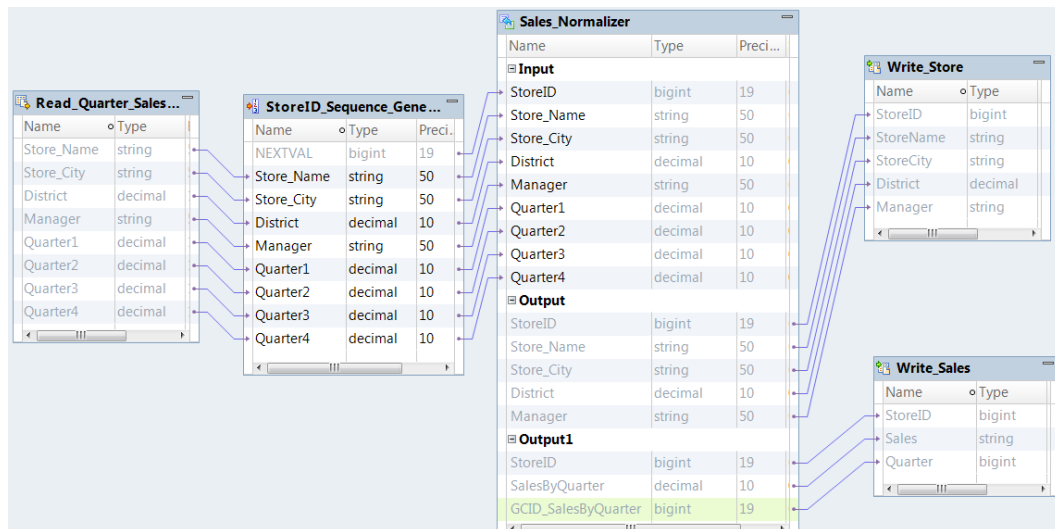
Cree una asignación con una transformación de lectura, una transformación de generador de secuencia, la transformación de normalizador y dos transformaciones de escritura.

Ejemplo de asignación de normalizador

Crear una asignación que contenga la transformación de normalizador para normalizar los datos de ventas trimestrales de ocurrencia múltiple desde un origen de archivo sin formato.

La transformación de normalizador genera una fila de salida independiente para cada trimestre de ventas y escribe las cantidades de ventas normalizadas en un destino de ventas. La transformación de normalizador escribe la información de las tiendas en un destino Tienda.

La siguiente figura muestra la asignación de la transformación de normalizador:



La asignación contiene los siguientes objetos:

Read_STORE

Un origen de datos que contiene campos de ocurrencia múltiple.

Transformación de generador de secuencia de StoreID

Una transformación de generador de secuencia que genera una clave de StoreID para vincular la tabla Tienda a la tabla Ventas.

Sales_Normalizer

Una transformación de normalizador que normaliza los datos de ventas de ocurrencia múltiple.

Write_Store

Un destino que recibe la información de las tiendas desde la transformación de normalizador.

Write_Sales

Un destino que recibe los números de ventas desde la transformación de normalizador.

Ejemplo de definición del normalizador

El origen es un archivo sin formato que contiene información de tiendas y datos de ventas trimestrales. Defina la estructura de los datos de origen en la vista **Normalizador**.

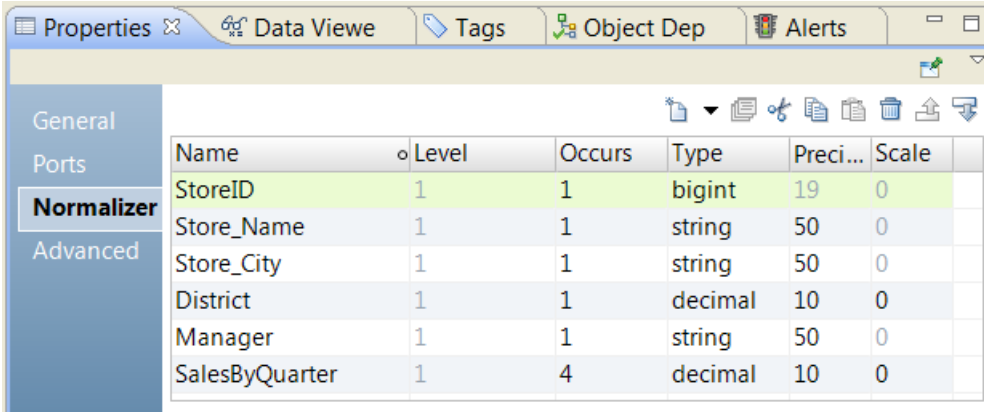
El archivo sin formato STORE contiene los siguientes datos de origen:

StoreID	Store_Name	Store_City	Distrito	Administrador	Quarter1	Quarter2	Quarter3	Quarter4
1	BigStore	Nueva York	Este	Robert	100	300	500	700
2	SmallStore	Phoenix	Oeste	Radhika	250	450	650	850

Añada el archivo sin formato a una asignación como una transformación de lectura y, a continuación, cree una transformación de normalizador vacía. Arrastre los puertos desde el objeto de datos Read_STORE a la transformación de normalizador para crear la definición del normalizador.

La vista **Normalizador** contiene una instancia de los campos Store_Name, Store_City, District y Manager. La vista **Normalizador** contiene cuatro instancias de un campo llamado QUARTER. Fusione los campos QUARTER para crear un campo SalesByQuarter que se produce cuatro veces.

La siguiente imagen muestra la definición de Normalizador con los campos Quarter fusionados:



The screenshot shows the 'Normalizer' task editor in SQL Server Data Tools. The 'Ports' tab is selected, showing a table of field definitions:

Name	Level	Occurs	Type	Preci...	Scale
StoreID	1	1	bigint	19	0
Store_Name	1	1	string	50	0
Store_City	1	1	string	50	0
District	1	1	decimal	10	0
Manager	1	1	string	50	0
SalesByQuarter	1	4	decimal	10	0

Ejemplo de grupos de entrada y salida del normalizador

Tras modificar la jerarquía de entrada, la transformación de normalizador tiene un grupo de entrada y un grupo de salida predeterminado. Deberá reorganizar los puertos de salida en dos grupos. Necesita un grupo que contenga la información de tiendas y un grupo que contenga la información de ventas.

El grupo de entrada contiene un puerto para cada campo del origen. El grupo de salida contiene puertos para los campos de tiendas y el puerto para el campo SalesByQuarter de ocurrencia múltiple. El grupo de salida también contiene un ID de columna generado, `GCID_SalesByQuarter`, que corresponde con el campo SalesByQuarter de ocurrencia múltiple.

Para devolver las ventas trimestrales a un destino diferente, cree un nuevo grupo en la vista **Resumen**. En el grupo Output1, añada los siguientes campos:

```
StoreID
SalesByQuarter
GCID_SalesByQuarter
```

Actualice el grupo de salida predeterminado. Quite los siguientes campos:

```
SalesByQuarter
GCID_SalesByQuarter
```

La siguiente imagen muestra el grupo de entrada y los grupos de salida en la vista **Resumen**:

	Name	Type	Precision	Scale	Location
Input					
1	StoreID	bigint	19	0	StoreID
2	Store_Name	string	50	0	Store_N...
3	Store_City	string	50	0	Store_City
4	District	decimal	10	0	District
5	Manager	string	50	0	Manager
6	Quarter1	decimal	10	0	SalesBy...
7	Quarter2	decimal	10	0	SalesBy...
8	Quarter3	decimal	10	0	SalesBy...
9	Quarter4	decimal	10	0	SalesBy...
Output					
1	StoreID	bigint	19	0	StoreID
2	Store_Name	string	50	0	Store_N...
3	Store_City	string	50	0	Store_City
4	District	decimal	10	0	District
5	Manager	string	50	0	Manager
Output1					
1	StoreID	bigint	19	0	StoreID
2	SalesByQuarter	decimal	10	0	SalesBy...
3	GCID_SalesByQuarter	bigint	19	0	SalesBy...

StoreID es la clave generada que vincula la información de tiendas con la información de ventas. Compruebe que los dos grupos de salida devuelven el StoreID.

Salida de asignación de ejemplo de normalizador

Añada la transformación de escritura a la asignación y conecte los puertos de salida de la transformación del normalizador a los objetos de datos.

Al ejecutar la asignación, la transformación de normalizador escribe las siguientes filas en el destino Tienda:

StoreID	Store_Name	Store_City	Distrito	Administrador
1	BigStore	Nueva York	Este	Robert
2	SmallStore	Phoenix	Oeste	Radhika

La transformación de normalizador escribe las siguientes filas en el destino Ventas:

StoreID	SalesByQuarter	GCID_SalesByQuarter
1	100	1

StoreID	SalesByQuarter	GCID_SalesByQuarter
1	300	2
1	500	3
1	700	4
2	250	1
2	450	2
2	650	3
2	850	4

Transformación de normalizador en un entorno no nativo

El procesamiento de transformaciones de normalizador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 33

Transformación de fusión

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de fusión, 560](#)
- [Cómo configurar una estrategia de fusión, 560](#)
- [Propiedades avanzadas de Transformación de fusión, 561](#)
- [Transformación de fusión en un entorno no nativo, 561](#)

Resumen de la transformación de fusión

La transformación de fusión es una transformación pasiva que lee los valores de datos de varias columnas de entrada y crea una columna de salida individual.

Utilice la transformación de fusión para crear datos en un formato preferido. Por ejemplo, puede combinar los campos Customer_Firstname y Customer_Surname para crear un campo llamado Customer_FullName.

Dentro de una transformación de fusión puede crear varias estrategias de fusión. La transformación de fusión proporciona un asistente que puede usarse para crear estrategias.

Cómo configurar una estrategia de fusión

Para configurar una estrategia de fusión, edite los ajustes en la vista **Estrategias** de una transformación de fusión.

1. Seleccione la vista **Estrategias**.
2. Haga clic en **Nueva**.
Se abre el asistente **Nueva estrategia**.
3. Haga clic en el campo **Entradas** para seleccionar los puertos de entrada para la estrategia.
4. Para definir el carácter de fusión que colocar entre los elementos fusionados, haga clic en **Elegir**. Si no se elige un carácter de fusión, la transformación de fusión utiliza de forma predeterminada el carácter de espacio.
5. Si lo desea, seleccione **Incluir cadenas vacías en salida fusionada** para incluir las cadenas de entrada vacías en la salida.
6. Haga clic en **Finalizar**.

Propiedades avanzadas de Transformación de fusión

Configure propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de fusión.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de fusión en un entorno no nativo

El procesamiento de transformaciones de fusión en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 34

Transformación de analizador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de analizador, 562](#)
- [Modos de transformación del analizador, 563](#)
- [Cuando utilizar una transformación de analizador, 563](#)
- [Uso de los datos de referencia en la transformación de analizador, 564](#)
- [Operaciones de análisis de tokens, 566](#)
- [Puertos del análisis de tokens, 567](#)
- [Propiedades del análisis de tokens, 568](#)
- [Modo de análisis basado en patrones, 570](#)
- [Configurar una estrategia de análisis de tokens, 571](#)
- [Configurar una estrategia de análisis de patrones, 572](#)
- [Propiedades avanzadas de Transformación de analizador, 573](#)
- [Transformación de analizador en un entorno no nativo, 573](#)

Resumen de la transformación de analizador

La transformación de analizador es una transformación pasiva que analiza los valores de los datos de entrada en nuevos puertos. La transformación escribe los valores en nuevos puertos según los tipos de información que contengan los valores y la posición de los valores en la cadena de entrada.

Utilice una transformación de analizador cuando desee cambiar la estructura de un conjunto de datos. La transformación de analizador puede añadir columnas a un conjunto de datos y escribir los valores de datos en nuevas columnas. Utilice una transformación de analizador cuando una columna de datos contenga varios valores en una sola columna y desee escribir los valores de datos en columnas individuales según el tipo de información que contengan.

La transformación de analizador analiza valores de datos en los puertos de salida que haya definido. Si la transformación puede identificar un valor de datos de entrada pero un puerto de salida definido no está disponible, la transformación escribe en un puerto de desbordamiento. Si la transformación no puede identificar un valor de datos de entrada, escribe en un puerto de datos sin analizar.

Modos de transformación del analizador

Cuando cree una transformación de analizador, seleccione el modo de análisis de tokens o el modo de análisis basado en patrones.

Seleccione uno de los siguientes modos:

- Modo de análisis de tokens. Utilice este modo para analizar los valores de entrada que coinciden con valores de los objetos de datos de referencia como conjuntos de tokens, expresiones regulares, modelos probabilísticos y tablas de referencia. Se pueden usar varias estrategias de análisis de tokens en una transformación.
- Modo de análisis basado en patrones. Utilice este modo para analizar los valores de entrada que coinciden con los valores de los conjuntos de patrones.

Cuando utilizar una transformación de analizador

Utilice la transformación de analizador cuando los campos de datos de una columna contengan más de un tipo de información y desee mover los valores del campo a nuevas columnas. La transformación de analizador le permite crear una nueva columna para cada tipo de información en un conjunto de datos.

Los ejemplos siguientes describen ciertos tipos de cambios estructurales que puede realizar con una transformación de analizador.

Crear nuevas columnas para los datos de contacto

Puede crear una estructura de datos que analice los datos de nombres desde una sola columna a varias columnas. Por ejemplo, puede crear columnas para tratamientos, nombres de pila, segundos nombres y apellidos.

Configure la transformación con un modelo probabilístico que represente las estructuras de los nombres de personas del puerto de entrada. Puede utilizar un ejemplo de los datos de los puertos de entrada para definir el modelo.

Cree una estrategia de análisis de tokens que se aplique al modelo probabilístico del puerto de entrada y escriba valores de nombre en nuevas columnas. La transformación escribe los valores de nombre en las nuevas columnas según la posición de cada valor en la cadena de entrada y el tipo de nombre que el valor representa.

Nota: también puede utilizar una estrategia de análisis basado en patrones para analizar los datos de contacto. Cuando configure una estrategia de análisis basado en patrones, defina los patrones que representan las estructuras de los nombres del puerto de entrada.

Crear columnas de dirección

Puede crear una estructura de datos que analice una sola columna de datos de dirección en varias columnas que describen una dirección de entrega.

Configure la transformación con las tablas de referencia que contienen elementos de dirección reconocibles, como códigos postales, nombres de estado y nombres de ciudades. Cree una estrategia de análisis de tokens que escriba cada elemento de dirección en un nuevo puerto.

No puede utilizar una tabla de referencia para analizar datos de dirección de calles desde una cadena de entrada porque el nombre de la calle y los datos del número son demasiado generales para que una tabla de referencia pueda capturarlos. Sin embargo, puede utilizar el puerto de desbordamiento para capturar estos datos. Cuando haya analizado todos los datos de ciudades, estados y códigos postales de una dirección, el resto de datos contienen la información de las calles.

Por ejemplo, utilice una estrategia de análisis de tokens para dividir la siguiente dirección en elementos de dirección:

```
123 MAIN ST NW STE 12 ANYTOWN NY 12345
```

La estrategia de análisis puede escribir los elementos de dirección en las columnas siguientes:

Nombre de columna	Datos
Desbordamiento	123 MAIN ST NW STE 12
Ciudad	ANYTOWN
Estado	NY
ZIP	12345

Crear columnas de datos de producto

Puede crear una estructura de datos que analice una sola columna de datos de productos en varias columnas que describen los detalles del producto e inventario.

Configure la transformación con conjuntos de tokens que contengan elementos de inventario, tales como dimensión, color y peso. Cree una estrategia de análisis de tokens que escriba cada elemento de inventario en un nuevo puerto.

Por ejemplo, utilice una estrategia de análisis de tokens para dividir la siguiente descripción de pintura en elementos de inventario independientes:

```
500ML Red Matt Exterior
```

La estrategia de análisis puede escribir los valores en las siguientes columnas:

Nombre de columna	Datos
Tamaño	500 ML
Color	Rojo
Estilo	Mate
Exterior	Y

Uso de los datos de referencia en la transformación de analizador

Informatica Developer se instala con varios objetos de datos de referencia que puede utilizar con la transformación de analizador. También puede crear objetos de datos de referencia en Developer Tool.

Cuando añade un objeto de datos de referencia a una transformación de analizador, la transformación escribe las cadenas que coinciden con un valor del objeto con las nuevas columnas que haya especificado.

La siguiente tabla describe los tipos de datos de referencia que puede utilizar:

Tipo de datos de referencia	Descripción
Conjuntos de patrones	Identifica los valores de datos según la posición relativa de cada valor en la cadena.
Modelos probabilísticos	Añade capacidades de coincidencia parcial para las operaciones de análisis de tokens. La transformación puede utilizar un modelo probabilístico para inferir el tipo de información de una cadena. Para habilitar las capacidades de coincidencia parcial, puede compilar el modelo probabilístico en la herramienta del desarrollador.
Tablas de referencia	Busca las cadenas que coinciden con las entradas en una tabla de base de datos.
Expresiones regulares	Identifica las cadenas que coinciden con las condiciones que ha definido. Puede utilizar una expresión regular para buscar una cadena dentro de una cadena más grande.
Conjuntos de tokens	Identifica las cadenas según los tipos de información que contienen. Informatica se instala con conjuntos de tokens y distintos tipos de definiciones de tokens, tales como definiciones de palabra, número de teléfono, código postal y código de producto.

Conjuntos de patrones

Un conjunto de patrones contiene expresiones que identifican patrones de datos en la salida de una operación de etiquetado con tokens. Puede emplear conjuntos de patrones para analizar el puerto de salida de datos agrupados y para escribir cadenas coincidentes en uno o más puertos de salida. Utilice conjuntos de patrones en transformaciones de analizador que utilizan el modo de análisis de patrones.

Por ejemplo, puede configurar una transformación de analizador para que utilice conjuntos de patrones que identifiquen nombres e iniciales. Esta transformación utiliza los conjuntos de patrones para analizar la salida de una transformación de etiquetador en el modo de etiquetado de tokens. Puede configurar la transformación de analizador para que escriba nombres e iniciales en la salida, en puertos diferentes.

Modelos probabilísticos

Un modelo probabilístico identifica tokens por los tipos de información que contengan y por sus posiciones en una cadena de entrada.

Un modelo probabilístico contiene valores de datos de referencia y valores de etiqueta. Los valores de datos de referencia representan los datos de un puerto de entrada que conecta a la transformación. Los valores de etiqueta describen los tipos de información que contienen los valores de datos de referencia. Puede asignar una etiqueta a cada valor de datos de referencia del modelo.

Para vincular los valores de datos de referencia a las etiquetas de un modelo probabilístico, debe compilar el modelo. El proceso de compilación genera una serie de asociaciones lógicas entre los valores de datos y las etiquetas. Cuando se ejecuta una asignación que lee el modelo, el Servicio de integración de datos aplica la lógica del modelo a los datos de entrada de la transformación. El Servicio de integración de datos devuelve la etiqueta que mejor describe la información de cada valor de datos.

Cree un modelo probabilístico en Developer tool. El repositorio de modelos almacena el objeto del modelo probabilístico. Developer tool escribe los valores de datos, las etiquetas y los datos de compilación en un archivo de la estructura del directorio de Informatica.

Nota: Si añade un modelo probabilístico a una operación de análisis de tokens y, a continuación, edita la configuración de la etiqueta en el modelo probabilístico, invalidará la operación. Cuando actualiza la

configuración de la etiqueta en un modelo probabilístico, no recrear ninguna operación de análisis que utilice el modelo.

Tablas de referencia

Una tabla de referencia es una tabla de base de datos que contiene al menos dos columnas. Una columna contiene la versión estándar o requerida de un valor de datos y otras columnas contienen versiones alternativas del valor. Cuando añade una tabla de referencia a una transformación, la transformación busca los datos de puertos de entrada para los valores que también aparecen en la tabla. Puede crear tablas con cualquier dato que sea útil para el proyecto de datos en el que esté trabajando.

Expresiones regulares

En el contexto de las operaciones de análisis, una expresión regular es una expresión que se puede usar para identificar una o más cadenas en datos de entrada. La transformación de analizador escribe cadenas identificadas en uno o más puertos de salida. Puede utilizar expresiones regulares en transformaciones de analizador que utilizan el modo de análisis de tokens.

Las transformaciones de analizador utilizan expresiones regulares para hacer coincidir patrones en datos de entrada y analizar todas las cadenas coincidentes con una o más salidas. Por ejemplo, puede utilizar una expresión regular para identificar todas las direcciones de correo electrónico en datos de entrada y analizar cada componente de la dirección de correo electrónico en una salida diferente.

Conjuntos de tokens

Un conjunto de tokens contiene expresiones que identifican tokens específicos. Puede utilizar conjuntos de tokens en transformaciones de analizador que utilizan el modo de análisis de tokens.

Emplee los conjuntos de tokens para identificar tokens como parte de las operaciones de análisis. Por ejemplo, puede utilizar un conjunto de tokens para analizar todas las direcciones de correo electrónico que utilizan un formato "AccountName@DomainName".

Operaciones de análisis de tokens

En el modo de análisis de tokens, la transformación de analizador analiza cadenas que coinciden con los datos de los conjuntos de tokens, expresiones regulares, modelos probabilísticos o entradas de la tabla de referencia.

Para realizar el análisis de tokens, añada estrategias en la vista **Estrategias** de la transformación. Se puede añadir una o más operaciones a cada estrategia. La transformación proporciona un asistente que puede usarse para crear estrategias.

Se pueden añadir los siguientes tipos de operación a una estrategia de análisis de tokens:

Analizar usando conjunto de tokens

Utilice conjuntos de tokens predefinidos o definidos por el usuario para analizar datos de entrada. Las operaciones con conjuntos de tokens pueden utilizar expresiones regulares personalizadas que escriben datos en una o más salidas.

También puede utilizar modelos probabilísticos para identificar y analizar los valores de datos de entrada.

Analizar usando tabla de referencia

Utilice las tablas de referencia para analizar datos de entrada.

La transformación lleva a cabo las operaciones en el orden que aparecen en la estrategia.

Puertos del análisis de tokens

Configure los puertos del análisis de tokens con ajustes apropiados para sus datos.

Una transformación de analizador en el modo de análisis de tokens tiene los siguientes tipos de puerto:

Entrada

Contiene datos que se pasan a la transformación de analizador. La transformación fusiona todos los puertos de entrada en una cadena de datos combinados mediante el **Carácter de unión de entrada** especificado en la ficha **Estrategias**. Si no se especifica un carácter de unión de entrada, la transformación utiliza de forma predeterminada un carácter de espacio.

Puertos de salida analizada

Puerto(s) definido(s) por el usuario que contienen cadenas analizadas correctamente. En los casos que varias estrategias de análisis utilicen la misma salida, la transformación fusiona la salida en una cadena de datos combinados mediante el **Carácter de unión de salida** especificado en la ficha **Estrategias**. Si no se especifica un carácter de unión de salida, la transformación utiliza de forma predeterminada un carácter de espacio.

Desbordamiento

Contiene cadenas analizadas correctamente que no caben en el número de salidas definido en la transformación. Por ejemplo, si la salida solamente tiene dos salidas "WORD", la cadena "John James Smith" da como resultado una salida desbordada de "Smith." La transformación de analizador crea un puerto de desbordamiento para cada estrategia que se añade.

Cuando selecciona la opción de desbordamiento detallado, la transformación crea un puerto de desbordamiento para cada etiqueta del modelo.

Sin analizar

Contiene cadenas que la transformación no puede analizar correctamente. La transformación de analizador crea un puerto no analizado para cada estrategia que se añade.

Puertos de salida en la coincidencia probabilística

Cuando configura una estrategia de análisis para utilizar técnicas de coincidencias probabilísticas, la transformación de analizador añade un puerto para almacenar las puntuaciones de coincidencia en cada puerto de salida.

La siguiente tabla describe los tipos de puerto:

Tipo de puerto	Puerto creado en la coincidencia probabilística
Puerto de salida analizado	salida [nombre de etiqueta] salida de puntuación [nombre de etiqueta]
Puerto de datos de desbordamiento	salida [desbordamiento de datos] salida de puntuación [desbordamiento de datos]
Puerto de datos sin analizar	salida [datos sin analizar] salida de puntuación [datos sin analizar]

Propiedades del análisis de tokens

Configure las propiedades para las operaciones de análisis de tokens la vista **Estrategias** en la transformación de analizador.

Propiedades generales

Las propiedades generales se aplican a todas las operaciones de análisis de tokens que haya definido en la estrategia. Utilice las propiedades generales para dar un nombre a la estrategia, especificar los puertos de entrada y salida y especificar si la estrategia permite técnicas de coincidencia probabilística.

En la tabla siguiente, se describen las propiedades generales.

Propiedad	Descripción
Nombre	Proporciona un nombre para la estrategia.
Entradas	Identifica los puertos de entrada que las operaciones de estrategia pueden leer.
Salidas	Identifica los puertos de salida en los que las operaciones de estrategia pueden escribir.
Descripción	Describe la estrategia. La propiedad es opcional.
Utilizar técnicas de coincidencias probabilísticas	Especifica que la estrategia puede utilizar un modelo probabilístico para identificar tokens.
Carácter de unión de entrada	Especifica el carácter utilizado para unir puertos de datos de entrada. La transformación fusiona todos los puertos de entrada en una cadena de datos combinados y analiza la cadena completa.
Carácter de combinación de salida	Especifica el carácter utilizado para combinar valores de datos de salida en el caso de que varias operaciones utilicen la misma salida.
Inversión habilitada	Configura la estrategia para analizar datos de derecha a izquierda. Esta propiedad está deshabilitada para la coincidencia probabilística.

Propiedad	Descripción
Permitir desbordamiento inverso	Configura la estrategia para analizar datos de izquierda a derecha. Esta propiedad está deshabilitada para la coincidencia probabilística.
Permitir desbordamiento detallado	Crea un campo de desbordamiento único para cada operación de análisis.
Delimitadores	Especifica el delimitador que separa los datos de entrada en tokens separados. El valor predeterminado es espacio.

Propiedades del modelo probabilístico

Puede seleccionar un modelo probabilístico en lugar de un conjunto de tokens al configurar una estrategia de análisis de tokens. Seleccione la operación **Analizar usando conjunto de tokens** y seleccione la opción para utilizar técnicas de coincidencias probabilísticas.

La tabla siguiente describe las propiedades de modelos probabilísticos:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Filtrar texto	Utiliza los caracteres o comodines que se especifican para filtrar la lista de conjuntos de tokens, los modelos probabilísticos o las expresiones regulares.
Modelo probabilístico	Identifica el modelo probabilístico que haya seleccionado.

Propiedades de la tabla de referencia

Las propiedades de la tabla de referencia se aplican al configurar una operación de etiquetado para usar una tabla de referencia.

La tabla siguiente describe las propiedades de la tabla de referencia:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Tabla de referencia	Especifica la tabla de referencia que utiliza la operación para analizar los valores de entrada.
Distinguir mayúsculas de minúsculas	Determina si el formato de mayúsculas y minúsculas de las cadenas de entrada debe coincidir con el de las entradas de la tabla de referencia.
Reemplazar coincidencias por valores válidos	Reemplaza los datos analizados con los datos de la columna Válida en la tabla de referencia.
Salidas	Especifica los puertos de salida para los datos analizados.

Propiedades de conjuntos de tokens

Las propiedades del conjunto de tokens se aplican al configurar una operación de análisis para utilizar conjuntos de tokens.

Seleccione la operación **Analizar usando conjunto de tokens** para analizar la entrada con conjuntos de tokens. Borre la opción para utilizar técnicas de coincidencias probabilísticas.

Las tablas siguientes describen las propiedades del conjunto de tokens:

Propiedad	Descripción
Nombre	Proporciona un nombre para la operación.
Conjuntos de tokens (solo una salida)	Especifica el conjunto de tokens que utiliza la operación para analizar datos. La operación escribe los datos en un solo puerto.
Expresión regular (una o varias salidas)	Especifica la expresión regular que utiliza la operación para analizar datos. La operación escribe los datos en varios puertos si encuentra varias cadenas en el campo de entrada.
Salidas	Identifica los puertos de salida en los que escriben las operaciones.

Puede añadir, editar, importar o eliminar un conjunto de tokens o una expresión regular. También puede filtrar la lista de los conjuntos de tokens.

La siguiente tabla describe las propiedades que se utilizan para realizar las tareas:

Propiedad	Descripción
Filtrar texto	Filtra la lista de conjuntos de tokens o de expresiones regulares. Utilice caracteres de texto y caracteres comodín como filtro.
Añadir	Utilícelo para definir un conjunto de tokens o una expresión regular personalizados.
Editar	Edita el contenido de un conjunto de tokens personalizado.
Importar	Importa una copia no reutilizable de un conjunto de tokens o una expresión regular desde una carpeta del repositorio de modelos. Si actualiza el objeto de origen del conjunto de tokens o la expresión regular, el servicio de integración de datos no actualizará la copia no reutilizable.
Eliminar	Elimina un conjunto de tokens personalizados o una expresión regular.

Modo de análisis basado en patrones

En el modo de análisis basado en patrones, la transformación de analizador analiza patrones formados por varias cadenas.

Puede utilizar los siguientes métodos para definir los patrones en el modo de análisis basado en patrones:

- Analice datos de entrada mediante patrones definidos en tablas de referencia. Puede utilizar una tabla de referencia de patrones desde la salida creada a partir de un perfil de una transformación de etiquetador que utiliza el modo de etiquetado de tokens.

- Analice datos de entrada con los patrones que defina.
- Analizar datos de entrada utilizando patrones que haya importado desde un conjunto de patrones reutilizables en el repositorio de modelos. Los cambios en el conjunto de patrones reutilizables no actualizan los datos que añada a la transformación de analizador.

Puede utilizar los caracteres comodín "+" y "*" para definir un patrón. Utilice caracteres "*" para hacer coincidir cualquier cadena y caracteres "+" para hacer coincidir una o más instancias de la cadena precedente. Por ejemplo, utilice "WORD+" para buscar varias instancias consecutivas de un token de palabra y utilice "WORD*" para buscar un token de palabra seguido de uno o más tokens de cualquier tipo.

Puede utilizar varias instancias de estos métodos dentro de la transformación de analizador. La transformación utiliza las instancias en el orden que aparecen en la lista de la vista **Configuración**.

Nota: En el modo de análisis basado en patrones, la transformación de analizador requiere la salida de una transformación de etiquetador que utiliza el modo de etiquetado de tokens. Cree y configure la transformación de etiquetador antes de crear una transformación de analizador que utiliza el modo de análisis basado en patrones.

Puertos de análisis basado en patrones

Configure los puertos de análisis basado en patrones con los ajustes apropiados a los datos.

Una transformación de analizador que utiliza el modo de análisis basado en patrones tiene los siguientes tipos de puerto:

Label_Data

Conecte este puerto al puerto `Labeled_Output` de una transformación de etiquetador que utilice el modo de etiquetado de tokens.

Tokenized_Data

Conecte este puerto al puerto `Tokenized_Data` de una transformación de etiquetador que utilice el modo de etiquetado de tokens.

Parse_Status

Si se encuentra una coincidencia para el patrón de entrada, el puerto emite el valor `Con coincidencia`. Si no se encuentra ninguna coincidencia, emite `Sin coincidencia`.

Desbordamiento

Cadenas analizadas correctamente que no encajan en el número de salidas definido en la transformación. Por ejemplo, si solamente se han definido dos salidas "WORD", la cadena "John James Smith" da como resultado un desbordamiento de "Smith" de forma predeterminada.

Analizada

Cadenas analizadas correctamente en puertos definidos por el usuario.

Configurar una estrategia de análisis de tokens

Para configurar una estrategia de análisis de tokens, abra una transformación de analizador en el modo de análisis de tokens y seleccione la vista **Estrategias**.

1. Seleccione la vista **Estrategias**.
2. Haga clic en **Nueva**.

- Se abre el asistente **Nueva estrategia** .
3. Haga clic en el campo **Entradas** para seleccionar los puertos para la estrategia.
 4. Edite las propiedades de la estrategia y haga clic en **Siguiente**.
 5. Elija una operación y haga clic en **Siguiente**.
 6. Configure las propiedades de la operación y seleccione puertos de salida para los datos analizados correctamente.
 7. Si lo desea, haga clic en **Siguiente** para añadir más operaciones a la estrategia.
 8. Después de añadir todas las operaciones a la estrategia, haga clic en **Finalizar**.
 9. Si lo desea, añada más estrategias a la transformación.
 10. Si lo desea, cambie el orden en que la transformación procesa las estrategias y las operaciones. Seleccione una estrategia u operación y haga clic en **Mover hacia arriba** o **Mover hacia abajo**.

Configurar una estrategia de análisis de patrones

Para configurar una estrategia de análisis de patrones, abra una transformación de analizador en modo de análisis de patrones y seleccione la vista **Patrones**.

Antes de configurar la transformación para analizar patrones, compruebe que la vista **Patrones** muestra los nombres de los puertos de salida que esperaba. La transformación de analizador analiza tokens en los puertos de salida que seleccione. Cree puertos de salida adicionales si es necesario.

1. Seleccione la vista **Patrones**.
2. Añada uno o más patrones a la estrategia. Puede añadir patrones de las siguientes formas:
 - Introducir valores de datos para crear un patrón. Haga clic en **Nuevo** y seleccione **Nuevo patrón**.
Si selecciona **Nuevo patrón**, haga clic en **Especificar patrones aquí** y especifique uno o varios tipos de token. Los tokens que ha introducido debe coincidir con la estructura de tokens de un campo de datos de entrada. Para añadir patrones tiene describir las estructuras de tokens del puerto de entrada.
 - Importar los valores de datos desde una tabla de referencia. Haga clic en **Nuevo** y seleccione **Nueva tabla de referencia**.
Si selecciona **Nueva tabla de referencia**, examine el repositorio de modelos y seleccione una tabla de referencia que contenga una lista de estructuras de token. La tabla de referencia debe tener dos columnas. La segunda columna de la tabla de referencia debe contener valores numéricos.
 - Importar valores de datos desde un conjunto de patrones. Haga clic en **Importar** y seleccione un conjunto de patrones reutilizable en el repositorio de modelos.
Si selecciona **Importar**, busque los conjuntos de contenido en el repositorio de modelos y seleccione un conjunto de patrones reutilizable.

Nota: puede usar el campo **Filtrar texto** para filtrar la lista de tablas de referencia y los conjuntos de patrones.

Puede mezclar conjuntos de patrones y tablas de referencia en los patrones de columna.
3. Asigne cada token de la columna de patrones a un puerto de salida.
 - Para asignar un token en un puerto de salida, haga doble clic en la columna de puerto y seleccione el nombre del token desde el menú.

- Para analizar varios tokens en una única salida, haga doble clic en la columna de puerto y seleccione **Personalizado**. Asigne tokens al puerto y seleccione el delimitador que desee utilizar.

Asigne los tokens de cada fila del patrón a uno o más puertos de salida.

4. Guarde la transformación.

Propiedades avanzadas de Transformación de analizador

Configure las propiedades para ayudarle a determinar cómo el servicio de integración de datos procesa los datos para la transformación de analizador.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de analizador en un entorno no nativo

El procesamiento de transformaciones de analizador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 35

Transformación de Python

La transformación de Python proporciona una interfaz para definir la función de transformación usando código Python.

Python es un lenguaje que utiliza una sintaxis sencilla, escritura dinámica y enlaces dinámicos, lo que hace que Python sea una opción ideal para aumentar la productividad o participar en el desarrollo rápido de aplicaciones. Cuando usa su código Python en una asignación de Data Engineering, dicho código se inserta en el código de Scala generado que el motor de Spark o Databricks Spark ejecuta para procesar conjuntos de datos grandes, diversos y en constante cambio.

También puede utilizar la transformación de Python para el aprendizaje automático. En la transformación, puede especificar un archivo de recursos que contiene un modelo previamente entrenado y cargar dicho modelo en el código Python. Por ejemplo, puede cargar un modelo previamente entrenado para clasificar datos de entrada o para crear predicciones.

Para poder utilizar la transformación de Python, configure las propiedades avanzadas de Spark correspondientes en las propiedades de conexión de Hadoop o Databricks. Luego, asegúrese de que los nodos de trabajo del clúster contienen una instalación de Python.

Para obtener más información sobre la instalación de Python, consulte la *Guía de Data Engineering Integration*.

Solo puede ejecutar la transformación de Python en el motor de Spark o de Databricks Spark. La transformación de Python no se puede ejecutar en el entorno nativo.

Para obtener más información acerca de la transformación de Python, consulte la *Guía del usuario de Data Engineering Integration*.

CAPÍTULO 36

Transformación de rango

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de rango, 575](#)
- [Transformaciones de rango en asignaciones dinámicas, 576](#)
- [Puertos de la transformación de rango, 577](#)
- [Puerto de rango, 578](#)
- [Definir agrupaciones por puertos, 578](#)
- [Memorias caché de rango, 580](#)
- [Propiedades avanzadas de la transformación de rango, 580](#)
- [Transformación de rango en un entorno no nativo, 581](#)

Resumen de la transformación de rango

La transformación de rango es una transformación activa que limita registros a un rango superior o inferior. Utilice una transformación de rango para devolver el mayor o el menor valor numérico de un puerto o grupo. O bien use una transformación de rango para devolver las cadenas de la parte superior o inferior de un orden de clasificación de una asignación.

Durante la ejecución de una asignación, el Servicio de integración de datos guarda en la memoria caché datos de entrada hasta que pueda realizar los cálculos de rango.

La transformación de rango difiere de las funciones de transformación MAX y MIN. La transformación de rango devuelve un grupo de valores de la parte superior o inferior, no solo un único valor. Por ejemplo, utilice una transformación de rango para seleccionar los diez primeros vendedores de un territorio en concreto. O bien, con el fin de generar un informe financiero, puede utilizar una transformación de rango para identificar los tres departamentos que menos gastan en salarios y gastos generales. Mientras que el lenguaje SQL ofrece muchas funciones diseñadas para manejar grupos de datos, no es posible realizar la identificación de estratos superiores o inferiores de un conjunto de filas haciendo uso de las funciones SQL estándar.

Conecte a la transformación todos los puertos que representen el mismo conjunto de filas. Las filas que queden dentro del rango, en función de algunas medidas que establezca al configurar la transformación, pasarán a través de la transformación de rango.

Como transformación activa, la transformación de rango puede modificar el número de filas que pasan a través de ella. Puede pasar 100 filas a la transformación de rango, pero solo puede hacer una selección para crear el rango de las 10 filas principales. Las diez filas superiores pasan de la transformación de rango a otra transformación.

No puede enlazar puertos de una transformación con la transformación de rango. A su vez, también puede crear variables locales y escribir expresiones de no agregado.

Valores de las cadenas de establecimiento de rango

Puede configurar la transformación de rango para devolver los valores superiores o inferiores de un puerto de cadena. El servicio de integración de datos ordena las cadenas en función del orden de clasificación seleccionado para la asignación utilizada.

Cuando configure la aplicación que contiene la asignación, seleccione el orden de clasificación que el servicio de integración de datos utiliza para ejecutar la asignación. Puede seleccionar un lenguaje binario o un idioma específico, como el francés o el alemán. Si selecciona binario, el servicio de integración de datos calcula el valor binario de cada cadena y ordena las cadenas mediante los valores binarios. Si selecciona un idioma, el servicio de integración de datos ordena las cadenas por orden alfabético mediante el orden de clasificación del idioma.

Propiedades de la transformación de rango

Cuando crea una transformación de rango, puede configurar las siguientes propiedades:

- Introduzca un directorio de la memoria caché.
- Seleccione el rango superior o inferior.
- Seleccione el puerto de entrada/salida que contiene los valores que se usan para determinar el rango. Puede seleccionar solo un puerto para definir un rango.
- Seleccione el número de filas en las que desee crear un rango.
- Defina los grupos para crear un rango, como por ejemplo los diez productos menos caros de cada fabricante.

Transformaciones de rango en asignaciones dinámicas

Puede utilizar una transformación de rango en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados.

Si hace referencia a un puerto generado en la transformación de rango y dicho puerto no existe en tiempo de ejecución, la asignación fallará.

Si especifica un puerto dinámico como puerto de rango, ese puerto dinámico no podrá tener más de un puerto generado.

Si especifica un puerto dinámico como puerto de agrupación, el servicio de integración de datos considerará todos los puertos generados como puertos de agrupación. La asignación no es válida si especifica un puerto generado como puerto de agrupación y especifica el puerto dinámico primario como puerto de rango o como puerto de agrupación.

Puede parametrizar el puerto de rango y la agrupación por puertos. Utilice un parámetro de tipo de puerto para el puerto de rango. Utilice un parámetro de tipo de lista de puertos para la agrupación por puertos.

Puertos de la transformación de rango

La transformación de rango incluye puertos de entrada, de entrada/salida o de salida que están conectados a otra transformación de la asignación. La transformación también incluye puertos de transferencia y de variable.

Una transformación de rango cuenta con los siguientes tipos de puertos:

Entrada

Recibe los datos de las transformaciones de nivel superior. Usted puede diseñar puertos de entrada como puertos de entrada/salida. La transformación debe tener al menos un puerto de entrada.

Puerto dinámico

Puerto que puede recibir varias columnas para crear un número dinámico de puertos generados. Un puerto generado es un puerto dentro de un puerto dinámico que representa una única columna. Puede crear puertos dinámicos de entrada, de salida y de variable.

Salida

Pasa datos a las transformaciones de nivel inferior. Puede diseñar puertos de salida como puertos de entrada/salida. La transformación debe tener al menos un puerto de salida.

Transferencia

Pasa los datos inalterados.

Variable

Se utiliza para variables locales. Puede utilizar un puerto de variable para guardar valores o cálculos para utilizar en una expresión. Los puertos de variable no pueden ser ni puertos de entrada ni de salida. Pasan datos dentro de la transformación.

Índice de rango

Developer Tool crea un puerto RANKINDEX para cada transformación de rango. El servicio de integración de datos utiliza el puerto de índice de rango para guardar la posición en el rango de cada fila del grupo.

Por ejemplo, puede crear una transformación de rango para identificar los 50 empleados que más cobran de la empresa. Identifique la columna SALARY como el puerto de entrada/salida que se utiliza para medir los rangos y configure la transformación para filtrar todas las filas y quedarnos solo con las 50 primeras.

Una vez que la transformación de rango identifique todas las filas que formen parte de un rango superior o inferior, asigna valores del índice de rango. En el caso de los 50 primeros empleados en función de su salario, el que más cobra recibe el índice de rango 1. El siguiente mejor pagado recibe el índice de rango 2, y así sucesivamente. Al medir un rango inferior, como por ejemplo los diez productos con el precio más bajo del inventario, la transformación de rango asigna un índice de rango en orden ascendente (del más bajo al más alto). Así, el producto más barato recibiría el índice de rango 1.

Si dos valores de rango coinciden, ambos reciben el mismo valor en el índice de rango y la transformación omite el siguiente valor. Por ejemplo, si desea ver las primero cinco tiendas del país y dos de ellas presentan las mismas ventas, los datos que se devuelvan pueden parecerse a:

RANKINDEX	SALES	STORE
1	10000	Orange
1	10000	Brea
3	90000	Los Angeles

RANKINDEX	SALES	STORE
4	80000	Ventura

RANKINDEX es solo un puerto de salida. Puede pasar el índice de rango a otra transformación de la asignación o directamente a un destino.

Puerto de rango

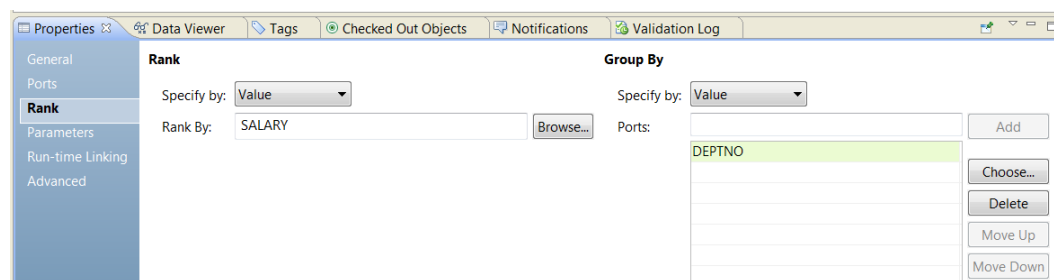
El puerto de rango determina la columna por la que se van a clasificar los valores.

Debe designar un puerto de entrada/salida o de salida como el puerto de rango. Por ejemplo, cree una transformación de rango para clasificar a los primeros empleados de cada departamento según sus salarios. El puerto de salario contiene el salario de cada empleado. Puede designar puerto de entrada/salida de salario como puerto de rango.

Seleccione el puerto de rango en la ficha **Rango** de la vista **Propiedades**. Puede utilizar un parámetro para el puerto de rango. Para ello, seleccione **Especificar por parámetro**. Busque o cree un parámetro de puerto. El valor predeterminado del parámetro es el nombre de un puerto o de un puerto generado.

Debe enlazar el puerto de rango con otra transformación.

La siguiente imagen muestra la ficha **Rango**:



Nota: El puerto de rango no admite los tipos de datos binarios.

Definir agrupaciones por puertos

Puede configurar la transformación de rango para crear grupos de filas con rango.

Por ejemplo, si desea seleccionar los diez productos más caros por fabricante, en primer lugar debería establecer un grupo para cada fabricante. En el panel **Agrupar por** de la ficha **Rango**, puede definir puertos de entrada, de entrada/salida o de salida como la agrupación por puertos.

Por cada valor único del puerto de grupo, la transformación crea un grupo de filas dentro de la definición de rango (superior o inferior y un número en concreto en cada rango).

La transformación de rango cambia el número de filas de dos formas diferentes. Al filtrar casi todas las filas de un rango superior o inferior, se reduce el número de filas que se transfieren a la transformación. En la definición de grupos, cree un grupo de filas con rango para cada grupo.

Por ejemplo, si crea una transformación de rango que clasifica los cinco mejores vendedores agrupados por trimestre, el índice de rango enumera los vendedores del 1 al 5 de cada trimestre:

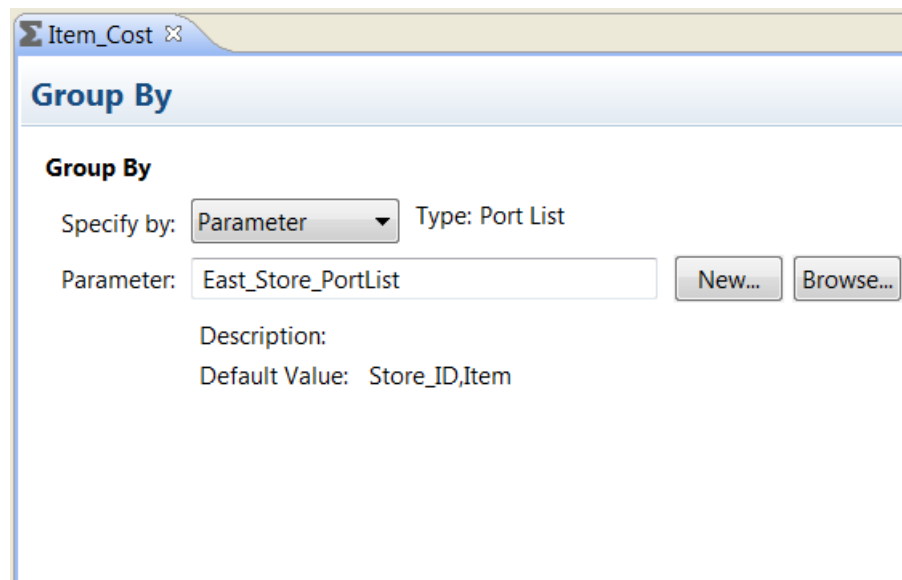
RANKINDEX	SALES_PERSON	SALES	QUARTER
1	Sam	10,000	1
2	Mary	9,000	1
3	Alice	8,000	1
4	Ron	7,000	1
5	Alex	6,000	1

Establezca el número de filas que quiera incluir en una clasificación en la ficha **Avanzado** de la vista **Propiedades**.

Parámetros de la agrupación por puertos

Puede configurar un parámetro de lista de puertos que contenga uno o más puertos que quiera incluir en un grupo. Para crear un parámetro de lista de puertos, seleccione puertos de una lista de puertos en la transformación.

La siguiente imagen muestra la ficha **Agrupar por** cuando utiliza un parámetro para identificar los puertos en el grupo:



Puede buscar un parámetro de lista de puertos o hacer clic en **Nuevo** para crear un parámetro de lista de puertos. Si decide crear un parámetro de lista de puerto, puede seleccionar los puertos de una lista de puertos en la transformación.

Memorias caché de rango

Cuando se ejecuta una asignación que utiliza una transformación de rango, el Servicio de integración de datos crea una memoria caché de índice y una memoria caché de datos en la memoria para procesar la transformación. Si el Servicio de integración de datos necesita más espacio del que hay disponible en la memoria caché, almacena los datos desbordados en archivos de memoria caché.

Al ejecutar una asignación que utiliza una transformación de rango, el Servicio de integración de datos compara una fila de entrada con las filas de la memoria caché de datos. Si la fila de entrada deja fuera de rango una fila de la memoria caché, el Servicio de integración de datos sustituye ésta por la fila de entrada. Si configura la transformación de rango a filas del grupo, el Servicio de integración de datos establece un rango en las filas de cada grupo.

El Servicio de integración de datos crea las siguientes memorias caché para la transformación de rango:

- Memoria caché de índice que almacena valores de grupo tal y como están configurados en los puertos de agrupación.
- Memoria caché de datos que almacena información basada en los puertos de agrupación.

Propiedades avanzadas de la transformación de rango

Configure propiedades para ayudar a determinar cómo el Servicio de integración de datos procesa los datos para la transformación de rango.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Superior/Inferior

Especifica si desea el rango superior o inferior de una columna.

Número de rangos

Número de filas para incluir en el rango superior o inferior.

Comparación de cadenas que distinguen mayúsculas de minúsculas

Especifica si el Servicio de integración de datos utiliza comparaciones de cadenas que distinguen mayúsculas de minúsculas cuando crea rangos de cadenas. Desactive esta opción para hacer que el Servicio de integración de datos ignore esta distinción en las cadenas. Esta opción está seleccionada de forma predeterminada.

Directorio de la memoria caché

Directorio en el que el Servicio de integración de datos crea los archivos de memoria caché de índice y de datos. Compruebe que el directorio existe y que tiene suficiente espacio en disco para los archivos de memoria caché.

Introduzca varios directorios separados por punto y coma para aumentar el rendimiento durante la creación de particiones en la memoria caché. La creación de particiones en la memoria caché crea una memoria caché independiente para cada partición que procesa la transformación.

El valor predeterminado es el parámetro del sistema CacheDir. Para esta propiedad, puede configurar otro parámetro del sistema o definido por el usuario.

Tamaño de la memoria caché de datos de rango

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de datos para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Tamaño de la memoria caché de índice de rango

Cantidad de memoria que el Servicio de integración de datos asigna a la memoria caché de índice para la transformación al inicio de la ejecución de la asignación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Auto.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

TEMAS RELACIONADOS

- [“Tamaño de la memoria caché” en la página 74](#)

Transformación de rango en un entorno no nativo

El procesamiento de transformaciones de rango en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de rango en el motor de Blaze

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

La memoria caché de datos de la transformación de rango se ha optimizado para que utilice la longitud variable para almacenar los tipos de datos binarios y de cadena que pasen por la transformación de rango. La optimización está habilitada en tamaños de registros de hasta 8 MB. Si el tamaño de los registros es mayor que 8 MB, la optimización de la longitud variable estará deshabilitada.

Si se usa la longitud variable para almacenar datos que pasan por la transformación de rango en la memoria caché de datos, la transformación de rango se optimiza para usar la entrada ordenada, y se inserta una transformación de ordenación de paso antes de la transformación de rango en la asignación en tiempo de ejecución. Para ver la transformación de ordenación, vea la asignación optimizada o vea el plan de ejecución en el entorno de validación de Blaze.

Durante la optimización de la memoria caché de datos, la memoria caché de datos y la memoria caché de índice de la transformación de rango están establecidas en Automático. La memoria caché de ordenación de la transformación de ordenación está establecida en el mismo tamaño que la memoria caché de datos de la

transformación de rango. Para configurar la memoria caché de ordenación, debe configurar el tamaño de la memoria caché de datos de la transformación de rango.

Transformación de rango en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La distinción entre mayúsculas y minúsculas está deshabilitada.

Optimización de la memoria caché de datos

La memoria caché de datos no se puede optimizar para que la transformación almacene datos usando la longitud variable.

Transformación de rango en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- Una transformación de rango está en el mismo canal de transmisión que una transformación de búsqueda pasiva configurada con una condición de búsqueda de desigualdad.
- Una transformación de rango está en un nivel superior a una transformación de combinación.
- Un canal de transmisión contiene más de una transformación de rango.
- Un canal de transmisión contiene una transformación de agregación y una transformación de rango.

Transformación de rango en el motor de Databricks Spark

Algunas reglas de procesamiento del motor de Databricks Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La distinción entre mayúsculas y minúsculas está deshabilitada.

Optimización de la memoria caché de datos

La memoria caché de datos no se puede optimizar para que la transformación almacene datos usando la longitud variable.

CAPÍTULO 37

Transformación de lectura

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de lectura, 583](#)
- [Propiedades de la transformación de lectura , 584](#)
- [Sincronizar objetos de datos relacionales, 587](#)
- [Cambiar el objeto de datos de origen, 588](#)
- [Parámetros de la transformación de lectura, 590](#)
- [Restricciones , 591](#)
- [Crear una transformación de lectura, 591](#)

Resumen de la transformación de lectura

La transformación de lectura es una transformación pasiva que lee datos de un origen. Las transformaciones de lectura no se pueden reutilizar.

Las transformaciones de lectura se pueden crear a partir de un objeto de datos físicos o un objeto de datos lógicos. Si desea crear una transformación de lectura a partir de un objeto de datos físico que haya importado desde un origen de adaptador de PowerExchange, el editor de asignaciones podría solicitarle que especifique una operación de lectura antes de poder crear una transformación de lectura desde el objeto de datos.

Se pueden configurar diferentes propiedades de una transformación de lectura según el tipo de objeto de datos que haya utilizado para crear la transformación. Por ejemplo, si crea una transformación de lectura a partir de un objeto de datos relacionales, podrá configurar reemplazos de SQL y definir restricciones. Las propiedades que puede configurar también dependen de si hay parámetros configurados para la transformación.

Las transformaciones de lectura pueden contener orígenes dinámicos. Puede configurar una transformación de lectura de forma que actualice dinámicamente sus puertos, metadatos y otras propiedades. Para obtener información sobre cómo configurar orígenes dinámicos, consulte el capítulo sobre asignaciones dinámicas de la *Guía de asignación de Informática Developer*.

Propiedades de la transformación de lectura

Después de crear una transformación de lectura, puede configurar sus propiedades.

Las propiedades de la transformación de lectura se configuran en las fichas de propiedades. Las fichas disponibles dependerán del tipo de origen que la transformación de lectura representa.

En la siguiente tabla se describe cada ficha de propiedad y se indica el tipo de origen correspondiente a cada ficha:

Ficha de propiedad	Descripción	Tipo de origen
General	Sirve para especificar las propiedades y el comportamiento de la transformación. En los orígenes de objeto de datos relacionales y personalizados, sincronice los puertos de entrada de la transformación con el origen.	Todo
Objeto de datos	Sirve para especificar el origen de datos de la transformación.	<ul style="list-style-type: none">- Relacional- Archivo sin formato- Objeto de datos personalizados
Puertos	Sirve para establecer la definición de puerto según el objeto de datos asociado.	<ul style="list-style-type: none">- Relacional- Objeto de datos personalizados- Objeto de datos lógicos
Formato	Sirve para introducir la configuración de un origen de datos de archivo sin formato.	Archivo sin formato
Consulta	Sirve para especificar una consulta al origen.	<ul style="list-style-type: none">- Relacional- Objeto de datos personalizados- Objeto de datos lógicos
Tiempo de ejecución	Sirve para definir el comportamiento en tiempo de ejecución.	<ul style="list-style-type: none">- Relacional- Archivo sin formato- Objeto de datos personalizados
Orígenes	Sirve para seleccionar las tablas de origen y configurar los detalles de origen.	<ul style="list-style-type: none">- Relacional- Objeto de datos personalizados
Parámetros de objeto de datos	Sirve para establecer las propiedades de parámetros.	<ul style="list-style-type: none">- Archivo sin formato- Objeto de datos personalizados- Objeto de datos lógicos

Ficha de propiedad	Descripción	Tipo de origen
Vinculación en tiempo de ejecución	Sirve para configurar un vínculo de grupo a grupo entre las transformaciones que usan un parámetro, una directiva de vínculo o ambos, a fin de saber qué puertos vincular en tiempo de ejecución.	Todo
Avanzado	Sirve para establecer el nivel de seguimiento y el orden de fila. En un origen relacional, puede establecer la opción para crear o reemplazar la tabla de destino en tiempo de ejecución.	Todo

Propiedades generales

Puede configurar el nombre y la descripción de la transformación de lectura. También puede configurar las siguientes propiedades:

Cuando cambien los metadatos de columna

Disponible en los orígenes relacionales. Seleccione una de las siguientes opciones:

- Sincronizar puertos de salida. Developer tool actualiza los puertos de salida de la transformación de lectura con los cambios de metadatos que el repositorio de modelos almacena en relación con el objeto de datos.
- No sincronizar. La transformación de lectura no refleja los cambios de metadatos en el objeto de datos.

Objeto de datos físicos

Disponible en los orígenes de archivo sin formato y personalizados. Es el objeto que se ha utilizado para crear la transformación.

Puede seleccionar el nombre del objeto de datos y configurar sus propiedades.

Propiedades del objeto de datos

En la ficha Objeto de datos, puede especificar o cambiar el origen de la transformación de lectura, así como convertir en dinámicos los orígenes de objeto de datos relacionales, personalizados y de archivo sin formato.

Puede configurar las siguientes propiedades:

Especificar por

Para especificar los metadatos y columnas de origen de la transformación de lectura, seleccione una de las siguientes opciones:

- Valor. La transformación de lectura utiliza el objeto de datos asociado para especificar los metadatos y columnas de origen.
- Parámetro. La transformación de lectura utiliza un parámetro para especificar los metadatos y columnas de origen.

Objeto de datos

Si ha creado la transformación de lectura a partir de un objeto de datos existente, el campo mostrará el nombre de ese objeto. Haga clic en **Examinar** para cambiar el objeto de datos y asociarlo a la transformación de lectura.

En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos

Si habilita esta opción, el Servicio de integración de datos obtiene los cambios de definición de datos y metadatos de las tablas de destino para la transformación de lectura.

Propiedades de consulta

Configure una consulta SQL a un objeto de datos personalizados o recurso relacional.

Al configurar las propiedades de la ficha **Consulta**, decida entre configurar las propiedades simples o avanzadas.

En la vista de propiedades **Simple**, puede configurar la instrucción SQL predeterminada como una instrucción DISTINCT y editar las condiciones de sugerencia, combinación, filtro y ordenación de la instrucción.

En la vista de propiedades **Avanzado**, puede definir una consulta SQL personalizada. Puede seleccionar valores de las columnas del objeto de datos asociado o de parámetros, o bien crear un parámetro de cero para representar un objeto de datos.

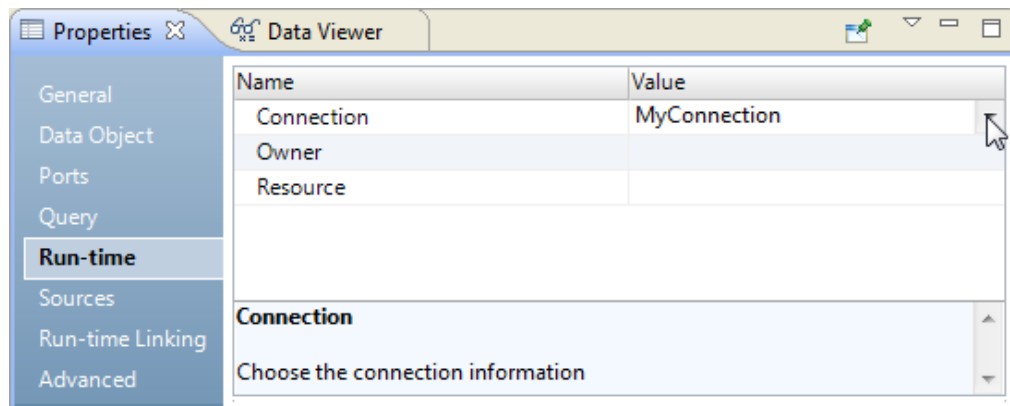
Propiedades de tiempo de ejecución

Puede configurar las siguientes propiedades de transformación de lectura en la ficha **Tiempo de ejecución**:

Conexión

Disponible en los orígenes relacionales. Conexión que utiliza la transformación. Haga clic a la derecha del campo para cambiar la conexión.

La siguiente imagen muestra la ubicación del botón desplegable donde debe hacer clic:



Propiedades de los orígenes

Configure los detalles de los orígenes de los objetos de datos personalizados y recursos relacionales. Puede cambiar la definición del objeto de datos relacionales después de importarlo al repositorio. Puede añadir y quitar puertos, definir claves principales y configurar las relaciones entre varios objetos de datos relacionales en el repositorio.

En la ficha **Orígenes** puede configurar los siguientes ajustes:

Todos los orígenes

Utilice los botones Añadir y Quitar para añadir y quitar los orígenes adicionales de la transformación.

Ficha General

Cambie el nombre y la descripción del origen seleccionado. Haga clic en el nombre de origen para cambiar otros detalles.

Ficha Claves

Designe las columnas de recurso como claves.

Ficha Relaciones

Añada y quite relaciones entre varios recursos relacionales.

Propiedades avanzadas

Configure propiedades avanzadas para determinar la manera en que el servicio de integración de datos procesa los datos de la transformación de lectura.

Configure las siguientes propiedades en la ficha Avanzadas:

Nivel de seguimiento

Controle la cantidad de detalle incluido en el archivo de registro de asignación.

PreSQL

Comando SQL que el servicio de integración de datos ejecuta en la base de datos de origen antes de leer el origen.

Developer tool no valida el SQL.

PostSQL

Comando SQL que el servicio de integración de datos ejecuta en la base de datos de origen después de escribir en el destino.

Developer tool no valida el SQL.

Restricciones

Instrucciones SQL para las restricciones de integridad referencial en el nivel de tabla. Solo se aplica a los orígenes relacionales.

Sincronizar objetos de datos relacionales

Puede sincronizar objetos de datos físicos cuando sus orígenes cambien. Cuando sincronice un objeto de datos físicos, Developer tool volverá a importar los metadatos del objeto desde el origen que se haya seleccionado.

Puede sincronizar todos los objetos de datos físicos. Cuando sincronice objetos de datos relacionales u objetos de datos personalizados, puede conservar o reemplazar las relaciones de clave que se hayan definido en Developer tool.

Elija entre los distintos métodos disponibles para sincronizar objetos de asignación:

Sincronizar un recurso relacional.

Para sincronizar un objeto de datos físicos, haga clic con el botón derecho en el objeto en la vista **Explorador de objetos** y seleccione **Sincronizar**.

Sincronizar puertos de transformación con el objeto de datos físicos.

En la ficha Objeto de datos de una transformación, seleccione la opción **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos**.

En tiempo de ejecución, el servicio de integración de datos obtiene los cambios de definición de datos y metadatos del origen de datos y actualiza la definición de objetos de datos en el Repositorio de modelos.

Para obtener una vista previa del modo en que el servicio de integración de datos obtiene los cambios de definición de datos y metadatos, vea la asignación con los parámetros resueltos.

Sincronizar puertos cuando los metadatos cambien.

En la ficha General de una transformación, seleccione la opción para sincronizar puertos. El nombre exacto de esta opción dependerá del tipo de transformación que configure. Por ejemplo, en una transformación de lectura, la opción es **Cuando los metadatos cambien, sincronizar puertos de salida**.

Cuando se ejecute la asignación, el servicio de integración de datos sincronizará los metadatos de columna de la transformación con los metadatos del origen de datos.

Cambiar el objeto de datos de origen

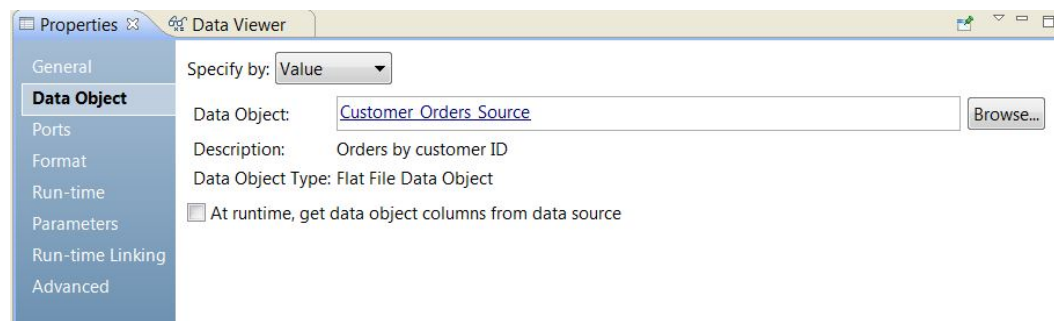
Una transformación de lectura se basa en un objeto de datos físicos o en un objeto de datos lógicos del repositorio de modelos. Puede cambiar el objeto de datos cuando configure una transformación de lectura. Puede parametrizar el objeto de datos para cambiar el objeto de datos en tiempo de ejecución. Así, por ejemplo, puede probar una asignación con un archivo de origen distinto del archivo de origen que use en una ejecución de asignación de producción.

Al crear una transformación a partir de un objeto de datos físicos, la información sobre el objeto de datos aparece recogida en la ficha **Objeto de datos** de las propiedades de la transformación. Puede hacer clic en el nombre del objeto de datos para ver la definición del objeto de datos físicos del repositorio de modelos.

Si desea cambiar el objeto de datos de la transformación, busque otro objeto de datos físicos en el repositorio de modelos. Cuando se cambia el objeto de datos, la transformación utiliza las propiedades de tiempo de ejecución y propiedades avanzadas del objeto de datos que seleccione.

La estructura del objeto de datos se puede actualizar en tiempo de ejecución según los cambios en el origen de datos. El origen de datos es el archivo físico o la tabla de base de datos que el objeto de datos representa. Cuando se habilita el servicio de integración de datos para obtener columnas de datos del origen de datos, dicho servicio examina la estructura del origen de datos. Luego, actualizará los puertos de objeto de datos en la instancia de transformación según el origen de datos. El servicio de integración de datos no cambia la definición del objeto de datos físicos del repositorio de modelos.

En la siguiente imagen se muestra la ficha **Objeto de datos**:



La ficha **Objeto de datos** tiene los siguientes campos:

Especificar por

Elija **Valor** para introducir un nombre de objeto de datos específico. Elija **Parámetro** para parametrizar el objeto de datos.

Objeto de datos

Nombre del objeto de datos en el repositorio de modelos. Puede hacer clic en el vínculo **Objeto de datos** para abrir la definición del objeto de datos del repositorio de modelos. También puede hacer clic en **Examinar** para buscar otro objeto de datos en el repositorio de modelos.

Descripción

Descripción del objeto de datos en el repositorio. Solo lectura.

Tipo de objeto de datos

Describe el tipo del objeto de datos, como archivo sin formato, relacional o personalizado.

En tiempo de ejecución, obtener las columnas de objetos de datos del origen de datos

El servicio de integración de datos obtiene los cambios de definición de datos y metadatos del archivo de datos o de la tabla a la que hace referencia el objeto de datos, y actualiza la estructura del objeto de datos de la instancia de transformación en tiempo de ejecución.

Para obtener una vista previa del modo en que el servicio de integración de datos obtiene los cambios de definición de datos y metadatos en tiempo de ejecución, vea la asignación con los parámetros resueltos.

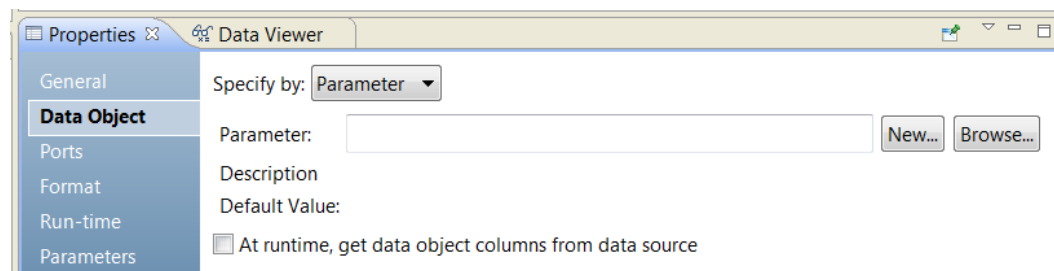
Parametrizar la transformación de lectura

Puede parametrizar la transformación de lectura y cambiar el objeto de datos en tiempo de ejecución.

Para parametrizar un objeto de datos, elija **Especificar por parámetro** en la ficha **Objeto de datos**. Las propiedades en la ficha **Objeto de datos** cambian.

Para parametrizar el objeto de datos, cree un parámetro de tipo de recurso o busque un parámetro de recurso que ya haya creado. El valor predeterminado de parámetro es el nombre del objeto de datos físicos en el repositorio de modelos. Al crear un valor de parámetro predeterminado, hay que seleccionar un nombre de objeto de datos físicos de una lista de objetos de datos en el repositorio.

La siguiente imagen muestra la ficha **Objeto de datos** cuando el objeto de datos se especifica mediante un parámetro:



La ficha **Objeto de datos** tiene las siguientes opciones por parámetro:

Parámetro

Nombre de un parámetro de recurso que ha configurado como el objeto de datos. Solo lectura.

Descripción

Descripción del parámetro. Solo lectura.

Nuevo

Cree un parámetro de recurso. Busque y seleccione un objeto de datos en el repositorio de modelos como el valor predeterminado de parámetro.

Examinar

Busque un parámetro de recurso y seleccione el parámetro.

Valor predeterminado

Valor predeterminado del parámetro de recurso configurado para el objeto de datos. El valor predeterminado es un nombre de objeto de datos físicos y la ruta a ese objeto en el repositorio de modelos. Solo lectura.

Parámetros de la transformación de lectura

Puede parametrizar algunas de las propiedades de una transformación de lectura, así como algunas de las propiedades del objeto de datos físicos reutilizable a partir del cual se crea la transformación de lectura.

Al crear un objeto de datos físicos, puede configurar las propiedades de lectura y escritura. Los parámetros que configure en las propiedades de lectura de un objeto de datos físicos aparecen en la ficha **Parámetros de objeto de datos** de la transformación de lectura al agregar el objeto de datos a una asignación.

Puede configurar los parámetros de las siguientes propiedades de lectura del objeto de datos:

- Directorio del archivo de control
- Nombre del archivo de control
- Escala predeterminada
- Delimitador
- Delimitador de archivo sin formato
- Directorio del archivo de fusión
- Nombre del archivo de origen
- Directorio del archivo de origen

Después de agregar el objeto de datos físicos a una asignación, puede consultar los parámetros en la ficha **Parámetros de objeto de datos** de la transformación de lectura. Puede exponer estos parámetros como parámetros de asignación para que reemplacen los valores de parámetro en tiempo de ejecución.

Nota: No pueden anidar parámetros definidos por el usuario dentro de un origen parametrizado. Si el objeto de datos de origen es parametrizado, no puede exponer un parámetro definido por el usuario como un parámetro de asignación para anular los valores de parámetro en tiempo de ejecución. La asignación usa el valor predeterminado en su lugar.

Puede configurar los siguientes parámetros de asignación de la transformación de lectura:

- Conexión (relacional)
- Objeto de datos
- Orden de resolución de vínculos
- Nombre de recurso (relacional)

- Nombre de propietario de tabla (relacional)

Puede ver estos parámetros en la ficha **Parámetros de objeto de datos** de la asignación.

Restricciones

Una restricción es una expresión condicional que deben cumplir los valores de una fila de datos.

Al definir una restricción, hay que especificar una expresión que dé como resultado TRUE en cada fila de datos.

El Servicio de integración de datos puede leer las restricciones de los orígenes relacionales, de los objetos de datos lógicos, de los objetos de datos físicos o de las tablas virtuales. Para establecer una restricción en un objeto de datos físicos reutilizable, cree un objeto de datos personalizados.

Cuando el Servicio de integración de datos lee las restricciones, puede quitar filas que no tienen como resultado TRUE en las filas de datos según el método de optimización aplicado.

Antes de establecer una restricción, debe comprobar que los datos de origen cumplen la condición definida por la restricción. Por ejemplo, una base de datos de origen tiene una columna EDAD que tiene filas con $EDAD < 70$. Puede establecer una restricción con $EDAD < 70$ en la base de datos de origen. El Servicio de integración de datos lee registros de la base de datos de origen con la restricción $EDAD < 70$. Si el Servicio de integración de datos lee registros con $EDAD \geq 70$, puede quitar las filas con $EDAD \geq 70$.

En la base de datos, puede usar comandos SQL para definir las restricciones en el entorno de la base de datos cuando se conecta a la base de datos. El Servicio de integración de datos ejecuta el SQL del entorno de conexión cada vez que se conecta a la base de datos.

Crear una transformación de lectura

Al crear una transformación de lectura, debe elegir uno de los siguientes métodos en función del recurso a partir del que se crea la transformación:

Crear la transformación de un objeto de datos en el repositorio de modelos.

Realice los siguientes pasos para crear una transformación de lectura a partir de un objeto de datos del repositorio de modelos:

1. Abra la asignación en el editor.
2. Arrastre un objeto de datos desde el **Explorador de objetos** a la vista del editor.
3. Seleccione **Lectura** y haga clic en **Aceptar**.

La transformación de lectura en la asignación contiene los puertos y las propiedades del objeto de datos.

Crear la transformación con el editor de asignación.

Utilice este método si desea definir una configuración detallada de la transformación de lectura. También puede recurrir a este método para basar una transformación de lectura en un parámetro.

Para crear una transformación de lectura en el editor de asignación, consulte [“Crear una transformación de lectura en el editor de asignación” en la página 592](#).

Crear una transformación de lectura en el editor de asignación

Puede crear una transformación de lectura que represente el origen de los datos, de los metadatos de columna y de las propiedades en una asignación.

Realice los pasos siguientes:

1. Seleccione uno de los siguientes métodos para crear una transformación de lectura:
 - Haga clic con el botón derecho en el editor de asignación y seleccione **Añadir transformación**. Se abre el cuadro de diálogo **Añadir transformación**. Seleccione Transformación de lectura y haga clic en **Siguiente**.
 - Desplácese hacia abajo en la paleta de asignaciones para localizar el icono de transformación de lectura y haga doble clic en él. La siguiente imagen muestra el icono de transformación de lectura:



Se abre el cuadro de diálogo **Nueva transformación de lectura**.

2. Haga lo siguiente para utilizar un archivo sin formato, un recurso relacional o un objeto de datos personalizados como origen:
 - a. Seleccione **Objeto de datos físicos** como tipo de objeto de datos.
 - b. Haga clic en **Examinar** para seleccionar un archivo sin formato, un recurso relacional o un objeto de datos personalizados. Se abre la ventana **Seleccionar objeto de datos**.
 - c. Seleccione un objeto de datos y haga clic en **Aceptar**.
 - d. Opcionalmente, configure la transformación para que recupere columnas de objetos de datos del origen en tiempo de ejecución. Seleccione **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos**. El servicio de integración de datos actualizará los metadatos de columna de la transformación de lectura cuando se ejecute la asignación.
3. Haga lo siguiente para utilizar un parámetro como origen:
 - a. Seleccione **Crear usando un parámetro**.
 - b. Haga clic en **Nuevo** para crear un parámetro o en **Examinar** para seleccionar uno ya existente.
 - c. Seleccione un parámetro y haga clic en **Aceptar**.
 - d. Opcionalmente, configure la transformación para que recupere columnas de objetos de datos del origen en tiempo de ejecución. Seleccione **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos**. El servicio de integración de datos actualizará los metadatos de columna de la transformación de lectura cuando se ejecute la asignación.

4. Haga lo siguiente para utilizar un objeto de datos lógicos como origen:
 - a. Seleccione **Objeto de datos lógicos** como tipo de objeto de datos.
 - b. Haga clic en **Examinar** para seleccionar un objeto de datos y, a continuación, haga clic en **Aceptar**.
5. Opcionalmente, escriba un nombre para la transformación de lectura.
6. Haga clic en **Finalizar**.

CAPÍTULO 38

Transformación de relacional a jerárquica

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de relacional a jerárquica, 594](#)
- [Ejemplo: transformación de relacional a jerárquica, 595](#)
- [Puertos relacionales de entrada y vista Resumen, 597](#)
- [Puertos de transformación de relacional a jerárquica, 598](#)
- [Referencias a esquemas, 598](#)
- [Desarrollo de la transformación de relacional a jerárquica, 598](#)

Resumen de la transformación de relacional a jerárquica

La transformación de relacional a jerárquica procesa una entrada relacional y la transforma en salida jerárquica. Una transformación de relacional a jerárquica lee la entrada relacional de los puertos de entrada y transforma los datos en salidas jerárquicas en el puerto de salida de transformación. Para transformar la entrada relacional en salida jerárquica, utilice un objeto de esquema para definir la estructura jerárquica.

Puede utilizar el asistente de transformación de relacional a jerárquica para crear una estructura jerárquica que refleje los puertos de entrada relacionales. Puede ver la asignación para los puertos de salida jerárquica en la vista **Resumen** de la transformación.

Después de crear la transformación, puede pasar los datos desde el puerto de salida jerárquica a otra transformación de una asignación.

Nota: La transformación de relacional a jerárquica puede procesar hasta 10 000 elementos de esquema en un archivo .xsd. Para procesar más de 10 000 elementos, divida los datos en varios archivos.

En el modelo relacional, cada esquema de la tabla identifica una columna, llamada clave principal, para identificar cada fila de forma exclusiva. Puede identificar la relación entre cada fila de la tabla y una fila de otra tabla con una clave externa. El asistente genera claves cuando crea la transformación. Puede cambiar una transformación generada automáticamente y añadir, editar o eliminar puertos.

Puede vincular un puerto relacional de entrada con un nodo en la jerarquía. Vincule una clave principal desde el elemento o atributo relevantes en la jerarquía con un grupo relacional en la entrada. La clave principal identifica cada fila en las tablas relacionales.

Vincule una clave externa desde el elemento o atributo relevantes en la jerarquía con un grupo relacional en la entrada. Una clave externa en la entrada relacional es una columna en una tabla que apunta a la clave principal de otra tabla.

El puerto relacional de entrada y el nodo jerárquico deben tener tipos de datos compatibles.

Ejemplo: transformación de relacional a jerárquica

El departamento de finanzas de la empresa Electronics Superstore debe procesar las nóminas para los empleados de la compañía. Necesitan transformar los datos de los empleados, que se encuentran almacenados en una base de datos relacional, a un formato jerárquico que su sistema de pago pueda procesar.

La asignación debe utilizar una transformación de relacional a jerárquica que permita la entrada de la información de los empleados, como nombre del empleado, ID, dirección y cuenta bancaria, y genere la información en un formato jerárquico utilizable.

En la entrada relacional, el elemento Bank_ID es una clave principal en la tabla Empleado y una clave externa en la tabla Banco:

Employee_ID	Last_Name	First_Name	Dirección	Bank_ID	Bank_Account
9173327437	Sandrine	Jacques	74 Mobile Avenue	74845	8723487234
9174562342	Race	Tom	266 Crouse St.	9234734	45324734
8484526471	Jones	Charles	3815 LaValle Boulevard	389236	234638437
7023847265	Smith	Delilah	193 Short Drive	74845	8723463432
9174596725	Frederick	George	17 Serenity Road	9234734	6342636699

Bank_ID	Bank_Name	SWIFT_Code
74845	National Bank	9173327
9234734	International Bank	9174562
389236	Star National Bank	8484526

En la salida Pago en formato jerárquico, los elementos se combinan desde las tablas:

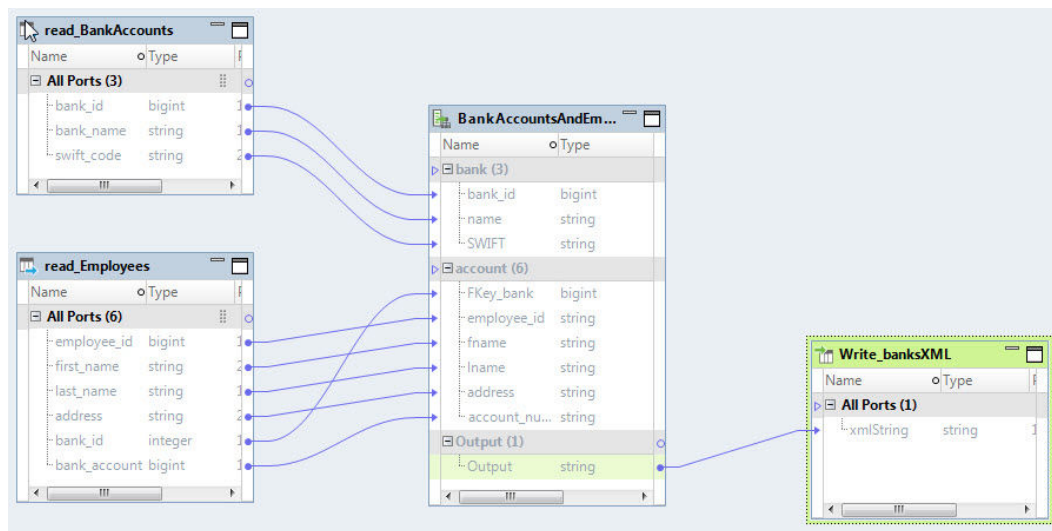
```
<banks>
  <bank name="National Bank" SWIFT="9173327">
    <account id="8723487234">
      <employee_id>9173327437</employee_id>
      <fname>Sandrine</fname>
      <lname>Jacques</lname>
      <address>74 Mobile Avenue</address>
    </account>
    <account id="8723463432">
      <employee_id>9082745558</employee_id>
      <fname>Delilah</fname>
      <lname>Smith</lname>
      <address>193 Short Drive</address>
    </account>
  </bank>
  <bank name="International Bank" SWIFT="9174562">
    <accounts>
      <account id="45324734">
        <employee_id>5534398889</employee_id>
        <fname>Race</fname>
        <lname>Tom</lname>
        <address>266 Crouse St.</address>
      </account>
      <account id="6342636699">
        <employee_id>9174596725</employee_id>
        <fname>Frédéric</fname>
        <lname>George</lname>
      </account>
    </accounts>
  </bank>
</banks>
```

```

        <address>17 Serenity Road</address>
      </account>
    </accounts>
  </bank>
  <bank name="Star National Bank" SWIFT="8484526">
    <accounts>
      <account id="234638437">
        <employee_id>8484526471</employee_id>
        <fname>Jones</fname>
        <lname>Charles</lname>
        <address>3815 LaValle Boulevard</address>
      </account>
    </accounts>
  </bank>
</banks>

```

La siguiente imagen muestra la asignación en este ejemplo:



La asignación contiene los siguientes objetos:

Read_BankAccounts

El origen que contiene la información bancaria.

Read_Employees

El origen que contiene los datos de los empleados.

BankAccountsAndEmployees_To_PaymentsSystemXML

Una transformación de relacional a jerárquica que transforma una entrada relacional que contiene la información del empleado y la información bancaria en un formato XML que el sistema de pagos puede procesar.

Write_BanksXML

Una ruta de acceso de destino al archivo que almacena los datos transformados cada vez que ejecuta la asignación.

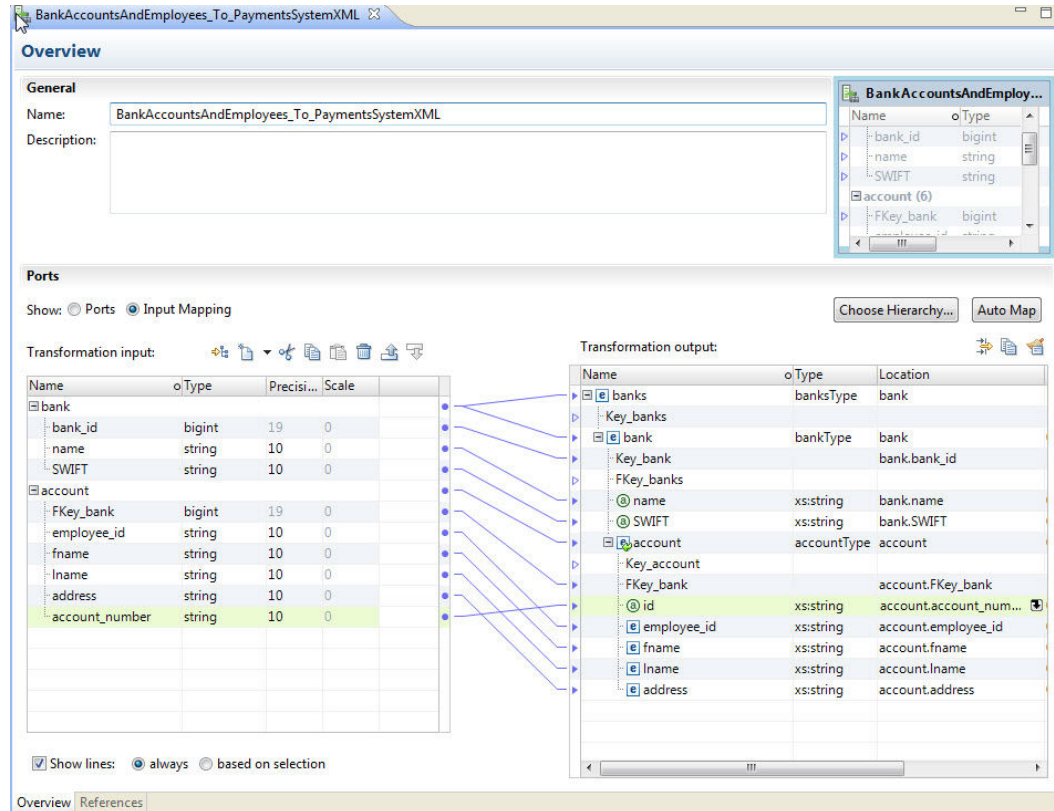
La asignación utiliza los archivos Read_BankAccount y Read_Employees para proporcionar la entrada relacional. La asignación procesa y transforma los datos mediante la transformación BankAccountsAndEmployees_To_PaymentsSystemXML. A continuación, la asignación almacena la salida en la ruta de acceso de destino que aparece en el archivo sin formato Write_BanksXML.

Puertos relacionales de entrada y vista Resumen

Para transformar datos relacionales en datos jerárquicos, el asistente crea una estructura jerárquica que refleja los puertos de entradas relacionales. Puede utilizar la vista **Resumen** para vincular los puertos relacionales a los puertos jerárquicos.

Para ver los vínculos entre la entrada relacional y la salida jerárquica, utilice la vista **Resumen**. Seleccione **Asignación de entrada**. El panel **Puertos** aparece en la vista **Resumen**.

La siguiente imagen muestra el panel **Puertos**:



A la izquierda del panel **Puertos** está el panel **Entrada de transformación** que contiene los elementos y grupos relacionales. A la derecha está el área **Salida de transformación** que contiene los nodos de esquema jerárquicos.

Puede crear puertos en el área **Entrada de transformación** y vincular elementos relacionales con los nodos del esquema. También puede arrastrar el puntero desde un nodo del esquema hacia un campo vacío del área **Entrada de transformación** para crear un puerto. Cuando conecte un puerto relacional con un nodo de esquema, Developer tool muestra un vínculo entre ellos.

Puertos de transformación de relacional a jerárquica

Los puertos de la transformación de relacional a jerárquica se definen en la vista **Resumen** de la transformación.

Una transformación de relacional a jerárquica puede leer salidas de datos relacionales desde un búfer. Los puertos de salida devuelven datos jerárquicos a un búfer.

Referencias a esquemas

Una transformación de jerárquica a relacional requiere un esquema jerárquico para definir la jerarquía de salida de la transformación. Para utilizar el esquema en la transformación, puede definir una referencia de esquema.

Puede definir referencias de esquema de la transformación en la vista **Referencias** de la transformación.

La transformación de relacional a jerárquica hace referencia a objetos de esquema en el repositorio de modelos. Los objetos de esquema pueden existir en el repositorio antes de crear la transformación.

Un esquema puede hacer referencia a esquemas adicionales. La vista **Referencias** muestra el espacio de nombres y el prefijo para cada esquema al que hace referencia la transformación de relacional a jerárquica.

Desarrollo de la transformación de relacional a jerárquica

Utilice el asistente de Nueva transformación para generar automáticamente una transformación de relacional a jerárquica. Elija un archivo de ejemplo de esquema o jerárquico para definir la jerarquía de salida.

Crear la transformación de relacional a jerárquica

1. En Developer tool, haga clic en **Archivo > Nuevo > Transformación**.
2. Seleccione la transformación de relacional a jerárquica y haga clic en **Siguiente**.
3. Introduzca un nombre para la transformación, busque la ubicación en el repositorio de modelos donde colocarla y haga clic en **Siguiente**.
4. Para seleccionar un esquema, seleccione uno de los siguientes métodos:
 - Para utilizar un esquema del repositorio de modelos para definir la jerarquía de salida, al lado del campo **Objeto de esquema**, examine para seleccionar el archivo de esquema en el repositorio.
 - Para importar un archivo de esquema, haga clic en **Crear un nuevo objeto de esquema**. En la ventana **Nuevo objeto de esquema**, puede examinar y seleccionar un archivo de esquema, o puede seleccionar crear un esquema desde un archivo jerárquico de muestra.
5. Elija la raíz de la jerarquía de salida. En el cuadro de diálogo **Raíz de jerarquía**, seleccione el elemento en el esquema que sea el elemento raíz del archivo jerárquico de salida. Para seleccionar el objeto raíz, puede añadir un archivo jerárquico de muestra. Para añadir un archivo de muestra, junto al campo **Archivo de ejemplo**, busque y seleccione el archivo del sistema de archivos.

6. Haga clic en **Finalizar**.
El asistente crea la transformación en el repositorio.

Crear los puertos

Configure los puertos en la vista **Resumen**.

1. Para ver la asignación, en el área **Puertos** de la vista **Resumen**, seleccione **Asignación de entrada**.
2. Seleccione el tipo de puerto de entrada, la precisión y la escala.
3. Expanda los árboles en la malla **Puertos**. A la izquierda, el panel **Entrada de transformación** muestra la entrada relacional y, a la derecha, el panel **Salida de transformación** muestra la salida jerárquica esperada.
4. Para definir un nodo como raíz, haga clic en **Seleccionar jerarquía**.
Developer tool solo muestra los nodos desde el nivel raíz y debajo del nivel raíz en el área **Entrada de transformación**.
5. Para ver las líneas que conectan los puertos con los nodos jerárquicos, haga clic en **Mostrar líneas**. Elija entre ver todas las líneas de conexión o solo las líneas de los puertos seleccionados.
6. Para añadir un grupo o un puerto de entrada al área **Entrada de transformación**, utilice uno de los siguientes métodos:
 - Arrastre un elemento simple o complejo del área **Salida de transformación** a una columna vacía del área **Entrada de transformación**. Si el nodo es un nodo de grupo, Developer tool agrega un grupo relacional sin puertos.
 - Para añadir un grupo relacional, seleccione una fila y haga clic con el botón derecho para seleccionar **Nuevo > Grupo**.
 - Para añadir un puerto relacional, haga clic con el botón derecho para seleccionar **Nuevo > Campo**.
7. Para borrar la configuración de nodos jerárquica de las ubicaciones de puertos, utilice uno de los siguientes métodos:
 - Seleccione uno o varios nodos en el área **Salida de transformación**, haga clic con el botón derecho y seleccione **Borrar**.
 - Seleccione una o más líneas que conecten los puertos de entrada relacionales con los nodos jerárquicos, haga clic con el botón derecho y seleccione **Eliminar**.
8. Para mostrar los puertos de salida en la jerarquía, haga clic en **Mostrar como jerarquía**. Cada grupo secundario aparece debajo del grupo primario.

CAPÍTULO 39

Transformación de consumidor de servicio web REST

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de consumidor de servicio web REST, 600](#)
- [Configuración de la transformación de consumidor de servicio web REST, 602](#)
- [Métodos HTTP, 604](#)
- [Puertos de la transformación de consumidor de servicio web REST, 607](#)
- [Asignación de entrada de la transformación de consumidor de servicio web REST, 610](#)
- [Asignación de salida de la transformación de consumidor de servicio web REST, 612](#)
- [Propiedades avanzadas de la transformación de consumidor de servicio web REST, 614](#)
- [Creación de la transformación de consumidor de servicio web REST, 615](#)
- [Análisis de un mensaje de respuesta JSON que contiene matrices, 616](#)

Resumen de la transformación de consumidor de servicio web REST

La transformación de consumidor de servicio web REST es una transformación activa que se conecta a un servicio web REST como un cliente del servicio web para acceder a los datos o transformarlos. Utilice una transformación de consumidor de servicio web REST para conectarse a un servicio web REST. La transformación de consumidor de servicio web REST puede enviar una solicitud a un servicio web REST y recibir una respuesta desde un servicio web REST.

La transformación de consumidor de servicio web REST se conecta a un servicio web mediante una URL que defina en la transformación o en una conexión HTTP. También puede utilizar una conexión HTTPS. Las transformaciones de consumidor de servicio web de REST pueden utilizar TLS 1.2, TLS 1.1 o TLS 1.0.

Un servicio web REST contiene un método HTTP para cada acción que el servicio web admite. Cuando el servicio de integración de datos se conecta a un servicio web REST, puede enviar una solicitud para obtener, publicar, colocar o eliminar datos. La solicitud puede actuar sobre recursos individuales o sobre colecciones de recursos. Después de que el servicio de integración de datos envía un mensaje de solicitud, recibe un mensaje de respuesta del servicio web.

La solicitud y los mensajes de respuesta contienen datos XML o JSON con elementos que pueden formar una jerarquía. Cuando una solicitud o un mensaje de respuesta contienen elementos de ocurrencia múltiple,

grupos de elementos constituyen niveles en la jerarquía XML o JSON. Los grupos están relacionados cuando un nivel está anidado dentro de otro.

En la transformación de consumidor de servicio web REST, la entrada y la salida del método definen la estructura de la solicitud y los mensajes de respuesta. La entrada y la salida del método incluyen asignaciones que definen cómo asignar los elementos del mensaje a los puertos de entrada y salida.

La transformación de consumidor de servicio web REST es compatible con el servidor proxy. También puede conectarse a una aplicación de Microsoft SharePoint con la transformación de consumidor de servicio web REST.

Ejemplo

Un almacén en línea define los recursos para una base de datos de productos. La base de datos identifica cada producto por el número de parte.

Los clientes de servicio web acceden a los detalles del producto a través de un servicio web REST. El servicio web utiliza la siguiente URL:

```
http://www.HypoStores.com/products/ProductDetails
```

Debe recuperar los detalles sobre un producto específico, tales como la descripción y el precio unitario, y pasar los detalles en una transformación de nivel inferior en una asignación. Cree una transformación de consumidor de servicio web REST para recuperar los detalles sobre un producto y páselos a otra transformación.

La siguiente tabla muestra los detalles de transformación que puede configurar:

Detalles de transformación	Valor
Método HTTP	Get
URL de base	http://www.HypoStores.com/products/ProductDetails
Puerto de argumento de entrada	Part_No
Puertos de salida	Descripción, Unit_Price
Salida de método	<La estructura del mensaje de respuesta.>

La salida del método incluye una asignación de salida que define cómo se asignan los elementos en el mensaje de respuesta a los puertos de salida.

Cuando el servicio de integración de datos envía la solicitud al servicio web, añade el valor en el puerto de argumento a la URL de base. Por ejemplo, para recuperar los detalles sobre la parte 0716, el servicio de integración de datos emplea la siguiente URL:

```
http://www.HypoStores.com/products/ProductDetails?Part_No=0716
```

Cuando el servicio de integración de datos recibe una respuesta, convierte la descripción del producto y el precio unitario en el mensaje de respuesta para los datos para los puertos de salida.

También puede pasar Part_No como un parámetro y sustituir el valor en mitad de secuencia cuando ejecute la asignación.

Proceso de transformación de consumidor de servicio web REST

La transformación de consumidor de servicio web REST crea un mensaje de solicitud en función de los datos en los puertos de entrada y la entrada del método. Convierte los elementos en el mensaje de respuesta para los datos para los puertos de salida en función de la salida del método.

Los puertos de entrada de la transformación de consumidor de servicio web REST contienen datos relacionales desde las transformaciones de nivel superior en una asignación. El servicio de integración de datos utiliza la entrada del método para convertir datos desde los puertos de entrada a elementos en el mensaje de solicitud.

Para conectarse al servicio web, el servicio de integración de datos lee la URL de base que se configuró en las propiedades de la transformación o la conexión HTTP. Identifica el recurso que desea obtener, enviar, colocar o eliminar añadiendo valores desde los puertos URL o puertos de argumento a la URL de base.

Cuando el servicio de integración de datos recibe una respuesta, pasa los datos en el mensaje de respuesta a los puertos de salida de la transformación. El servicio de integración de datos pasa datos en función de cómo está configurada la salida del método. Los puertos de salida contienen datos relacionales. El servicio de integración de datos envía los datos en los puertos de salida para las transformaciones de nivel inferior en la asignación o al destino.

Configuración de la transformación de consumidor de servicio web REST

Cuando se crea una transformación de consumidor de servicio web REST, debe seleccionar el método HTTP y definir la entrada y la salida del método. Si selecciona el método Get, no puede definir la entrada del método.

Los elementos de entrada en el mensaje de solicitud HTTP se asignan a puertos de entrada. Los elementos de salida en el mensaje de respuesta HTTP se asignan a puertos de salida. La herramienta Developer crea puertos para los elementos de primer nivel.

Cuando configure la transformación, realice las tareas siguientes:

1. Seleccione el método HTTP.
2. Configure los puertos que representan los elementos en el encabezado y cuerpo de los mensajes de solicitud y de respuesta.
3. Configure la asignación de entrada.
4. Configure la asignación de salida.
5. Configure las propiedades avanzadas tales como la conexión y la URL de base para el servicio web.

Si el servicio web REST requiere autenticación, cree un objeto de conexión HTTP.

Configuración de mensaje

El servicio de integración de datos genera los mensajes de solicitud e interpreta los mensajes de respuesta en función de la entrada y la salida del método y de los puertos configurados en la transformación de consumidor de servicio web REST.

Los puertos de entrada representan diferentes partes del mensaje de solicitud. Puede añadir puertos de entrada que identifiquen el recurso que desea recuperar o cambiar. También puede añadir puertos de entrada que representen encabezados HTTP, información de cookies y elementos en el mensaje de solicitud.

Los puertos de salida representan los elementos en el mensaje de respuesta que desea enviar a las transformaciones de nivel inferior o al destino en una asignación. Puede añadir puertos de salida que representen encabezados HTTP, información de cookies, código de respuesta y elementos en el mensaje de respuesta.

Identificación de recurso

Para identificar el recurso en una solicitud HTTP, el servicio de integración de datos añade los valores en puertos de entrada específicos a la URL de base. Define la URL de base en la conexión HTTP o en las propiedades de la transformación. Utilice la URL o puertos de argumento para identificar un recurso determinado.

Utilice puertos URL cuando el servicio web identifique un recurso a través de una cadena única de caracteres.

Por ejemplo, el servicio web REST de HypoStores identifica partes por el número de parte a través de la siguiente URL:

```
http://www.HypoStores.com/products/ProductDetails/<Part_No>
```

Para identificar una parte, defina los siguientes detalles de transformación:

1. Establezca la URL de base a la siguiente URL:

```
http://www.HypoStores.com/products/ProductDetails
```

2. Defina un puerto URL y pase el número de parte a la transformación a través del puerto URL.

Si la asignación pasa el número de parte 500 al puerto URL, el servicio de integración de datos emplea la siguiente URL en el mensaje de solicitud:

```
http://www.HypoStores.com/products/ProductDetails/500
```

Utilice puertos de argumento cuando el servicio web identifique la ubicación de un recurso a través de argumentos.

Por ejemplo, desea pasar un número de parte al servicio web REST de HypoStores a través del argumento "Part_No".

Para identificar una parte, defina los siguientes detalles de transformación:

1. Establezca la URL de base a la siguiente URL:

```
http://www.HypoStores.com/products/ProductDetails
```

2. Cree un puerto de argumento con el nombre de argumento "Part_No", y pase el número de parte a la transformación a través del puerto de argumento.

Si la asignación pasa el número de parte 600 al puerto de argumento, el servicio de integración de datos emplea la siguiente URL en el mensaje de solicitud:

```
http://www.HypoStores.com/products/ProductDetails?Part_No=600
```

Cree varios puertos de argumento para definir varios argumentos. El servicio de integración de datos separa cada argumento con un carácter de Y comercial (&).

Por ejemplo, desea recuperar los detalles de un empleado desde un servicio web REST y pasa el nombre y los apellidos del empleado a través de los argumentos "First_Name" y "Last_Name". Cree puertos de argumento con los nombres de argumento "First_Name" y "Last_Name". Si la asignación pasa el nombre "John Smith" a la transformación, el servicio de integración de datos utiliza una URL como la siguiente en el mensaje de solicitud:

```
http://www.HypoStores.com/employees/EmpDetails?First_Name=John&Last_Name=Smith
```

Si no especifica una URL ni un puerto de argumento, el servicio de integración de datos utiliza la URL de base de las propiedades de la transformación o la conexión HTTP para identificar el recurso. La URL de base de la conexión HTTP reemplaza a la URL de base de la transformación.

Métodos HTTP

Cuando crea una transformación de consumidor de servicio web REST, debe seleccionar el método HTTP que utiliza el servicio de integración de datos en el mensaje de solicitud. No se puede cambiar el método HTTP después de crear la transformación.

Configure la transformación para utilizar uno de los siguientes métodos HTTP:

Get

Recupera un recurso o un conjunto de recursos del servicio web. Por ejemplo, puede recuperar una tabla de productos o recuperar información sobre un producto.

Post

Envía datos a un servicio web. Utilice el método Post para crear un recurso o un conjunto de recursos. Por ejemplo, puede añadir los detalles de una nueva transacción de almacén.

Put

Reemplaza un recurso o un conjunto de recursos. Si los datos no existen, el método Put envía los datos. Por ejemplo, puede actualizar la dirección de envío de un cliente.

Delete

Elimina un recurso o un conjunto de recursos. Por ejemplo, puede eliminar el registro de un empleado que ya no trabaja para una organización.

Método HTTP Get

El servicio de integración de datos utiliza el método HTTP Get para recuperar datos desde un servicio web REST. Utilice el método Get para recuperar un recurso o un conjunto de recursos.

Cuando configura la transformación de consumidor de servicio web REST para utilizar el método Get, debe configurar los puertos de entrada, la salida del método y los puertos de salida. No se configura la entrada del método.

Ejemplo

Desea recuperar la descripción y el precio del número de parte 500 almacenados en la base de datos de productos de HypoStores. El servicio web utiliza la siguiente URL para identificar una parte:

```
http://www.HypoStores.com/products/ProductDetails?Part_No=<Part_No>
```

Introduzca la siguiente URL de base:

```
http://www.HypoStores.com/products/ProductDetails
```

La siguiente tabla muestra el puerto de entrada que puede definir:

Tipo de puerto	Nombre de argumento	Valor de entrada
Argumento	Part_No	500

La siguiente tabla muestra los puertos de salida que puede definir:

Tipo de puerto	Nombre de puerto	Valor de devolución
Salida	Part_Desc	...<desc>Bolígrafos de punta redonda ACME, paquete de 12, color negro, 0,7 mm</desc>...
Salida	Price_USD	...<price>9.89</price>...

Método HTTP Post

El servicio de integración de datos utiliza el método HTTP Post para enviar datos a un servicio web REST. El servicio web determina la función real que realiza el método Post. Puede usar el método Post para crear un recurso o un conjunto de recursos.

Cuando configure la transformación de consumidor de servicio web REST para utilizar el método Post, debe configurar los puertos de entrada, la entrada del método, la salida del método y los puertos de salida.

Ejemplo

Desea enviar la nueva parte 501 a la base de datos de productos de HypoStores. El servicio web utiliza la siguiente URL para el parte 501:

```
http://www.HypoStores.com/products/ProductDetails/501
```

Introduzca la siguiente URL de base:

```
http://www.HypoStores.com/products/ProductDetails
```

La siguiente tabla muestra los puertos de entrada que puede definir:

Tipo de puerto	Nombre de puerto	Valor de entrada
URL	URL_Part_No	501
Entrada	Part_Desc	Bolígrafos de punta redonda ACME, paquete de 12, color negro, 0,5 mm
Entrada	Price_USD	9.89

La siguiente tabla muestra los puertos de salida que puede definir:

Tipo de puerto	Nombre de puerto	Valor de devolución
Salida	Respuesta	<Respuesta devuelta por el servicio web>

Método HTTP Put

El servicio de integración de datos utiliza el método HTTP Put para actualizar los datos a través de un servicio web REST. Utilice el método Post para actualizar un recurso o un conjunto de recursos. Si los datos no existen, el servicio de integración de datos crea el recurso o el conjunto de recursos.

Cuando configure la transformación de consumidor de servicio web REST para utilizar el método Put, debe configurar los puertos de entrada, la entrada del método, la salida del método y los puertos de salida.

Ejemplo

Desea actualizar el precio unitario de la parte 501 en la base de datos de productos de HypoStores. El servicio web utiliza la siguiente URL para el parte 501:

`http://www.HypoStores.com/products/ProductDetails/501`

Introduzca la siguiente URL de base:

`http://www.HypoStores.com/products/ProductDetails`

La siguiente tabla muestra los puertos de entrada que puede definir:

Tipo de puerto	Nombre de puerto	Valor de entrada
URL	URL_Part_No	501
Entrada	Price_USD	9.99

La siguiente tabla muestra los puertos de salida que puede definir:

Tipo de puerto	Nombre de puerto	Valor de devolución
Salida	Respuesta	<Respuesta devuelta por el servicio web>

Método HTTP Delete

El servicio de integración de datos utiliza el método HTTP Delete para eliminar los datos a través de un servicio web REST. Utilice el método Delete para quitar un recurso o un conjunto de recursos.

Cuando configure la transformación de consumidor de servicio web REST para utilizar el método Delete, debe configurar los puertos de entrada, la entrada del método, la salida del método y los puertos de salida.

Ejemplo

Desea eliminar el número de parte 502 desde la base de datos de productos de HypoStores. El servicio web utiliza la siguiente URL para identificar una parte:

`http://www.HypoStores.com/products/ProductDetails?Part_No=<Part_No>`

Introduzca la siguiente URL de base:

`http://www.HypoStores.com/products/ProductDetails`

La siguiente tabla muestra el puerto de entrada que puede definir:

Tipo de puerto	Nombre de argumento	Valor de entrada
Argumento	Part_No	502

La siguiente tabla muestra los puertos de salida que puede definir:

Tipo de puerto	Nombre de puerto	Valor de devolución
Salida	Respuesta	<Respuesta devuelta por el servicio web>

Puertos de la transformación de consumidor de servicio web REST

Una transformación de consumidor de servicio web REST puede tener varios puertos de entrada y de salida. Puede crear puertos en grupos en función de la estructura de la jerarquía XML o JSON.

Cuando visualice los puertos de transformación, muestre los puertos si no necesita ver la jerarquía XML o JSON. Cuando se muestran los puertos, puede definir grupos y puertos, y asignar elementos desde la entrada y la salida del método a los puertos de entrada y salida.

Una transformación de consumidor de servicio web REST puede tener varios grupos de entrada y de salida. Cuando cree puertos y grupos y añada los puertos a los grupos. Defina los puertos en una jerarquía de grupo en función de la estructura de la jerarquía de entrada o salida en XML o JSON. Añada una clave para relacionar un grupo secundario con un grupo principal.

Todos los grupos, excepto el grupo situado en el nivel más bajo de la jerarquía, deben tener claves principales. Todos los grupos de la jerarquía, excepto el grupo raíz, deben tener claves externas.

La transformación posee un grupo de entrada raíz denominado RequestInput. Debe añadir una clave principal al grupo de entrada raíz. La clave debe contener valores string, bigint o integer. Puede configurar cualquier puerto en el grupo de entrada raíz como un puerto de transferencia.

Para asignar un elemento a un puerto, haga clic en el campo en la columna **Ubicación** y expanda la jerarquía en el cuadro de diálogo **Seleccionar ubicación**. A continuación, elija un elemento de la jerarquía.

Puertos de entrada

Los puertos de entrada representan datos desde una transformación de nivel superior o de un origen que desea pasar al servicio web. Puede configurar varios puertos de entrada. Cada puerto de entrada se asigna a un elemento en el mensaje de solicitud.

Para añadir un puerto de entrada, seleccione un grupo de entrada, haga clic en la flecha junto al botón **Nuevo** y seleccione **Campo**.

Puertos de salida

Los puertos de salida representan los elementos en el mensaje de respuesta que desea pasar a una transformación de nivel inferior o al destino. Puede configurar varios puertos de salida. Cada puerto de salida se asigna a un elemento en el mensaje de respuesta.

Para añadir un puerto de salida, seleccione un grupo de salida, haga clic en la flecha junto al botón **Nuevo** y seleccione **Campo**.

Puertos de transferencia

Los puertos de transferencia pasan datos a través de la transformación sin modificar los datos. Puede configurar cualquier puerto en el grupo de entrada raíz como un puerto de transferencia.

Para añadir un puerto de transferencia, añada un puerto para el grupo de entrada raíz. A continuación, haga clic con el botón derecho en el puerto y seleccione **Asignar**.

Puertos de argumento

Los puertos de argumento le permiten identificar un recurso cuando la URL para el recurso utiliza un argumento. Añada puertos de argumento al grupo de entrada raíz.

Un puerto de argumento tiene un nombre de puerto y un nombre de argumento. Si un nombre de argumento contiene un carácter que no está permitido en un nombre de puerto, introduzca un nombre de argumento que sea diferente del nombre de puerto. Por ejemplo, desea pasar el argumento "Cust-ID" para el servicio web, pero el servicio de integración de datos no permite el carácter de guión (-) en los nombres de puerto. Especifique "Cust-ID" como nombre de argumento y "CustID" como nombre del puerto.

El servicio de integración de datos añade los nombres de argumento y los valores para cada puerto de argumento a la URL de base como pares nombre = valor. Puede configurar varios puertos de argumento. El servicio de integración de datos separa los distintos argumentos de la solicitud con un carácter de Y comercial (&).

Por ejemplo:

```
http://www.HypoStores.com/customers/CustDetails?Last_Name=Jones&First_Name=Mary
```

Si define los puertos de argumento y los puertos URL en la transformación, el servicio de integración de datos añade los valores del puerto URL a la URL de base, seguidos por los nombres de argumento y los valores.

Para añadir un puerto de argumento, haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Puertos de argumento**. Introduzca el nombre de argumento y el nombre del puerto.

Puertos URL

Los puertos URL le permiten identificar un recurso mediante una URL estática. Para identificar un recurso, el servicio de integración de datos añade el valor del puerto URL a la URL de base.

Por ejemplo:

```
http://www.HypoStores.com/products/ProductDetails/<URL_port_value>
```

Añada puertos URL al grupo de entrada raíz.

Puede configurar varios puertos URL. El servicio de integración de datos separa los valores de cada puerto URL con un carácter de barra (/). Si define los puertos URL y los puertos de argumento en la transformación, el servicio de integración de datos añade los valores del puerto URL a la URL de base, seguidos por los nombres de argumento y los valores.

Para añadir un puerto URL, haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Puertos URL**.

Puertos de encabezado HTTP

Los puertos de encabezado HTTP representan encabezados HTTP en el mensaje de solicitud. Puede configurar varios puertos de encabezado HTTP.

Para pasar la información de encabezado al servicio web en la solicitud, añada el puerto al grupo de entrada raíz. Puede configurar un puerto de encabezado HTTP para el grupo de entrada raíz. Si añade un encabezado HTTP al grupo de entrada raíz, puede configurarlo como un puerto de transferencia.

Un puerto de encabezado HTTP tiene un nombre de puerto y un nombre de encabezado HTTP. Si un nombre de encabezado HTTP contiene un carácter que no está permitido en un nombre de puerto, introduzca un nombre de encabezado HTTP que sea diferente del nombre de puerto. Por ejemplo, desea pasar el nombre de encabezado "Content-Type" al servicio web, pero el servicio de integración de datos no permite el carácter

de guión (-) en los nombres de puerto. Especifique "Content-Type" como nombre del encabezado HTTP y "ContentType" como nombre del puerto.

Para añadir un puerto de encabezado HTTP, haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Encabezado HTTP**. Especifique un nombre de encabezado y un nombre de puerto.

Puertos de cookies

Puede configurar la transformación de consumidor de servicio web REST para que utilice autenticación por cookies. El servidor remoto del servicio web hace un seguimiento de los usuarios del consumidor de servicio web mediante cookies. Es posible aumentar el rendimiento cuando una asignación llama a un servicio web varias veces.

Para pasar la información de cookies al servicio web en la solicitud, añada el puerto al grupo de entrada raíz. Puede configurar un puerto de cookies para el grupo de entrada raíz. Si añade un puerto de cookies al grupo de entrada raíz, puede configurarlo como un puerto de transferencia.

Para extraer la información de cookies desde la respuesta, añada un puerto de cookies a un grupo de salida. Puede configurar un puerto de cookies para cada grupo de salida.

Cuando se proyecta el puerto de cookies a un mensaje de solicitud de servicio web, el proveedor del servicio web devuelve un valor de cookie en el mensaje de respuesta. Puede pasar el valor de cookie a otra transformación situada en un nivel inferior en la asignación o puede guardar el valor de la cookie en un archivo. Cuando se guarda el valor de la cookie en un archivo, puede configurar la cookie como entrada de la transformación de consumidor de servicio web REST.

Para añadir un puerto de cookies, haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Otros puertos**. A continuación, seleccione **Cookie** y haga clic en **Aceptar**.

Puertos de salida XML

Los puertos de salida XML representan las respuestas del servicio web. Los puertos de salida XML son puertos de cadena.

Añada un puerto de salida XML a un grupo de salida. Puede configurar un puerto de salida XML para cada grupo de salida.

Haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Otros puertos**. A continuación, seleccione **XML de salida** y haga clic en **Aceptar**.

Puertos de código de respuesta

Los puertos de código de respuesta representan los códigos de respuesta HTTP del servicio web. Los puertos de código de respuesta son puertos de entero.

Añada un puerto de código de respuesta a un grupo de salida. Puede configurar un puerto de código de respuesta para cada grupo de salida.

Para añadir un puerto de código de respuesta, seleccione un grupo de salida, haga clic con el botón derecho en el grupo de entrada raíz y seleccione **Nuevo > Otros puertos**. A continuación, seleccione **Código de respuesta** y haga clic en **Aceptar**.

Asignación de entrada de la transformación de consumidor de servicio web REST

Cuando visualice los puertos de transformación, muestre la asignación de entrada para ver la jerarquía de entrada del método. Cuando se muestra la asignación de entrada, puede definir grupos y puertos de entrada, y asignar puertos de entrada a elementos de la entrada del método.

La asignación de entrada incluye las siguientes áreas:

Puertos

Cree los grupos de entrada de transformación y los puertos de entrada en el área **Puertos**.

Entrada de método

El área **Entrada de método** muestra los elementos en el mensaje de solicitud que la transformación de consumidor de servicio web REST envía al servicio web. Si utiliza un objeto de esquema para crear la transformación, el objeto de esquema define la jerarquía de entrada de método.

Después de crear puertos de entrada, asigne los puertos de entrada desde el área **Puertos** a los elementos en el área **Entrada de método**. Cuando asigne un puerto de entrada a un elemento en la entrada de método, la ubicación del puerto aparece en la columna Ubicación en el área **Entrada de método**.

La herramienta Developer asigna elementos del primer nivel de la entrada de método a puertos de entrada cuando se opta por asignar el primer nivel de la jerarquía de entrada. Asimismo, la herramienta Developer crea los puertos que necesita para realizar la asignación. Si el primer nivel de la jerarquía contiene un elemento principal de ocurrencia múltiple con uno o varios elementos secundarios de ocurrencia múltiple, la herramienta Developer no asigna el primer nivel de la jerarquía.

Puede optar por ver las líneas que conectan los puertos de entrada a los elementos de la entrada de método.

Reglas y directrices para asignar puertos de entrada a elementos

Revise las siguientes reglas cuando asigne puertos de entrada a elementos en la jerarquía de entrada de método:

- Puede asignar un puerto de entrada a un elemento de la jerarquía. Puede asignar el mismo puerto a un número indeterminado de claves de la jerarquía.
- El puerto de entrada y el elemento deben tener tipos de datos compatibles.
- Puede asignar puertos de un grupo de entrada a varios niveles de jerarquía en la entrada de método.
- Debe asignar puertos de entrada a las claves en la entrada de método. Cualquier puerto que se asigne a una clave debe ser del tipo de datos string, integer o bigint. Asigne los datos a las claves en todos los niveles de la entrada de método situados encima del nivel de jerarquía que esté incluyendo en el mensaje de solicitud. Incluya las claves externas para todos los niveles situados por encima e incluyendo el nivel que esté asignando.

Nota: No es necesario que asigne puertos de entrada a claves si únicamente está asignando el nivel más bajo de la jerarquía de entrada de método.

- Debe asignar el elemento raíz RequestInput al elemento secundario del grupo Rest_Consumer_input para la definición de entrada de método.
- Puede asignar varios puertos de entrada con valores string, integer o bigint a una clave en el área **Entrada de método** para crear una clave compuesta. Cuando se hace clic en el campo **Ubicación** para una clave compuesta, puede volver a ordenar los puertos de entrada o eliminar uno de los puertos.

- Si el servicio web genera un documento JSON, asegúrese de que xmlRoot es el primer nodo en la jerarquía de respuesta. Si xmlRoot no es el primer nodo para un servicio web con respuesta JSON, pueden aparecer valores nulos.

Revise la siguiente regla cuando asigne los puertos de entrada a elementos con la entrada de método como JSON:

- Asegúrese de que los datos que pasa al puerto de entrada no contiene ceros iniciales seguidos de un número. Si los datos contienen ceros iniciales, se truncan en los valores correspondientes de la solicitud de JSON generada, y el tipo de dato del valor cambia de tipo cadena a tipo numérico.

Asignación de puertos de entrada a la entrada del método

Cuando se muestra la asignación de entrada de transformación, puede definir grupos y puertos de entrada y asignar puertos de entrada a elementos de entrada de operación.

1. Abra una transformación de consumidor de servicio web REST.
2. En la vista **Puertos**, abra la asignación de entrada.
3. Defina una clave principal para el grupo de entrada raíz.
4. Para añadir un grupo o un puerto de entrada al área **Puertos**, utilice uno de los siguientes métodos:

Método	Descripción
Arrastre un elemento.	Arrastre un grupo o un elemento secundario desde el área Entrada de método a una columna vacía en el área Puertos . Si arrastra un grupo al área Puertos , la herramienta Developer añade un grupo sin puertos.
Añada manualmente un grupo o puerto.	Para añadir un grupo, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Grupo . Para añadir un puerto, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Campo .
Arrastre un puerto desde otra transformación.	En el editor, arrastre un puerto desde otra transformación a la transformación de consumidor de servicio web REST.
Copie un puerto.	Seleccione puertos de otra transformación y cópielos al área Puertos . Para copiar puertos, puede utilizar métodos abreviados de teclado o puede utilizar los botones Copiar y Pegar de la herramienta Developer.
Seleccione Asignar primer nivel de la jerarquía .	La herramienta Developer asigna elementos que se encuentran en el primer nivel de la entrada de método a puertos y grupos de entrada. La herramienta Developer también crea los puertos y grupos de entrada para realizar la asignación.

5. Si crea manualmente un puerto o lo copia desde otra transformación, haga clic en la columna **Ubicación** del área **Entrada de método** y elija un puerto de la lista.
6. Para asignar puertos de entrada como una clave compuesta, utilice uno de los siguientes métodos:

Método	Descripción
Arrastre puertos de entrada.	Seleccione dos o más puertos de entrada y arrástrelos hasta una clave en la jerarquía de entrada de método.
Seleccione puertos de entrada desde el cuadro de diálogo Seleccionar ubicación .	Haga clic en la columna Ubicación de una clave en la jerarquía de entrada de método y luego seleccione los puertos de entrada.

7. Para borrar las ubicaciones de los elementos, utilice uno de los siguientes métodos:

Método	Descripción
Haga clic en Borrar .	Seleccione uno o más elementos en el área Entrada de método y haga clic en Borrar .
Elimine las líneas que conectan los puertos con los elementos.	Seleccione una o más líneas que conectan los puertos de entrada con los elementos en la entrada de método y pulse Eliminar .

Asignación de salida de la transformación de consumidor de servicio web REST

Cuando visualice los puertos de transformación, muestre la asignación de salida para ver la jerarquía de salida de método. Cuando se muestra la asignación de salida, puede definir grupos y puertos de salida y asignar elementos de salida de método a puertos de salida.

La asignación de salida incluye las siguientes áreas:

Salida de método

El área **Salida de método** muestra los elementos en el mensaje de respuesta que el servicio web devuelve a la transformación de consumidor de servicio web REST. Si utiliza un objeto de esquema para crear la transformación, el objeto de esquema define la jerarquía de salida de método.

Puertos

Cree los puertos y los grupos de salida de transformación en el área **Puertos**.

Después de crear los puertos de salida, asigne los elementos desde el área **Salida de método** a los puertos del área **Puertos**. Cuando se asigna un elemento desde la salida de método a un puerto de salida, la ubicación del elemento aparece en la columna **Ubicación** del área **Puertos**.

La herramienta Developer asigna elementos del primer nivel de la salida de método a puertos de salida cuando se opta por asignar el primer nivel de la jerarquía de salida. Asimismo, la herramienta Developer crea los puertos que necesita para realizar la asignación. Si el primer nivel de la jerarquía contiene un elemento principal de ocurrencia múltiple con uno o varios elementos secundarios de ocurrencia múltiple, la herramienta Developer no asigna el primer nivel de la jerarquía.

Puede elegir que se muestren los puertos de salida de una jerarquía. Cada grupo secundario aparece debajo del grupo principal. También puede optar por ver las líneas que conectan los elementos de la salida de método a los puertos de salida.

Si el objeto de esquema se elimina del repositorio, la herramienta Developer guarda la ubicación de los elementos del método en la asignación de salida. Cuando se muestra la asignación de salida, el área **Puertos** sigue mostrando la ubicación de los elementos del método en la columna **Ubicación** para los puertos de salida. Si asocia otro esquema con la transformación, la herramienta Developer comprueba si todas las ubicaciones son válidas. La herramienta Developer borra la ubicación de los elementos del método en el área **Puertos** de la asignación de salida si la ubicación ya no es válida.

Reglas y directrices para asignar elementos a puertos de salida

Revise las siguientes reglas cuando asigne elementos de la jerarquía de salida de método a puertos de salida:

- El elemento de salida de método y el puerto de salida deben tener tipos de datos compatibles.
- No se puede asignar un elemento a más de un puerto de salida en un grupo.
- Cada puerto de salida debe tener una ubicación válida, a excepción de que el puerto sea un puerto de transferencia.
- Si arrastra un elemento secundario de ocurrencia múltiple a un puerto de salida vacío, debe relacionar el grupo con otros grupos de salida. Cuando se selecciona un grupo, Developer tool crea claves para relacionar los grupos.
- Cuando se arrastra un elemento de ocurrencia múltiple a un grupo que contiene el elemento principal, puede configurar el número de ocurrencias de elemento secundario que se van a incluir. También puede reemplazar el grupo principal con el grupo secundario de ocurrencia múltiple en la salida de transformación.
- Si el servicio web genera un documento JSON, asegúrese de que xmlRoot es el primer nodo en la jerarquía de respuesta. Si xmlRoot no es el primer nodo para un servicio web con respuesta JSON, pueden aparecer valores nulos en los puertos de salida.

Personalizar opciones de vista

Puede cambiar la jerarquía de salida de método para mostrar los puertos de cookies, los puertos de transferencia y las claves en el área **Salida de método**. También puede mostrar construcciones de agrupación que definen cómo ordenar elementos.

Haga clic en **Personalizar vista** en el área **Salida de método**. Habilite una de las siguientes opciones:

Secuencia, Selección y Todo

Muestra una línea que indica si una definición de elemento es secuencia, selección o todo.

Los elementos en un grupo de secuencia deben estar en el orden especificado en la jerarquía.

En el mensaje de respuesta debe aparecer, como mínimo, un elemento en un grupo selección.

Los elementos en un grupo todo deben incluirse en el mensaje de respuesta.

Claves

Ver las claves en el área **Salida de método**. El área **Salida de método** incluye claves para cada grupo.

Puede añadir una clave a un puerto de salida en el área **Puertos**.

Puertos de transferencia

El área **Salida de método** muestra los puertos de transferencia. Los puertos de transferencia son puertos que transfieren datos a través de la transformación sin modificar los datos. Puede proyectar puertos de transferencia desde la salida de método a cualquiera de los grupos de salida de la transformación de consumidor de servicio web REST. Un puerto de transferencia recibe los datos sólo una vez, por lo que el puerto se encuentra en el nivel raíz de los mensajes de respuesta.

Asignación de la salida del método a puertos de salida

Cuando se muestra la asignación de salida de transformación, puede definir grupos y puertos de salida y asignar elementos de salida de método a puertos de salida.

1. Abra una transformación de consumidor de servicio web REST.

- En la vista **Puertos**, abra la asignación de salida.
- Para añadir un grupo o un puerto de salida al área **Puertos**, utilice uno de los siguientes métodos:

Método	Descripción
Arrastre un elemento.	Arrastre un grupo o un elemento secundario del área Salida de método a una columna vacía del área Puertos . Si arrastra un grupo al área Puertos , la herramienta Developer añade un grupo sin puertos.
Añada manualmente un grupo o puerto.	Para añadir un grupo, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Grupo . Para añadir un puerto, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Campo .
Arrastre un puerto desde otra transformación.	En el editor, arrastre un puerto desde otra transformación a la transformación de consumidor de servicio web REST.
Copie un puerto.	Seleccione puertos de otra transformación y cópielos al área Puertos . Para copiar puertos, puede utilizar métodos abreviados de teclado o puede utilizar los botones Copiar y Pegar de la herramienta Developer.

- Si crea manualmente un puerto o lo copia desde otra transformación, haga clic en la columna **Ubicación** del área **Puertos** y elija un elemento de la lista.
- Para borrar las ubicaciones de los puertos, utilice uno de los siguientes métodos:

Método	Descripción
Haga clic en Borrar .	Seleccione uno o más puertos en el área Puertos y haga clic en Borrar .
Elimine las líneas que conectan los elementos con los puertos.	Seleccione una o más líneas que conectan los elementos en la salida de método con los puertos de salida y pulse Eliminar .

Propiedades avanzadas de la transformación de consumidor de servicio web REST

Configure las propiedades que ayudan a determinar cómo el servicio de integración de datos procesa los datos para la transformación de consumidor de servicio web REST.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Conexión

Identifica el objeto de conexión HTTP para conectarse al servicio web. Cree y edite la conexión HTTP en la herramienta Developer. Cuando configure una conexión HTTP, configure la URL de base, el tipo de seguridad que el servicio web requiere y un tiempo de espera para la conexión.

La transformación de consumidor de servicio web REST se conecta a un servicio web mediante una URL. Puede definir la URL en las propiedades de la transformación o en la conexión HTTP.

Configure una conexión HTTP en las siguientes circunstancias:

- No se utiliza un puerto de entrada de URL.
- El servicio web requiere autenticación HTTP o certificados SSL.
- Se desea cambiar el tiempo de espera predeterminado de la conexión.

Validación de esquema XML

Valida el mensaje de respuesta en tiempo de ejecución. **Seleccione Error en XML no válido o Sin validación.**

Entrada ordenada

Habilita el servicio de integración de datos para generar una salida sin procesar todos los datos de entrada. Habilite la entrada ordenada cuando los datos de entrada se ordenen según las claves en la jerarquía de entrada XML.

URL

La URL de base para el servicio web REST. La URL de base en la conexión HTTP reemplaza este valor.

Formatear

El formato de la respuesta de servicio web. Seleccione **XML** o **JSON** en función de la respuesta de servicio web.

Creación de la transformación de consumidor de servicio web REST

Puede crear una transformación de consumidor de servicio web REST reutilizable o no reutilizable. Las transformaciones reutilizables pueden existir en varias asignaciones. Las transformaciones no reutilizables existen en una única asignación.

Cuando se crea una transformación de consumidor de servicio web REST, puede definir los elementos y la jerarquía XML manualmente, o puede importar los elementos y la jerarquía desde un objeto de esquema. El objeto de esquema puede ser un archivo XML o un archivo de texto.

Crear una transformación de consumidor de servicio web REST

Cuando cree una transformación de consumidor de servicio web REST, seleccione un método y defina la entrada del método y la salida del método según el método que elija.

1. Para crear una transformación de consumidor de servicio web REST, utilice uno de los siguientes métodos:

Método	Descripción
Reutilizable	Seleccione un proyecto o una carpeta en la vista Explorador de objetos. Haga clic en Archivo > Nuevo > Transformación . Seleccione la transformación de consumidor de servicio web REST y haga clic en Siguiente .
No reutilizable	En una asignación o un mapplet, arrastre una transformación de consumidor de servicio web REST desde la paleta de transformación hasta el editor de asignaciones o de mapplets.

2. Introduzca el nombre de la transformación y seleccione la ubicación y el método HTTP.
3. Haga clic en **Siguiente**.
4. Para definir el método de entrada, utilice uno de los siguientes métodos:

Método	Descripción
Crear como vacío	Los elementos XML y la jerarquía se definen manualmente.
Crear desde un elemento en un objeto de esquema	Los elementos XML y la jerarquía se importan desde un objeto de esquema.

El área **Definición de entrada de método** muestra los grupos de entrada y los puertos de entrada de la transformación. El área **Asignación de entrada** muestra la jerarquía de mensajes de solicitud.

5. Defina los grupos y los puertos de entrada y asigne los puertos de entrada a los elementos de entrada.
6. Haga clic en **Siguiente**.
7. Para definir la salida del método, seleccione **Crear como vacío** o **Crear desde un elemento en un objeto de esquema**.

El área **Definición de salida de método** muestra los grupos de salida y los puertos de salida de la transformación. El área **Asignación de salida** muestra la jerarquía de mensajes de solicitud.

8. Defina los grupos y los puertos de salida y asigne los elementos a los puertos de salida.
9. Haga clic en **Finalizar**.

Análisis de un mensaje de respuesta JSON que contiene matrices

Cuando el elemento es un elemento secundario de tipo complejo y la ocurrencia máxima de este elemento no está limitada, el esquema no es válido. El analizador JSON restringe la extracción de varias instancias de un elemento.

La ocurrencia máxima de elementos secundarios en el tipo complejo debe ser 0 o 1 con el indicador de orden como opción para el tipo complejo en un esquema. Cuando se cambia la ocurrencia máxima como 1 para validar el esquema, se puede extraer una instancia del elemento a la vez.

Puede utilizar la ocurrencia máxima como no limitada en el indicador de orden de selección de un tipo complejo en el esquema.

Ejemplo de mensaje de respuesta JSON

Tiene el siguiente esquema donde el elemento de tipo complejo `xmlRoot` tiene el nombre de elemento `Likes` cuya ocurrencia máxima es ilimitada:

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="xmlRoot">
    <xs:complexType>
      <xs:all>
        <xs:element type="xs:byte" name="Age"/>
        <xs:element type="xs:string" name="FirstName"/>
        <xs:element type="xs:string" name="Likes" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:string" name="FamilyName"/>
      </xs:all>
    </xs:complexType>
  </xs:element>
</xs:schema>
```



```

    </xs:all>
  </xs:complexType>
</xs:element>
</xs:schema>

```

Puede cambiar la respuesta JSON con el siguiente formato:

```

<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="xmlRoot">
    <xs:complexType>
      <xs:choice maxOccurs="unbounded">
        <xs:element type="xs:byte" name="Age"/>
        <xs:element type="xs:string" name="FirstName"/>
        <xs:element type="xs:string" name="Likes" />
        <xs:element type="xs:string" name="FamilyName"/>
      </xs:choice>
    </xs:complexType>
  </xs:element>
</xs:schema>

```

`<xs:choice maxOccurs="unbounded">` permite que el contenido se repita una o más veces, en cualquier orden.

Matrices sin nombre en un mensaje de respuesta

Una transformación de consumidor de servicio web REST admite matrices sin nombre solo en un mensaje de respuesta, pero no en un mensaje de solicitud. Para analizar un esquema de matriz sin nombre establecido en la definición de salida del método, el elemento primario de los elementos de matriz de tipo `complexType` o simple debe ser de `xmlRoot` de nombre.

En una transformación de consumidor de servicio web REST, debe definir `xmlRoot` como elemento secundario del elemento `xmlRoot` con el valor máximo de ocurrencias establecido en sin enlazar y los elementos de la matriz sin nombre como elementos secundarios del elemento `xmlRoot`.

La siguiente imagen muestra la salida del método definida para la matriz sin nombre:

Ports Method input Method output

Show: Method output definition Output mapping

Method output definition

Name	Type	Min...	Ma...	Description	>>
Rest_Consume...	(Rest_Cons...				
xmlRoot	(xmlRoot)	1	1		
xmlRoot	(xmlRoot)	1	Un...		
emp	xs:string	1	1		
empid	xs:string	1	1		

CAPÍTULO 40

Transformación de enrutador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de enrutador, 618](#)
- [Transformaciones de enrutador en asignaciones dinámicas, 619](#)
- [Cómo trabajar con grupos, 620](#)
- [Cómo trabajar con puertos, 624](#)
- [Cómo conectar transformaciones de enrutador de una asignación, 624](#)
- [Propiedades avanzadas de la transformación de enrutador, 625](#)
- [Transformación de enrutador en un entorno no nativo, 625](#)

Resumen de la transformación de enrutador

La transformación de enrutador es una transformación activa que enruta los datos en varios grupos de salida en función de una o más condiciones. Enrute los grupos de salida a transformaciones distintas o a destinos distintos en la asignación.

Una transformación de enrutador es similar a una de filtro, ya que ambas utilizan una condición para comprobar datos. Una transformación de filtro comprueba los datos para aplicar una condición y anula las filas de datos que no cumplen dicha condición. Una transformación de enrutador comprueba los datos para aplicar una o varias condiciones y puede enrutar las filas de datos que no cumplen ninguna de las condiciones en un grupo de salida predeterminado.

Si necesita comprobar los mismos datos de entrada en función de varias condiciones, utilice una transformación de enrutador en una asignación en lugar de crear varias transformaciones de filtro para llevar a cabo la misma tarea. La transformación de enrutador resulta más eficaz. Por ejemplo, para comprobar los datos basándose en tres condiciones, puede usar una transformación de enrutador en lugar de tres transformaciones de filtro. Cuando se utiliza una transformación de enrutador en una asignación, el Servicio de integración de datos procesa una vez los datos de entrada. Si utiliza varias transformaciones de filtro en una asignación, el Servicio de integración de datos procesa los datos de entrada para cada transformación.

Una transformación de enrutador está formada por grupos de entrada y salida, puertos de entrada y salida, condiciones del filtro de grupo y propiedades avanzadas que usted configura en Developer tool.

Cuando el motor de Spark ejecuta una asignación con una transformación de enrutador, dicho motor procesa la canalización de asignación ascendente una vez y prepara los datos en HDFS para que puedan ser utilizados por cada rama descendente.

La siguiente figura muestra un ejemplo de transformación de enrutador y sus componentes:

Nombre	Tipo	Longitud/...
Input (4)		
COUNTRY	string	10
CUSTOMER_NO	string	10
FIRSTNAME	string	10
LASTNAME	string	10
Default (4)		
COUNTRY	string	10
CUSTOMER_NO	string	10
FIRSTNAME	string	10
LASTNAME	string	10
France (4)		
COUNTRY	string	10
CUSTOMER_NO	string	10
FIRSTNAME	string	10
LASTNAME	string	10
Japan (4)		
COUNTRY	string	10
CUSTOMER_NO	string	10
FIRSTNAME	string	10
LASTNAME	string	10
USA (4)		
COUNTRY	string	10
CUSTOMER_NO	string	10
FIRSTNAME	string	10
LASTNAME	string	10

1. Grupo de entrada
2. Puertos de entrada
3. Grupo de salida predeterminado
4. Grupos de salida definidos por el usuario

Transformaciones de enrutador en asignaciones dinámicas

Puede utilizar una transformación de enrutador en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados en la condición de filtro del grupo.

Si utiliza un puerto dinámico en la condición de filtro del grupo, ese puerto podría contener más de un puerto generado. La condición de filtro de grupo se expande para incluir cada puerto generado. Cada puerto generado debe ser de un tipo válido en la expresión.

Puede parametrizar la condición de filtro de grupo. Utilice un parámetro de tipo de expresión para especificar el filtro.

Cómo trabajar con grupos

Una transformación de enrutador tiene los siguientes tipos de grupos:

- Entrada
- Salida

Grupo de entrada

La transformación de enrutador incluye un grupo de entrada simple. El grupo de entrada incluye todos los puertos de entrada que agregue a la transformación.

Grupos de salida

La transformación de rango incluye los siguientes tipos de grupos de salida:

Grupos definidos por el usuario

Cree un grupo definido por el usuario para probar una condición en función de los datos de entrada. Un grupo definido por el usuario está formado por puertos de salida y una condición del filtro de grupo. Puede crear y editar grupos definidos por el usuario en la ficha **Grupos** mediante Developer Tool. Cree un grupo definido por el usuario para cada condición que desee especificar.

El servicio de integración de datos utiliza la condición para evaluar cada una de las filas de los datos de entrada. Comprueba las condiciones de cada grupo definido por el usuario antes de procesar el grupo predeterminado. El servicio de integración de datos determina el orden de evaluación de cada condición en función del orden de los grupos de salida conectados. El servicio de integración de datos procesa los grupos definidos por el usuario que están conectados a una transformación o a un destino en una asignación.

Si una fila cumple más de una condición del filtro de grupo, el servicio de integración de datos pasa esta fila varias veces.

El grupo predeterminado

Tras la creación de un grupo definido por el usuario, Developer Tool crea el grupo predeterminado. Developer Tool no le permite editar ni eliminar el grupo predeterminado. Este grupo no tiene una condición del filtro de grupo asociada a él. Si todas las condiciones de grupo dan como resultado FALSE, el servicio de integración de datos pasa la fila al grupo predeterminado. Si desea que el servicio de integración de datos anule todas las filas del grupo predeterminado, no lo conecte a una transformación ni a un destino de una asignación.

Developer Tool eliminará el grupo predeterminado cuando elimine de la lista el último grupo definido por el usuario.

Developer Tool copia información de la propiedad desde los puertos de entrada del grupo de entrada para crear un grupo de puertos de salida para cada grupo de salida. No puede cambiar ni eliminar los puertos de salida ni sus propiedades.

Utilizar las condiciones del filtro de grupo

Puede realizar una prueba de los datos en función de una o varias condiciones del filtro de grupo. Puede crear condiciones del filtro de grupo en la ficha **Grupos** mediante el editor de expresiones.

Puede especificar cualquier expresión que devuelva un valor simple. A su vez, también puede introducir una constante para la condición. Una condición del filtro de grupo devuelve para cada fila que pase a través de la

transformación los valores TRUE o FALSE, dependiendo de si la fila cumple la condición que se ha especificado. Cero (0) es equivalente a FALSE. Todo valor distinto de cero es el equivalente de TRUE. Puede utilizar un solo puerto numérico como una condición de filtro. El servicio de integración de datos pasa las filas de datos que devuelven el resultado TRUE a cada transformación o destino que está asociado a cada grupo definido por el usuario.

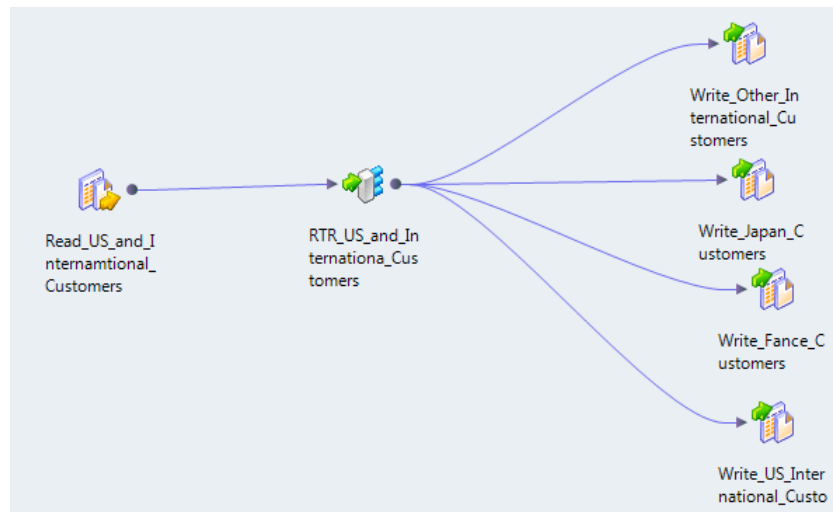
Nota: No se puede utilizar un solo puerto dinámico para devolver un valor booleano.

Por ejemplo, tiene cliente en nueve países distintos y desea realizar diferentes cálculos con los datos de tres países. Puede utilizar una transformación de enrutador de una asignación para filtrar dichos datos a tres transformaciones de expresión distintas.

Puede utilizar parámetros como elementos en la condición del filtro de grupo. Puede usar parámetros del sistema o parámetros definidos por el usuario. Puede crear parámetros en el editor de expresiones y añadirlos a la expresión.

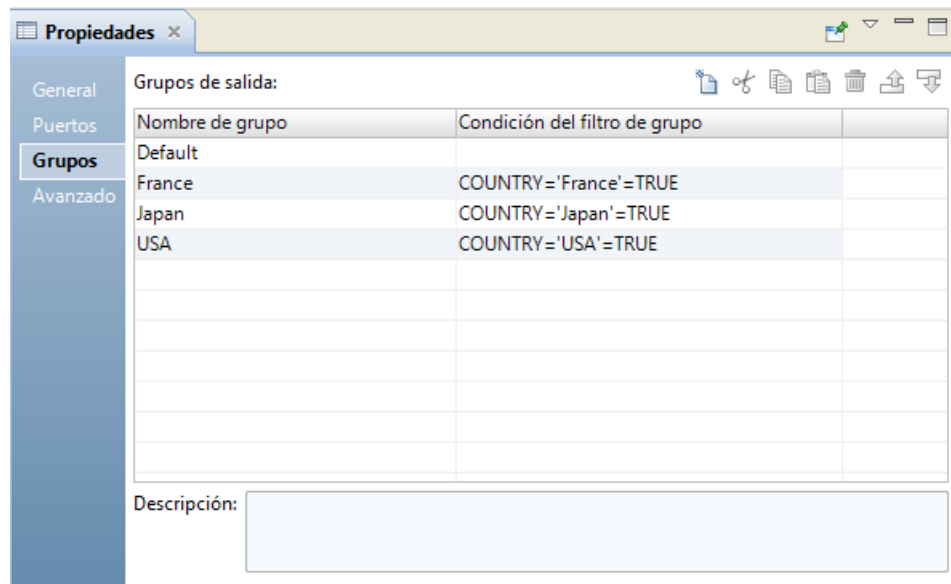
El grupo predeterminado no tiene una condición del filtro de grupo. Sin embargo, puede crear una transformación de expresión para realizar un cálculo en función de los datos de los otros seis países.

La siguiente figura muestra una asignación con una transformación de enrutador que filtra datos basándose en varias condiciones:



Para realizar varios cálculos con los datos de tres países diferentes, cree tres grupos definidos por el usuario y especifique tres condiciones del filtro de grupo en la ficha **Grupos**.

En la siguiente tabla se muestran las condiciones del filtro de grupo que filtra los datos del cliente:



En la siguiente tabla se muestran las condiciones del filtro de grupo que filtran los datos del cliente:

Nombre de grupo	Condición del filtro de grupo
Francia	customer_name='France'=TRUE
Japón	customer_name='Japan'=TRUE
EE. UU.	customer_name='USA'=TRUE

En la asignación, el servicio de integración de datos pasa las filas de datos que devuelven el resultado TRUE en cada transformación o destino asociado con cada grupo definido por el usuario, como por ejemplo Japón, Francia y EE. UU. El servicio de integración de datos pasa la fila al grupo predeterminado si el resultado de todas las condiciones es FALSE. A continuación, el servicio de integración de datos pasa los datos de los otros seis países a la transformación o al destino asociado al grupo predeterminado. Si desea que el servicio de integración de datos anule todas las filas del grupo predeterminado, no lo conecte a una transformación ni a un destino de una asignación.

La transformación de enrutador pasa los datos a través de cada grupo que cumpla la condición. Si los datos cumplen tres condiciones del grupo de salida, la transformación de enrutador pasa los datos a través de tres grupos de salida.

Por ejemplo, puede configurar las siguientes condiciones de grupo en una transformación de enrutador:

Nombre de grupo	Condición del filtro de grupo
Grupo de salida 1	employee_salary > 1000
Grupo de salida 2	employee_salary > 2000

Si la transformación de enrutador procesa datos de una fila de entrada con employee_salary=3000, enruta los datos a través de los grupos de salida 1 y 2.

Puertos dinámicos en condiciones de filtro de grupo

Puede utilizar un puerto dinámico en una condición del filtro de grupo. El Servicio de integración de datos aplica la condición de filtro a cada puerto generado en el puerto dinámico.

Por ejemplo, un puerto dinámico, MyDynamicPort, contiene tres puertos de decimales:

```
Salary  
Bonus  
Stock
```

Podría configurar la siguiente condición del filtro de grupo:

```
MyDynamicPort > 100
```

La condición del filtro de grupo se expande a la siguiente expresión:

```
Salary > 100 AND Bonus > 100 AND Stock > 100
```

Cada puerto generado debe ser de un tipo válido en la expresión.

Parametrizar el filtro de grupo

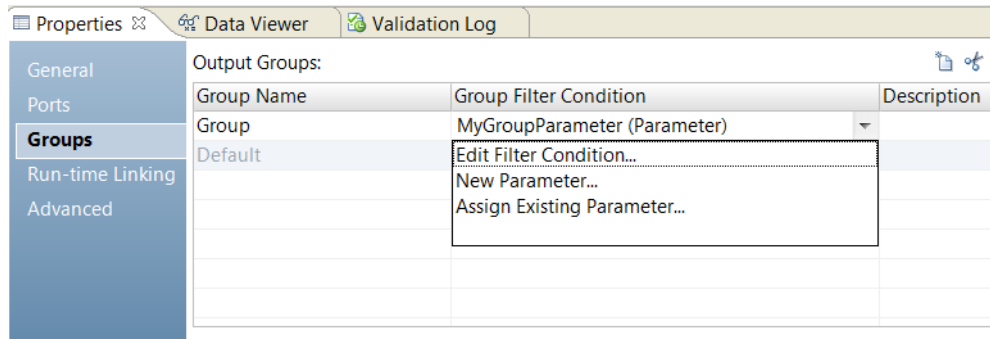
Puede utilizar un parámetro de expresión para definir un filtro de grupo. Un parámetro de expresión es un parámetro que contiene una expresión completa.

Por ejemplo, un parámetro de expresión puede contener la expresión

```
Employee_Salary > 1000.
```

Para parametrizar el filtro de grupo, seleccione **Nuevo parámetro** en el campo **Condición del filtro de grupo** de la ficha **Grupos**. Defina el parámetro y escriba la expresión en el editor de expresiones. Un parámetro de expresión no puede contener otros parámetros.

En la siguiente imagen se muestra la ficha **Grupos** en las propiedades de transformación:



Cuando se usa un parámetro de expresión, Developer tool no puede validar la expresión. Si la expresión no es válida en el tiempo de ejecución, podría producirse un error de asignación.

Cómo añadir grupos

Al añadir un grupo, Developer Tool copia la información de las propiedades de los puertos de entrada en los de salida.

1. Haga clic en la ficha **Grupos**.
2. Haga clic en el botón **Nuevo**.
3. Introduzca un nombre para el grupo en la sección **Nombre de grupo**.

4. Haga clic en el campo **Condición del filtro de grupo** para abrir el **Editor de expresiones**.
5. Introduzca la condición del filtro de grupo.
6. Haga clic en **Validar** para comprobar la sintaxis de la condición.
7. Haga clic en **Aceptar**.

Cómo trabajar con puertos

Una transformación de enrutador tiene puertos de entrada y de salida. Los puertos de entrada se encuentran en el grupo de entrada, mientras que los de salida están en los grupos de salida.

Puede crear puertos de entrada copiándolos de otra transformación o creándolos de forma manual en la ficha **Puertos**.

Developer Tool crea puertos de salida copiando las siguientes propiedades de los puertos de entrada:

- Nombre de puerto
- Tipo de datos
- Precisión
- Escala
- Valor predeterminado

Al realizar modificaciones en los puertos de entrada, Developer Tool actualiza los puertos de salida para reflejar los cambios. No puede editar ni eliminar los puertos de salida.

Developer Tool crea nombres de puertos de salida en función de los nombres de los puertos de entrada. Para cada puerto de entrada, Developer Tool crea un puerto de salida correspondiente en cada grupo de salida.

Cómo conectar transformaciones de enrutador de una asignación

Cuando conecte transformaciones a una transformación de enrutador en una asignación, tenga en cuenta las siguientes reglas:

- Puede conectar un grupo a una transformación o destino.
- Puede conectar un puerto de salida de un grupo a varias transformaciones o a varios destinos.
- Puede conectar varios puertos de salida de un grupo a varias transformaciones o a varios destinos.
- No puede conectar más de un grupo a un destino o a una transformación de grupo de entrada simple.
- No puede conectar más de un grupo a una transformación de varios grupos de entrada, excepto en el caso de transformaciones de incorporación, en las que puede conectar cada grupo de salida a un grupo de entrada diferente.

Propiedades avanzadas de la transformación de enrutador

Configure las propiedades para ayudar a determinar cómo el servicio de integración de datos procesa los datos para la transformación de enrutador.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de enrutador en un entorno no nativo

El procesamiento de transformaciones de enrutador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 41

Transformación de generador de secuencia

Este capítulo incluye los siguientes temas:

- [Introducción a la transformación de generador de secuencia, 626](#)
- [Puertos de generador de secuencia, 626](#)
- [Propiedades de transformación de generador de secuencia, 629](#)
- [Propiedades avanzadas del generador de secuencia, 631](#)
- [Objeto de datos de secuencia, 632](#)
- [Cómo crear una transformación de generador de secuencia, 635](#)
- [Preguntas frecuentes, 636](#)
- [Transformación de generador de secuencia en un entorno no nativo, 637](#)

Introducción a la transformación de generador de secuencia

La transformación de generador de secuencia es una transformación pasiva que genera valores numéricos. Utilice la transformación de generador de secuencia para crear valores de clave principal únicos, reemplazar claves principales que falten o realizar un ciclo a través de un intervalo secuencial de números.

El generador de secuencia contiene puertos de transferencia y un puerto de salida. Puede conectar el puerto NEXTVAL a los puertos de entrada de otras transformaciones. El servicio de integración incrementa la secuencia cuando se ejecuta la asignación.

Puede crear una transformación de generador de secuencia en función de una secuencia nueva o un objeto de datos de secuencia. Un objeto de datos de secuencia es un objeto que crea y mantiene una secuencia de valores.

Puertos de generador de secuencia

La transformación de generador de secuencia tiene puertos de transferencia y un puerto de salida, NEXTVAL. No se puede editar ni eliminar el puerto de salida.

Puertos de transferencia

Puede añadir un puerto a la transformación de generador de secuencia como un puerto de transferencia. Los puertos de transferencia son puertos de entrada y salida que reciben datos de entrada y devuelven los mismos datos a una asignación sin modificarlos.

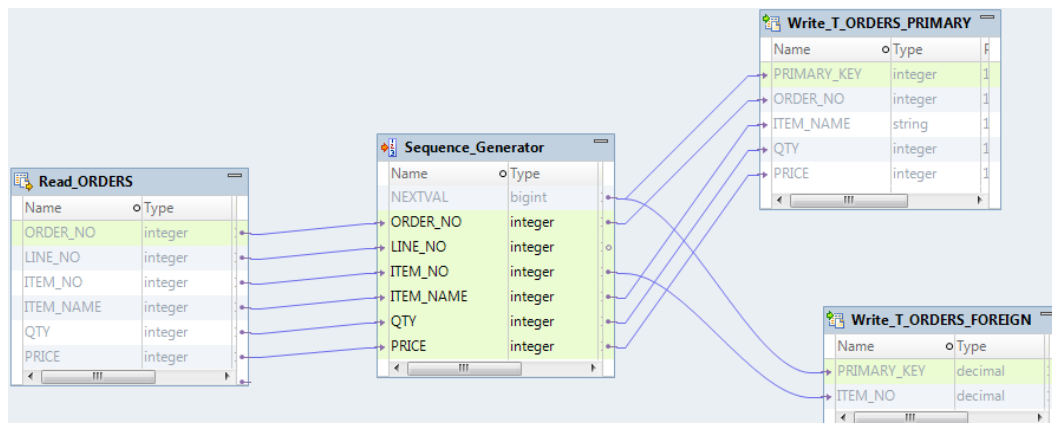
Deberá añadir al menos un puerto de entrada a la transformación y conectarla a un origen o una transformación de nivel superior antes de vincular el puerto de salida NEXTVAL a los destinos. Para añadir un puerto de transferencia a la transformación, arrastre un puerto desde un origen o una transformación de nivel superior en la asignación a la transformación de generador de secuencia.

Puerto NEXTVAL

Puede conectar NEXTVAL a una transformación para generar valores únicos para cada fila en la transformación. Conecte el puerto NEXTVAL a una transformación o destino de nivel inferior para generar una secuencia de números. Si conecta NEXTVAL a varias transformaciones, el Servicio de integración genera la misma secuencia de números para cada transformación.

Puede conectar el puerto NEXTVAL para generar la secuencia basándose en las propiedades Valor inicial y Valor de incremento. Si el generador de secuencia no está configurado para realizar un ciclo a través de la secuencia, el puerto NEXTVAL genera números de secuencia hasta el valor final configurado.

La siguiente imagen muestra una asignación con el puerto NEXTVAL de la transformación de generador de secuencia conectado a un origen y a dos destinos para generar valores de clave principal y externa.



Cuando configure la transformación de generador de secuencia con Valor inicial = 1 y Valor de incremento = 1, el Servicio de integración genera los mismos valores de clave principal para las tablas de destino T_ORDERS_PRIMARY y T_ORDERS_FOREIGN:

Crear claves

Puede crear valores de claves principales o externas con la transformación de generador de secuencia conectando el puerto NEXTVAL a un destino o transformación de nivel inferior. Puede utilizar un intervalo de valores de 1 a 9.223.372.036.854.775.807 con el intervalo más pequeño de 1. También puede crear claves compuestas para identificar cada fila de la tabla.

Para crear una clave compuesta, puede configurar el Servicio de integración para que efectúe un ciclo a través de un conjunto de valores más pequeño. Por ejemplo, si hay tres tiendas que generan números de pedido, configure una transformación de generador de secuencia para que efectúe un ciclo por los valores de 1 a 3, con un incremento de 1. Cuando conecte el puerto ORDER_NO a la transformación de generador de secuencia, los valores generados crearán claves compuestas únicas.

El siguiente ejemplo muestra claves compuestas y números de pedido:

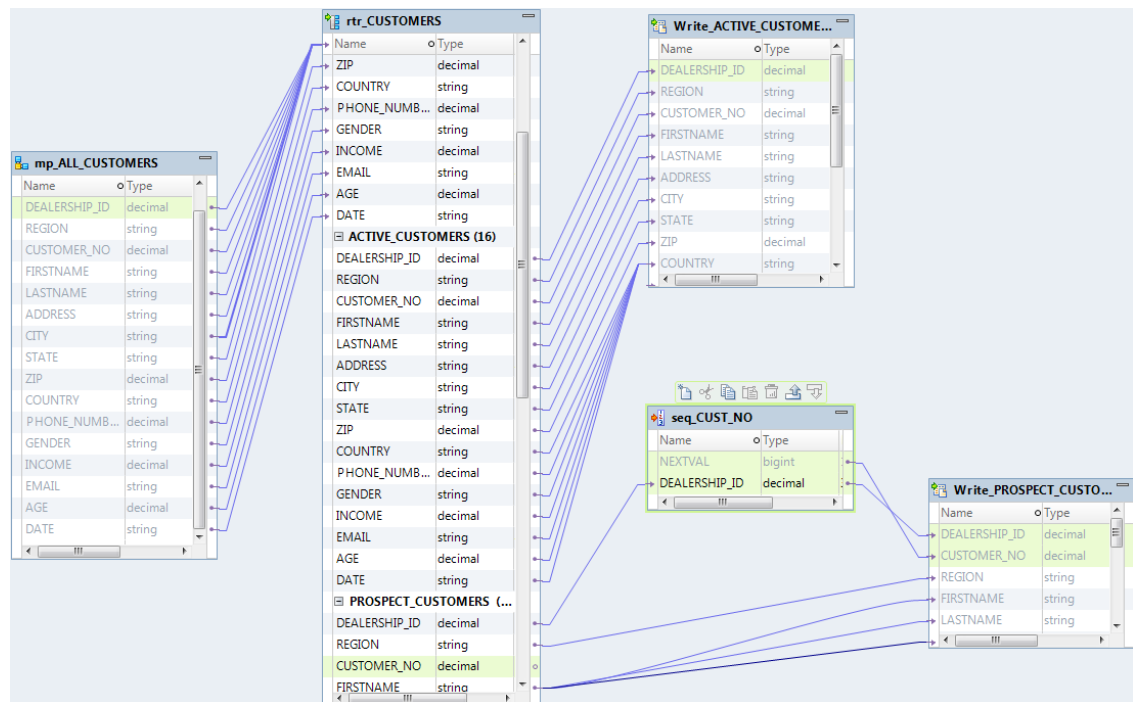
COMPOSITE_KEY	ORDER_NO
1	12345
2	12345
3	12345
1	12346
2	12346
3	12346

Reemplazar valores que faltan

Cuando se utiliza una transformación de generador de secuencia para reemplazar claves que faltan, también puede utilizar una transformación de enrutador para filtrar valores nulos de las columnas que tienen valores asignados. Puede conectar la transformación de enrutador a la transformación de generador de secuencia y utilizar NEXTVAL para generar una secuencia de valores numéricos para llenar los valores nulos.

Por ejemplo, para reemplazar valores nulos en una columna CUSTOMER_NO, cree una asignación con un origen que contenga datos de clientes. Añada una transformación de enrutador para filtrar los clientes con los números de cliente asignados a partir de los que tienen valores nulos. Añada una transformación de generador de secuencia para generar valores CUSTOMER_NO únicos. Añada los destinos de cliente donde escribir los datos.

La siguiente imagen muestra una asignación que reemplaza los valores nulos en la columna CUSTOMER_NO:



Propiedades de transformación de generador de secuencia

Configure las propiedades de transformación que utiliza el servicio de integración para generar valores secuenciales.

En la tabla siguiente se enumeran las propiedades que se configuran para un objeto de datos de secuencia y una secuencia nueva:

Propiedad	Descripción
Valor inicial	Valor inicial de la secuencia generada que desea que utilice el servicio de integración si se usa la opción de ciclo. Si selecciona Ciclo, el servicio de integración efectúa un ciclo que regresa a este valor cuando alcanza el valor final. El valor predeterminado es 0. El valor máximo es 9 223 372 036 854 775 806.
Valor final	Valor máximo generado por el servicio de integración. Si el servicio de integración alcanza este valor durante la sesión y la secuencia no está configurada para efectuar un ciclo, la sesión no se completará correctamente. El valor máximo es 9 223 372 036 854 775 807.
Valor de incremento	Diferencia entre dos valores consecutivos del puerto NEXTVAL. El valor predeterminado es 1. Debe ser un entero positivo. El valor máximo es 2 147 483 647.
Ciclo	Si está habilitada, el servicio de integración realiza un ciclo por todo el intervalo de la secuencia y se vuelve a iniciar con el valor inicial. Si está deshabilitada, el servicio de integración detiene la secuencia al llegar al valor final configurado. El servicio de integración interrumpe la sesión con errores de desbordamiento si se alcanza el valor final y aún tiene filas que procesar.

Valor inicial

Utilice Ciclo para generar una secuencia de repetición, tales como los números del 1 al 12, que corresponden a los meses de un año.

1. Introduzca el valor mínimo en la secuencia que desea que utilice el Servicio de integración para el valor inicial.
2. Introduzca el valor más alto que se utilizará para Valor final.
3. Seleccione Ciclo.

Mientras realiza el ciclo, el Servicio de integración alcanza el valor final configurado de la secuencia, lo ajusta e inicia el ciclo de nuevo, comenzando por el valor inicial configurado.

Valor final

El valor final es el valor máximo que desea que genere el Servicio de integración. Si el Servicio de integración alcanza el valor final y el generador de secuencia no está configurado para realizar un ciclo a través de la secuencia, la ejecución de la asignación genera un error de desbordamiento.

Establezca el valor final en cualquier número entero entre 1 y 9.233.372.036.854.775.807. Si conecta el puerto NEXTVAL a un puerto de enteros de nivel inferior, establezca el valor final en un valor que no sea

mayor que el valor máximo de entero. Por ejemplo, si conecta el puerto NEXTVAL a un puerto de enteros pequeños, establezca el valor final en un máximo de 32.767. Si NEXTVAL supera el valor máximo del tipo de datos para el puerto de nivel inferior, se produce un error de la asignación.

Valor de incremento

El servicio de integración genera una secuencia en el puerto NEXTVAL basada en las propiedades Valor actual e Incrementar por en la transformación de generador de secuencia.

La propiedad Valor actual es el valor con el que el servicio de integración inicia la creación de la secuencia para cada sesión: Incrementar por es el número entero que el servicio de integración añade al valor existente para crear el nuevo valor de la secuencia. De forma predeterminada, el Valor actual es 1 e Incrementar por es 1.

Por ejemplo, puede crear una transformación de generador de secuencia con un valor actual de 1.000 y un incremento de 10. Si pasa tres filas a través de la asignación, el servicio de integración genera el siguiente conjunto de valores:

```
1000  
1010  
1020
```

Realizar ciclos a través de un intervalo de valores

Puede establecer un intervalo de valores para la transformación de generador de secuencia. Si utiliza la opción de ciclo, la transformación de generador de secuencia repite el intervalo cuando llega al valor final.

Por ejemplo, si establece el intervalo de secuencia para que se inicie en 10 y finalice en 50, y establece un valor de incremento de 10, la transformación de generador de secuencia creará los valores 10, 20, 30, 40 y 50. La secuencia vuelve a comenzar en 10.

Propiedades avanzadas del generador de secuencia

Configure las propiedades avanzadas que el servicio de integración utiliza para generar valores secuenciales.

En la siguiente tabla se enumeran las propiedades avanzadas que se pueden configurar para un objeto de datos de secuencia y una secuencia nueva:

Propiedad	Descripción
Restablecer	<p>Si está habilitada, el servicio de integración restablece el objeto de datos de la secuencia al valor inicial cuando finaliza la ejecución de la asignación. Si está deshabilitada, el servicio de integración incrementa el valor actual después de finalizar la ejecución de la asignación y utiliza el valor en la siguiente ejecución de asignación.</p> <p>Esta propiedad está deshabilitada para las transformaciones de generador de secuencia reutilizables y para las transformaciones de generador de secuencia no reutilizables que utilizan un objeto de datos de secuencia reutilizable.</p>
Nivel de seguimiento	<p>Nivel de detalle de la transformación que el servicio de integración escribe en el registro de la asignación. Puede elegir Simplificado, Normal, Inicialización detallada o Datos detallados. El valor predeterminado es Normal.</p>
Mantener orden de filas	<p>Mantenga el orden de las filas de los datos de entrada en la transformación. Seleccione esta opción si el servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas. El valor predeterminado es false.</p>

Restablecer

Si configura una transformación de generador de secuencia no reutilizable para utilizar la propiedad Restablecer, el servicio de integración utilizará el valor inicial original para cada ejecución de asignación. De lo contrario, el Servicio de integración incrementa el valor actual y utiliza ese valor en la siguiente ejecución de asignación.

Por ejemplo, puede configurar una transformación de generador de secuencia para crear valores de 1 a 1.000 con un incremento de 1, un valor inicial de 1 y elegir Restablecer. Durante la primera ejecución de asignación, el Servicio de integración genera los números de 1 a 234. En cada ejecución de asignación posterior, el Servicio de integración vuelve a generar números que comienzan con el valor inicial de 1.

Si no restablece, el Servicio de integración actualiza el valor actual a 235 al final de la primera ejecución. La próxima vez que utiliza la transformación de generador de secuencia, el primer valor generado es 235.

Nota: La opción Restablecer está deshabilitada para las transformaciones de generador de secuencia reutilizable.

Mantener orden de filas

Mantenga el orden de las filas de los datos de entrada en la transformación. Seleccione esta opción si el servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas.

Cuando el servicio de integración de datos realiza optimizaciones, puede hacer que se pierda un orden establecido anteriormente en la asignación. Puede establecer el orden en una asignación con un origen de archivo sin formato ordenado, un origen relacional ordenado o una transformación de ordenación. Al configurar una transformación para mantener el orden de las filas, el servicio de integración de datos tiene en cuenta esta configuración cuando realiza optimizaciones para la asignación. El servicio de integración de

datos realiza optimizaciones para la transformación si puede mantener el orden. El servicio de integración de datos no realiza optimizaciones para la transformación si la optimización cambia el orden de las filas.

Objeto de datos de secuencia

Un objeto de datos de secuencia crea y mantiene una secuencia de valores numéricos. La transformación de generador de secuencia utiliza el objeto de datos de secuencia para generar los valores de la transformación.

Puede utilizar un objeto de datos de secuencia reutilizable en varias transformaciones de generador de secuencia. Todas las transformaciones de generador de secuencia que utilizan el mismo objeto de datos de secuencia utilizan la misma secuencia de valores si se ejecutan en el mismo Servicio de integración. También puede utilizar un objeto de datos de secuencia reutilizable en una transformación de generador de secuencia no reutilizable. Puede utilizar un objeto de datos de secuencia no reutilizable en una transformación de generador de secuencia no reutilizable.

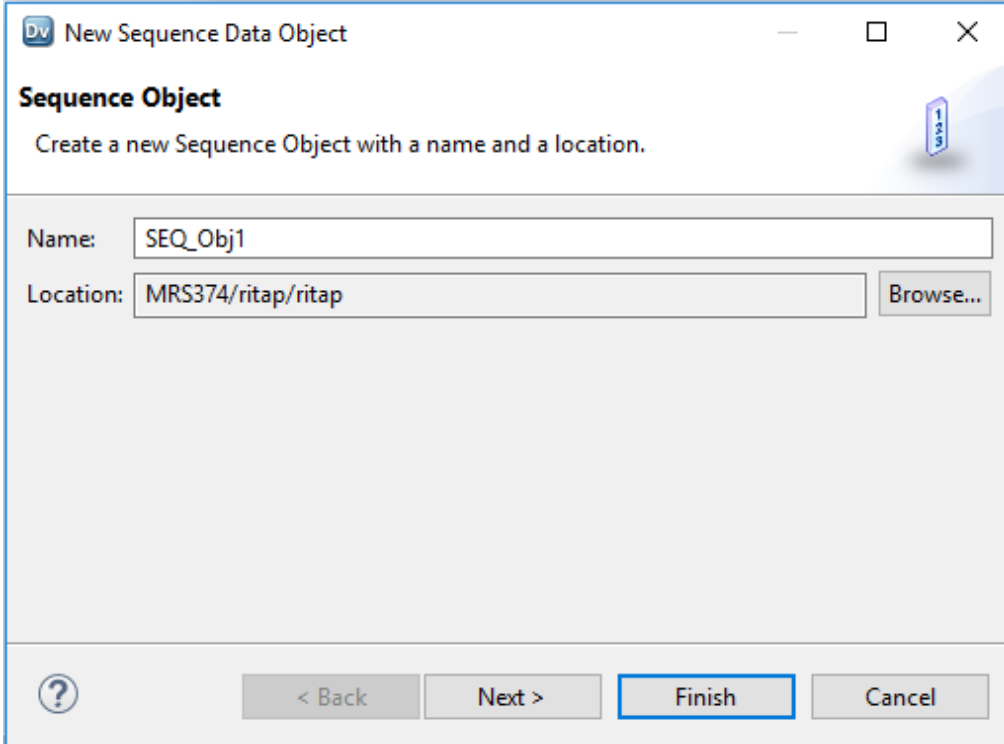
Por ejemplo, va a crear varias asignaciones que escriben en el mismo campo de clave principal en una tabla relacional. Cada asignación utiliza la misma transformación de generador de secuencia reutilizable, que utiliza el mismo objeto de datos de secuencia reutilizable y se ejecuta en el mismo Servicio de integración. Cada asignación escribe valores únicos en el campo de clave principal.

Crear un objeto de datos de secuencia

Para utilizar un objeto de datos de secuencia y crear una transformación de generador de secuencia, cree el objeto de datos de secuencia, configure las propiedades del objeto y seleccione el objeto en el cuadro de diálogo Transformación de generador de secuencia.

1. En el editor de asignación, desplácese hacia abajo en la paleta de asignación para localizar la transformación de generador de secuencia y arrastrarla hasta la asignación.
Se abre el asistente para **nuevas transformaciones**.
2. Haga clic en **Nuevo objeto de datos de secuencia**.

Se abre el asistente **Nuevo objeto de datos**.



New Sequence Data Object

Sequence Object

Create a new Sequence Object with a name and a location.

Name: SEQ_Obj1

Location: MRS374/ritap/ritap

3. Especifique un nombre para el objeto de secuencia de datos.
La convención de nomenclatura para los objetos de datos de secuencia es SEQ_<nombre de objeto de datos>.
4. Haga clic en **Siguiente** para configurar las propiedades del objeto de datos de secuencia.

Si crea una transformación de generador de secuencia desde el objeto, la transformación usará las propiedades que introduzca para el objeto de datos. En la siguiente imagen se muestran las propiedades que puede configurar:

Dv New Sequence Data Object

Sequence Object
Enter the Sequence

Start Value: 0

End Value: 9223372036854775807

Increment Value: 1

Cycle:

? < Back Next > Finish Cancel

5. Cuando haya configurado las propiedades del objeto de datos, podrá crear una transformación de generador de secuencia mediante el objeto de datos de secuencia. Al crear la transformación, nombre la transformación de generador de secuencia y seleccione **Elegir un objeto de secuencia existente**. Vaya hasta el objeto de datos y haga clic en **Aceptar**.

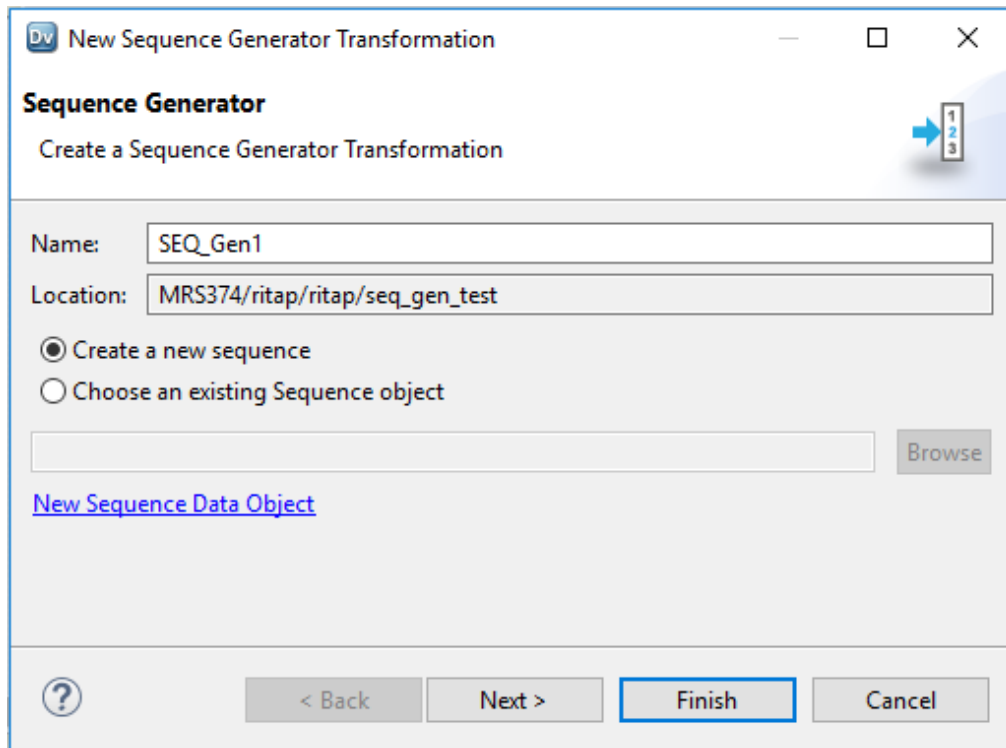
La transformación de generador de secuencia aparece en el editor de asignaciones con un puerto de solo salida NEXTVAL. Puede conectar el puerto NEXTVAL a un destino o una transformación de nivel inferior para generar una secuencia de números.

Cómo crear una transformación de generador de secuencia

Para utilizar una transformación de generador de secuencia en una asignación, añádala a la asignación, configure las propiedades de la transformación y luego conecte NEXTVAL a una o varias transformaciones.

1. En el editor de asignación, desplácese hacia abajo en la paleta de asignación para localizar la transformación de generador de secuencia y arrastrarla hasta la asignación.

Se abre el asistente para **nuevas transformaciones**.



New Sequence Generator Transformation

Sequence Generator
Create a Sequence Generator Transformation

Name: SEQ_Gen1

Location: MRS374/ritap/ritap/seq_gen_test

Create a new sequence
 Choose an existing Sequence object

Browse

[New Sequence Data Object](#)

? < Back Next > Finish Cancel

2. Introduzca un nombre para la transformación de generador de secuencia.
La convención de nomenclatura para las transformaciones de generador de secuencia es SEQ_<nombre de transformación>.
3. Elija crear una secuencia nueva o utilizar un objeto de secuencia existente.

- Para crear una secuencia nueva, seleccione **Crear una nueva secuencia**. Haga clic en **Siguiente** para configurar las propiedades de secuencia. En la siguiente imagen se muestran las propiedades que puede configurar:

- Para utilizar un objeto de secuencia existente, seleccione **Elegir un objeto de secuencia existente**. Vaya hasta el objeto de secuencia y haga clic en **Aceptar**.

La transformación de generador de secuencia aparece en el editor de asignaciones con un puerto de solo salida NEXTVAL. Puede conectar el puerto NEXTVAL a un destino o una transformación de nivel inferior para generar una secuencia de números.

Preguntas frecuentes

¿Puedo cambiar una transformación de generador de secuencia no reutilizable para hacer que sea reutilizable?

No puede hacer que la transformación sea reutilizable, pero puede cambiar la transformación para que utilice un objeto de datos de secuencia. El objeto de datos de secuencia mantiene la integridad de la secuencia, independientemente del número de transformaciones que utilice.

¿Puedo colocar una transformación de generador de secuencia no reutilizable en un mapplet?

No, no puede hacerlo. Los mapplets son objetos reutilizables, de modo que todos los objetos del mapplet también deben ser reutilizables. Utilice en su lugar una transformación de generador de secuencia reutilizable.

Transformación de generador de secuencia en un entorno no nativo

El procesamiento de transformaciones de generador de secuencia en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible cuando el dominio y el clúster de Databricks residen en la misma red virtual en entornos de AWS o Azure Databricks.

Transformación de generador de secuencia en el motor de Blaze

Una asignación con una transformación de generador de secuencia consume recursos de forma significativa cuando se cumplen las siguientes condiciones:

- La transformación está configurada para mantener el orden de las filas.
- La asignación se ejecuta en una sola partición.

Transformación de generador de secuencia en el motor de Spark

La transformación de generador de secuencia no mantiene el orden de las filas en los datos de salida. Si habilita la propiedad **Mantener orden de filas** en la transformación, el servicio de integración de datos la omite.

Nota: El generador de secuencias de Spark con mantener orden de filas funciona bien si el conjunto de datos es relativamente pequeño y puede ejecutarse o forzarse a ejecutarse en una sola partición.

CAPÍTULO 42

Transformación de ordenación

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de ordenación, 638](#)
- [Transformaciones de ordenación en asignaciones dinámicas, 639](#)
- [Cómo desarrollar una transformación de incorporación, 639](#)
- [Puertos de transformación de ordenación, 640](#)
- [Ficha Ordenar, 640](#)
- [Configurar claves de ordenación, 640](#)
- [Propiedades avanzadas de la transformación de ordenación, 643](#)
- [Memoria caché de ordenación, 644](#)
- [Cómo crear una transformación de ordenación, 644](#)
- [Ejemplo de transformación de ordenación, 645](#)
- [Transformación de ordenación en un entorno no nativo, 647](#)

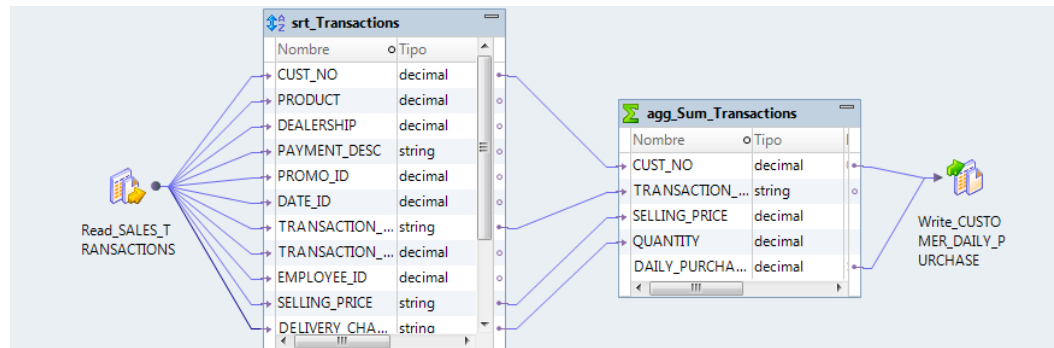
Resumen de la transformación de ordenación

Utilice una transformación de ordenación para ordenar datos de forma ascendente o descendente de acuerdo con una clave de ordenación especificada. Puede configurar la transformación de ordenación para la ordenación con distinción de mayúsculas y minúsculas y, para la salida distinta. La transformación de ordenación es una transformación activa.

Cuando se crea una transformación de ordenación se especifican puertos como claves de ordenación, y se configura cada uno de los puertos de clave de ordenación para que clasifique en orden ascendente o descendente. El Servicio de integración de datos ordena cada puerto secuencialmente cuando se especifican varios puertos para la clave de ordenación.

Por ejemplo, necesita crear una factura para el total de ventas de clientes a partir de una base de datos de clientes. Utilice una transformación de ordenación en la tabla de ventas de clientes para ordenar los datos en orden descendente según el número de cliente. Utilice el resultado de la transformación de ordenación como entrada de la transformación de agregación. Puede aumentar el rendimiento de una transformación de agregación con la opción de entrada ordenada.

La siguiente figura muestra la asignación:



Transformaciones de ordenación en asignaciones dinámicas

Puede utilizar una transformación de ordenación en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados.

Puede hacer referencia a un puerto dinámico o a un puerto generado en la transformación de ordenación. Sin embargo, si el puerto generado no existe en tiempo de ejecución, la asignación fallará.

Si utiliza un puerto dinámico como clave de ordenación, el Servicio de integración de datos tendrá en cuenta todos los puertos generados en el puerto dinámico y el orden de los puertos generados.

Puede parametrizar las claves de ordenación. Utilice un parámetro de lista de ordenación para la clave de ordenación.

Cómo desarrollar una transformación de incorporación

Cuando se desarrolla una transformación de incorporación, se deben tener en cuenta factores como los puertos de clave de ordenación, filas de salida distintas y los criterios de ordenación con distinción de mayúsculas y minúsculas.

Tenga en cuenta los siguientes factores cuando desarrolle una transformación de ordenación:

- Los puertos que se desea configurar como claves de ordenación y la dirección de la ordenación.
- Si se desea una ordenación que distinga mayúsculas de minúsculas.
- Si se desea tener en cuenta valores nulos como prioridad de ordenación.
- Si se desean filas de salida distintas.
- El valor del tamaño de la memoria caché que se desea establecer.

Puertos de transformación de ordenación

Al crear puertos en la transformación de ordenación, se crean puertos de entrada y de salida puertos de forma predeterminada. La transformación de ordenación devuelve los mismos puertos de salida como puertos de entrada.

Puede definir puertos dinámicos en una transformación de ordenación. Un puerto dinámico puede recibir distintas columnas de datos de una transformación de nivel superior de una asignación. Esto permite a la transformación de ordenación ordenar las filas que contienen columnas distintas.

Defina las claves de ordenación en la ficha **Ordenar** de la vista **Propiedades**.

Ficha Ordenar

Defina la clave de ordenación en la ficha **Ordenar** de la vista **Propiedades** de la transformación de ordenación. Seleccione uno o más puertos que quiera usar como criterios de ordenación.

El Servicio de integración de datos ordena los datos según el orden de los puertos en la ficha Ordenar. Configure para ordenar los datos en orden ascendente o descendente. El valor predeterminado es ascendente.

Si la transformación de ordenación se configura para filas de salida distintas, Developer tool configura todos los puertos como parte de la clave de ordenación. El Servicio de integración de datos descarta las filas duplicadas durante la operación de ordenación.

Configurar claves de ordenación

Defina la clave de ordenación en la ficha **Ordenar** de la vista **Propiedades** de la transformación.

Este es el inicio del concepto.

En la siguiente imagen se muestra la ficha **Ordenar**:

Sort

Output: All rows Distinct rows only

Sort Keys

Specify by: Value ▼

Ports: Add

Department	Ascending (A)
Employee	Ascending (A)

Choose...
Delete
Move Up
Move Down

La ficha **Ordenar** contiene las siguientes opciones:

Salida

Elija si desea obtener todas las filas ordenadas o descartar las filas duplicadas. Las filas duplicadas son aquellas en las que todos los valores de columna son los mismos.

Especificar por

Seleccione **Valor** o **Parámetro**. Seleccione **Valor** para usar los nombres de puerto. Seleccione **Parámetro** para utilizar un parámetro de lista de ordenación.

Añadir

Se acepta el nombre del puerto que escriba manualmente. Debe escribir un nombre de puerto válido antes de hacer clic en **Añadir**.

Elegir

Haga clic en **Elegir** para seleccionar los puertos que quiera añadir a la clave de ordenación. Developer tool proporciona una lista de los puertos de la transformación para su selección.

Subir y Bajar

Puede cambiar el orden de puertos del grupo. Seleccione el nombre de puerto y haga clic en uno de los botones para subirlo o bajarlo en la ordenación.

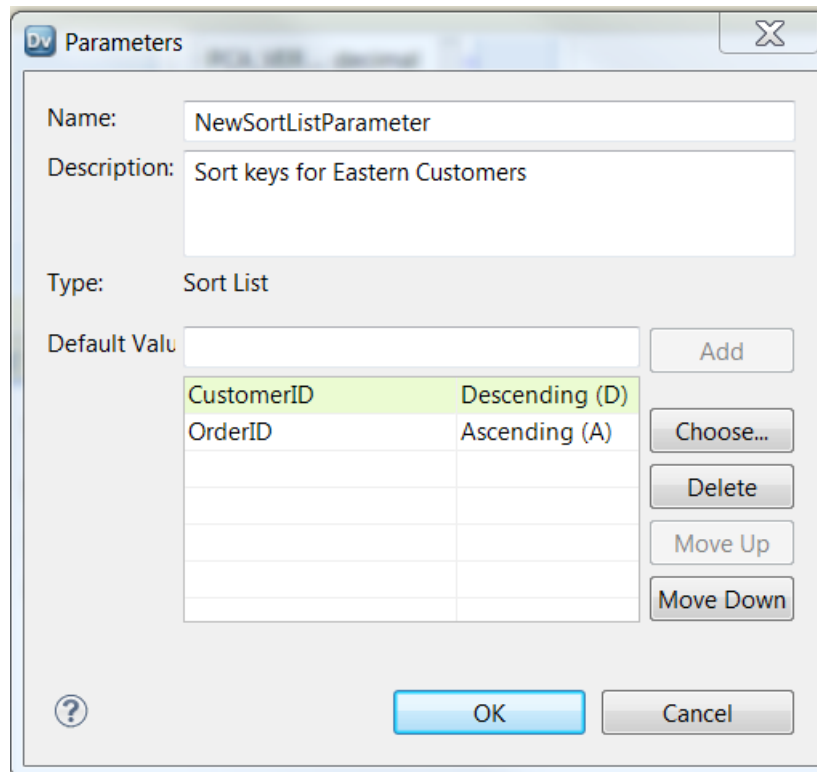
Parametrizar las claves de ordenación

Puede crear un parámetro de lista de ordenación que contenga una lista de puertos para las claves de ordenación.

Si la transformación de ordenación está en una asignación dinámica, puede que contenga puertos generados. Puede parametrizar las claves de ordenación. Cree un parámetro de lista de ordenación que contenga una lista de puertos por la que establecer la ordenación.

En la ficha **Ordenar** de las propiedades de la transformación, elija **Especificar por Parámetro**. Haga clic en **Nuevo** para crear un parámetro.

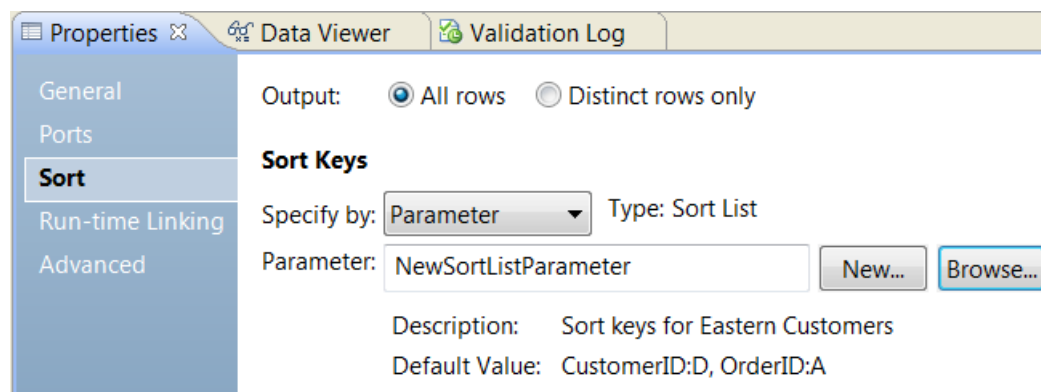
La siguiente imagen muestra el cuadro de diálogo **Parámetros**:



Elija los puertos o puertos generados de las claves de ordenación. Puede elegir un tipo de orden ascendente o descendente.

Puede introducir manualmente los nombres de puerto. Escriba el nombre del puerto en el campo **Valor predeterminado** y haga clic en **Añadir**. Developer tool añade el nombre del puerto a la lista de ordenación.

La siguiente imagen muestra la ficha **Ordenar** después de configurar un parámetro para las claves de ordenación:



Propiedades avanzadas de la transformación de ordenación

En las propiedades avanzadas de la transformación de ordenación se pueden especificar criterios de clasificación adicionales. El Servicio de integración de datos aplica las propiedades a todos los puertos de clave de clasificación. Las propiedades de la transformación de ordenación también determinan los recursos del sistema que asigna el Servicio de integración de datos cuando clasifica datos.

La siguiente sección describe las propiedades avanzadas para una transformación de ordenación:

Distinguir mayúsculas de minúsculas

Determina si el Servicio de integración de datos tiene en cuenta el formato de mayúsculas y minúsculas en la clasificación de los datos. Cuando se habilita la propiedad Distinguir mayúsculas de minúsculas, el Servicio de integración de datos ordena antes los caracteres en mayúsculas que los caracteres en minúsculas. De forma predeterminada, Developer tool define Distinguir mayúsculas de minúsculas.

Valor nulo como inferior

Trata un valor nulo como inferior a cualquier otro valor. Habilite la propiedad si desea que el Servicio de integración de datos trate un valor como inferior respecto a cualquier otro valor cuando realiza la operación de ordenación.

Tamaño de la memoria caché de ordenación

Cantidad de memoria que el Servicio de integración de datos asigna al inicio de la ejecución de la asignación para realizar la operación de ordenación. El Servicio de integración de datos pasa todos los datos entrantes a la transformación de ordenación antes de realizar la operación de ordenación. Seleccione Auto para que el Servicio de integración de datos calcule automáticamente los requisitos de memoria en tiempo de ejecución. Introduzca un valor específico en bytes al ajustar el tamaño de la memoria caché. El valor predeterminado es Automático.

Directorio de trabajo

Directorio donde el Servicio de integración de datos almacena datos temporalmente si la cantidad de datos entrantes es mayor que el tamaño de la memoria caché de ordenación. Una vez que el Servicio de integración de datos ha ordenado los datos, elimina los archivos temporales.

Introduzca varios directorios separados por punto y coma para aumentar el rendimiento durante la creación de particiones en la memoria caché. La creación de particiones en la memoria caché crea una memoria caché independiente para cada partición que procesa la transformación.

El valor predeterminado es el parámetro del sistema TempDir. Puede configurar otro parámetro del sistema o definido por el usuario en este campo.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

TEMAS RELACIONADOS

- [“Tamaño de la memoria caché” en la página 74](#)

Memoria caché de ordenación

El Servicio de integración de datos crea una caché en la memoria para ejecutar la transformación de ordenación. El Servicio de integración de datos pasa todos los datos entrantes a la transformación de ordenación antes de realizar la operación de ordenación. Si el Servicio de integración de datos necesita más espacio del que hay disponible en la memoria caché, almacena temporalmente los datos en el directorio de trabajo de la transformación de ordenación.

Si no se configura el tamaño de la memoria caché para ordenar todos los datos en la memoria, aparece una advertencia en el registro de la sesión que indica que el Servicio de integración de datos realizó varios pases por los datos de origen. El Servicio de integración de datos hace varios pases por los datos cuando tiene que paginar información en el disco para completar la ordenación. El mensaje especifica la cantidad de memoria necesaria para un solo pase, que es cuando el Servicio de integración de datos lee los datos una vez y realiza la ordenación en la memoria sin necesidad de paginar en el disco. Para optimizar el rendimiento de la asignación, configure el tamaño de la memoria caché para que el Servicio de integración de datos solo tenga que hacer un pase por los datos.

Si la cantidad de datos entrantes es mayor que el tamaño de la memoria caché de ordenación, el Servicio de integración de datos almacena datos temporalmente en el directorio de trabajo de la transformación de ordenación. El Servicio de integración de datos necesita un espacio en disco de como mínimo el doble de la cantidad de datos entrantes cuando ordena datos en el directorio de trabajo.

Para obtener el mejor rendimiento, configure el tamaño de la memoria caché de ordenación con un valor menor o igual que la cantidad de memoria física disponible en el equipo que ejecuta la asignación. Para ordenar los datos utilizando una transformación de ordenación, asigne al menos 16 MB (16 777 216 bytes) de memoria física. De forma predeterminada, el tamaño de la memoria caché de ordenación se establece como Auto.

Optimizar la memoria caché de ordenación

La memoria caché de ordenación se ha optimizado para que utilice la longitud variable para almacenar los tipos de datos binarios y de cadena que pasen por la transformación de ordenación.

La longitud variable reduce la cantidad de datos que el servicio de integración de datos almacena en la memoria caché de ordenación, así como el consumo de espacio en disco en el equipo del servicio de integración de datos.

Por ejemplo, suponga que tiene almacenados datos sobre sus clientes. Algunos clientes tienen nombres más largos que otros. Si el servicio de integración de datos utiliza la longitud fija para almacenar los datos de los nombres de clientes, podría almacenarlos en 20 caracteres por cada nombre. Si, en cambio, utiliza la longitud variable, podría almacenar datos con una longitud media de 10 caracteres.

Cómo crear una transformación de ordenación

Puede crear una transformación de ordenación reutilizable o no reutilizable.

Cómo crear una transformación de ordenación reutilizable

Cree una transformación de ordenación reutilizable para utilizarla en varias asignaciones o mapplets.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de ordenación.
4. Haga clic en **Siguiente**.
5. Especifique un nombre para la transformación.
6. Haga clic en **Finalizar**.
La transformación aparece en el editor.
7. Haga clic en **Nueva** para añadir un puerto a la transformación.
8. Edite el puerto para definir el nombre, el tipo de datos y la precisión.
9. En la ficha **Ordenar**, elija la puertos por los que quiera ordenar o seleccione un parámetro de lista de ordenación.
10. Haga clic en la vista **Avanzadas** y edite las propiedades de la transformación.

Cómo crear una transformación de ordenación no reutilizable

Cree una transformación de ordenación no reutilizable como parte de una asignación o de un mapplet.

1. En una asignación o en un mapplet, arrastre una transformación de ordenación desde la paleta Transformación hasta el editor.
La transformación aparece en el editor.
2. En la vista **Propiedades**, edite el nombre y la descripción de la transformación.
3. En la ficha **Puertos**, haga clic en **Nuevo** para añadir puertos a la transformación.
4. Edite los puertos para definir el nombre, el tipo de datos y la precisión.
5. Seleccione **Clave** para indicar el puerto como clave de ordenación.
6. Haga clic en la vista **Avanzadas** y edite las propiedades de la transformación.

Ejemplo de transformación de ordenación

Usted tiene una tabla de base de datos PRODUCT_ORDERS que contiene información sobre todos los pedidos efectuados por el cliente.

ORDER_ID	ITEM_ID	ITEM	QUANTITY	PRICE
43	123456	ItemA	3	3.04
41	456789	ItemB	2	12.02
43	000246	ItemC	6	34.55
45	000468	ItemD	5	0.56
41	123456	ItemA	4	3.04

ORDER_ID	ITEM_ID	ITEM	QUANTITY	PRICE
45	123456	ItemA	5	3.04
45	456789	ItemB	3	12.02

Utilice la transformación de ordenación en PRODUCT_ORDERS y especifique ORDER_ID como clave de ordenación descendente.

Después de ordenar los datos, el servicio de integración de datos pasa las siguientes filas fuera de la transformación de ordenación:

ORDER_ID	ITEM_ID	ITEM	QUANTITY	PRICE
45	000468	ItemD	5	0.56
45	123456	ItemA	5	3.04
45	456789	ItemB	3	12.02
43	123456	ItemA	3	3.04
43	000246	ItemC	6	34.55
41	456789	ItemB	2	12.02
41	123456	ItemA	4	3.04

Necesita averiguar el importe total y la cantidad de artículos de cada pedido. Puede utilizar el resultado de la transformación de ordenación como entrada de la transformación de agregación. Utilice la entrada ordenada en la transformación de agregación para aumentar el rendimiento.

Si no se utiliza la entrada ordenada, el servicio de integración de datos realiza los cálculos de agregados a medida que lee las filas. El servicio de integración de datos almacena datos para cada grupo hasta que lee el origen por completo para asegurarse de que todos los cálculos de agregados se realizan de forma precisa. Si se utiliza la entrada ordenada y no se ordenan previamente los datos correctamente, se obtienen resultados inesperados.

La transformación de agregación tiene los puertos agrupar por ORDER_ID e ITEM con la opción de entrada ordenada seleccionada. Cuando se pasan los datos desde la transformación de ordenación, la transformación de agregación agrupa ORDER_ID para calcular el importe total de cada pedido.

ORDER_ID	SUM
45	54.06
43	216.42
41	36.2

Transformación de ordenación en un entorno no nativo

El procesamiento de transformaciones de ordenación en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de ordenación en el motor de Blaze

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- El destino está configurado para que mantenga el orden de las filas, y la transformación de ordenación no está conectada directamente a un destino de archivo sin formato.

Ordenación paralela

El servicio de integración de datos permite la ordenación paralela con las siguientes restricciones:

- La asignación no incluye otra transformación entre la transformación de ordenación y el destino.
- El tipo de datos de las claves de ordenación no cambia entre la transformación de ordenación y el destino.
- Cada clave de ordenación de la transformación de ordenación debe estar vinculada a una columna en el destino.

Orden global

El motor de Blaze puede realizar ordenaciones globales en las siguientes situaciones:

- La transformación de ordenación está conectada directamente a destinos de archivos sin formato.
- El destino está configurado para mantener el orden de las filas.
- La clave de ordenación no es de un tipo de datos binario.

Si alguna de las condiciones no se cumple, el motor de Blaze realiza una ordenación local.

Optimización de la memoria caché de datos

Si se inserta una transformación de ordenación antes de la transformación de agregación o de rango para optimizar la memoria caché de datos de agregación o de rango, el tamaño de la memoria caché de ordenación es igual que el de la memoria caché de datos de la transformación de agregación o de rango. Para configurar la memoria caché de ordenación, debe configurar el tamaño de la memoria caché de datos de la transformación de agregación o de rango.

Transformación de ordenación en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones no se realiza si la distinción entre mayúsculas y minúsculas está deshabilitada.

El servicio de integración de datos registra una advertencia y omite la transformación de ordenación en las siguientes situaciones:

- Existe una discrepancia de tipo entre el destino y las claves de ordenación de la transformación de ordenación.
- La transformación contiene claves de ordenación que no están conectadas al destino.
- La transformación de escritura no está configurada para mantener el orden de las filas.
- La transformación no está en un nivel directamente superior al de la transformación de escritura.

Valores nulos

El servicio de integración de datos trata los valores nulos como bajos aunque haya configurado la transformación para que trate los valores nulos como altos.

Optimización de la memoria caché de datos

La memoria caché de ordenación no se puede optimizar para que almacene datos usando la longitud variable.

Transformación de ordenación en una asignación de transmisión

Las asignaciones de transmisión tienen reglas de procesamiento adicionales que no aplican a las asignaciones por lotes.

Validación de asignaciones

La validación de asignaciones genera un error en la siguiente situación:

- Una transformación de ordenación en una asignación de transmisión no debe tener una transformación de agregación de nivel superior.

Directrices generales

Tenga en cuenta las siguientes directrices generales:

- La asignación se ejecuta en el modo de salida completo si contiene una transformación de ordenación.
- Para mantener el orden de clasificación global, compruebe que el destino tiene solo una partición. El origen puede tener una o varias particiones.

Transformación de ordenación en el motor de Databricks Spark

Algunas reglas de procesamiento del motor de Databricks Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Validación de asignaciones

La validación de asignaciones no se realiza si la distinción entre mayúsculas y minúsculas está deshabilitada.

El servicio de integración de datos registra una advertencia y omite la transformación de ordenación en las siguientes situaciones:

- Existe una discrepancia de tipo entre el destino y las claves de ordenación de la transformación de ordenación.
- La transformación contiene claves de ordenación que no están conectadas al destino.
- La transformación de escritura no está configurada para mantener el orden de las filas.
- La transformación no está en un nivel directamente superior al de la transformación de escritura.

Valores nulos

El servicio de integración de datos trata los valores nulos como bajos aunque haya configurado la transformación para que trate los valores nulos como altos.

Optimización de la memoria caché de datos

La memoria caché de ordenación no se puede optimizar para que almacene datos usando la longitud variable.

Ordenación paralela

El servicio de integración de datos permite la ordenación paralela con las siguientes restricciones:

- La asignación no incluye otra transformación entre la transformación de ordenación y el destino.
- El tipo de datos de las claves de ordenación no cambia entre la transformación de ordenación y el destino.
- Cada clave de ordenación de la transformación de ordenación debe estar vinculada a una columna en el destino.

CAPÍTULO 43

Transformación de SQL

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de SQL, 650](#)
- [Puertos de transformación de SQL, 651](#)
- [Propiedades avanzadas de Transformación de SQL, 655](#)
- [Consulta de transformación de SQL, 657](#)
- [Cardinalidad entre fila de entrada y fila de salida, 660](#)
- [Optimización del filtro con la transformación de SQL, 663](#)
- [Ejemplo de transformación de SQL con una consulta SQL, 665](#)
- [Procedimientos almacenados, 669](#)
- [Conexión de transformación de SQL, 674](#)
- [Cómo crear manualmente una transformación de SQL, 675](#)
- [Creación de una transformación de SQL desde un procedimiento almacenado, 676](#)

Resumen de la transformación de SQL

La transformación de SQL procesa consultas SQL en la parte intermedia de una asignación. Puede ejecutar consultas SQL desde la transformación de SQL o puede configurar la transformación de SQL para ejecutar procedimientos almacenados desde una base de datos.

Puede pasar los valores de puertos de entrada a los parámetros en la consulta o procedimiento almacenado. La transformación puede insertar, eliminar, actualizar y recuperar filas de una base de datos. Puede ejecutar instrucciones DDL de SQL para crear una tabla o soltar una tabla en la parte intermedia de una asignación. La transformación de SQL es una transformación activa. La transformación puede devolver múltiples filas para cada fila de entrada.

Puede importar un procedimiento almacenado desde una base de datos a la transformación de SQL. Cuando se importa el procedimiento almacenado, la herramienta Developer crea los puertos de transformación que corresponden a los parámetros en el procedimiento almacenado. La herramienta Developer también crea la llamada al procedimiento almacenado de forma automática.

Para configurar una transformación de SQL para ejecutar un procedimiento almacenado, realice las tareas siguientes:

1. Defina las propiedades de la transformación, incluido el tipo de base de datos al que conectarse.

2. Importe un procedimiento almacenado para definir los puertos y cree la llamada al procedimiento almacenado.
3. Defina manualmente los puertos para los conjuntos de resultados o los procedimientos almacenados adicionales que necesite ejecutar.
4. Añada las llamadas a los procedimientos almacenados adicionales en el editor de SQL.

Puede configurar una consulta SQL en el editor de SQL de transformaciones. Cuando ejecute la transformación de SQL, la transformación procesa la consulta, devuelve filas y devuelve cualquier error de base de datos.

Para configurar una transformación de SQL para ejecutar una consulta, realice las tareas siguientes:

1. Defina las propiedades de la transformación, incluido el tipo de base de datos al que conectarse.
2. Defina los puertos de entrada y de salida.
3. Cree una consulta SQL en el editor de SQL.

Después de configurar la transformación, configure la transformación de SQL en una asignación y conecte los puertos de nivel superior. Obtenga una vista previa de los datos para comprobar los resultados.

Puertos de transformación de SQL

Cuando se crea una transformación de SQL, la herramienta Developer crea el puerto `SQLError` de forma predeterminada. Añada puertos de entrada, puertos de salida y puertos de transferencia en la vista **Puertos**.

La transformación de SQL es compatible con los siguientes tipos de puerto:

Entrada

Recibe datos de origen que pueden usarse en una consulta SQL.

Salida

Devuelve datos de base de datos desde una consulta SQL `SELECT`.

Transferencia

Los puertos de entrada-salida que pasan datos de origen a través de la transformación sin modificarlos.

SQLError

Devuelve errores SQL desde la base de datos. Si no se produce ningún error, se devuelve `NULL`.

NumRowsAffected

El puerto devuelve el número total de filas de la base de datos que se ven afectadas por las instrucciones de consulta `INSERT`, `DELETE` y `UPDATE` para una fila de entrada. La herramienta Developer crea este puerto cuando se opta por incluir las estadísticas de actualización en la fila de salida.

Valor de devolución

Recibe el valor de devolución de un procedimiento almacenado.

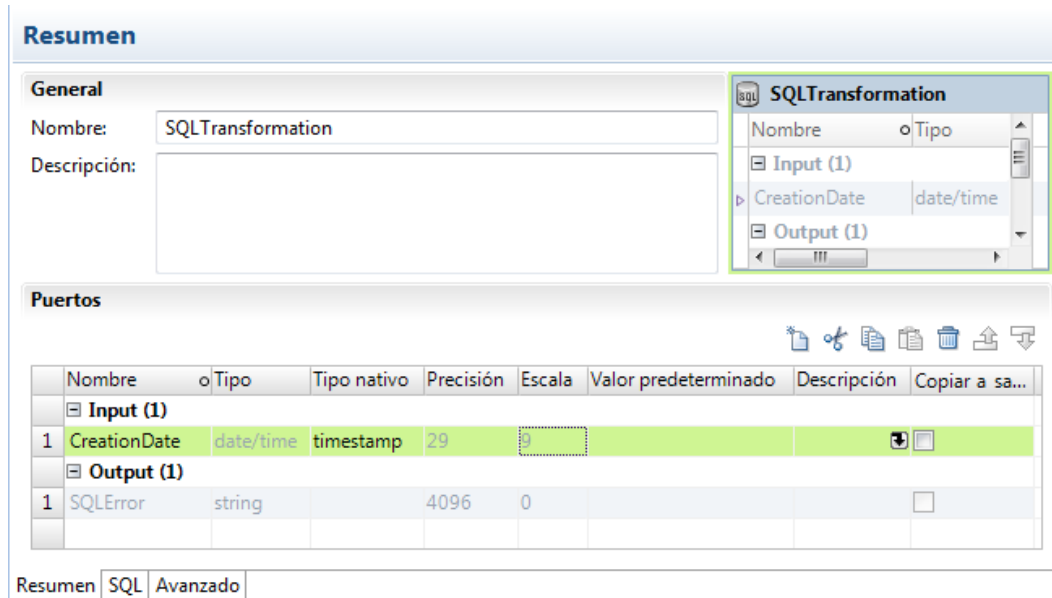
Puertos de entrada

Puede hacer referencia a los puertos de entrada de la transformación de SQL con el enlace de parámetros en cualquier tipo de instrucción SQL o procedimiento almacenado. Puede crear puertos de entrada en la transformación de SQL para los datos que no prevé pasar a puertos de salida.

Debe añadir manualmente los puertos si está configurando una consulta SQL que tiene parámetros de entrada. Cuando importe un procedimiento almacenado a la transformación de SQL, la transformación de SQL crea los puertos de entrada. Puede añadir puertos de transferencia para pasar datos a través de la transformación sin modificarlos.

Puede añadir puertos en la vista **Resumen**. Cuando añada un puerto, especifique el tipo de datos nativo para el puerto. El tipo de datos nativo es el tipo de datos válido para la base de datos a la que se está conectando. Cuando se configura un tipo de datos nativo, aparece un tipo de datos de transformación. Si arrastra filas hasta la transformación de SQL, la herramienta Developer establece el tipo de datos nativo basándose en los tipos de datos válidos para la base de datos a la que se está conectando. Compruebe que los tipos de datos de las columnas que se usan en la consulta sean los mismos tipos de datos que las columnas de la base de datos.

La siguiente figura muestra el puerto de entrada **CreationDate** en una transformación de SQL reutilizable:



Para añadir puertos de entrada, haga clic en **Entrada** en el panel **Puertos**. Haga clic en **Nuevo**.

Nota: Si selecciona **Copiar en salida** para un puerto, el puerto de entrada se convierte en un puerto de transferencia. Los puertos de transferencia aparecen en las secciones **Entrada** y **Salida** de la vista **Puertos**.

Puertos de salida

salida de transformación de SQL los puertos devuelven valores desde una instrucción de consulta o desde un procedimiento almacenado.

Debe definir los puertos de salida cuando configura manualmente una transformación de SQL. Defina un puerto de salida para cada parámetro de salida de procedimiento almacenado o para cada puerto que devuelve una instrucción SELECT.

Cuando importa un procedimiento almacenado, la herramienta Developer crea un puerto de salida para cada parámetro de salida que el procedimiento devuelve. Si el procedimiento devuelve un conjunto de resultados,

debe definir manualmente los puertos de salida en el conjunto de resultados. Un procedimiento almacenado puede devolver un conjunto de resultados y puede devolver los parámetros de salida que no forman parte del conjunto de resultados en la misma ejecución. Debe definir los puertos de salida para los campos del conjunto de resultados y para los parámetros de salida.

Cuando configure un puerto de salida, elija el tipo de datos nativo del puerto. El tipo de datos nativo del puerto de salida debe coincidir con el tipo de datos de la columna correspondiente en la base de datos. Cuando se configura el tipo de datos nativo, la herramienta Developer define el tipo de datos de la transformación para el puerto.

Por ejemplo, la transformación de SQL contiene la siguiente consulta SQL para una base de datos Oracle:

```
SELECT FirstName, LastName, Age FROM EMPLOYEES
```

Puede configurar los siguientes puertos de salida y los tipos de datos nativos en la transformación de SQL:

Puerto de salida	Tipos de datos nativos	Tipos de datos de transformación
FirstNm	varchar2	string
LastNm	varchar2	string
Age	number	double

El número de puertos de salida y el orden de los mismos debe coincidir con el número y el orden de las columnas que devuelve la consulta o el procedimiento almacenado. Cuando el número de puertos de salida es mayor que el número de columnas en la consulta o el procedimiento almacenado, los puertos adicionales devuelven un valor nulo. Cuando el número de puertos de salida es menor que el número de columnas en SQL, el servicio de integración de datos genera un error de fila.

Si se cambia el tipo de base de datos a la que se conecta la transformación, la herramienta Developer cambia los tipos nativos de los puertos de salida. La herramienta Developer puede no elegir el tipo de datos correcto para todos los puertos. Si se cambia el tipo de base de datos, compruebe que el tipo de datos nativo para cada puerto de salida sea el mismo tipo de datos que la columna de la base de datos. Por ejemplo, la herramienta Developer puede elegir nVarchar2 para una columna de base de datos Oracle. Es posible que tenga que cambiar el tipo de datos a varchar2.

Configure puertos de salida en la vista **Resumen** de la transformación de SQL.

Puertos de transferencia

Los puertos de transferencia son puertos de entrada-salida que pasan datos a través de la transformación sin cambiar los datos. La transformación de SQL devuelve datos en los puertos de transferencia independientemente de que una consulta SQL devuelva filas o no.

Cuando la fila de entrada contiene una instrucción SELECT, la transformación de SQL devuelve los datos en el puerto de transferencia para cada fila que devuelve desde la base de datos. Si el resultado de la consulta contiene varias filas, la transformación de SQL repite los datos de transferencia en cada fila.

Cuando una consulta no devuelve ninguna fila, la transformación de SQL devuelve los datos de columna de transferencia con valores nulos en las columnas de salida. Por ejemplo, las consultas que contienen instrucciones INSERT, UPDATE y DELETE no devuelven ninguna fila. Cuando la consulta genera errores, la transformación de SQL devuelve los datos de columna de transferencia, el mensaje SQLException y valores nulos en los puertos de salida.

No se puede configurar un puerto de transferencia para que devuelva datos desde una consulta SELECT.

Para crear un puerto de transferencia, cree un puerto de entrada y seleccione **Copiar en salida**. La herramienta Developer crea un puerto de salida y añade un sufijo “_output” al nombre del puerto. No se puede cambiar el puerto de salida que la herramienta Developer crea para un puerto de transferencia. No se puede crear un puerto de salida con el sufijo “_output”.

La siguiente figura muestra un puerto de transferencia Nombre en una transformación de SQL reutilizable:

The screenshot shows the 'Resumen' (Summary) tab of the SQL Transformation tool. The 'General' section shows the name 'SQLTransformation' and an empty description. The 'Puertos' (Ports) section contains a table with the following data:

	Nombre	Tipo	Tipo nativo	Precisión	Escala	Valor predeterminado	Descripción	Copiar a salida
Input (2)								
1	Name	string	nvarchar2	29	0			<input checked="" type="checkbox"/>
2	CreationDate	date/time	timestamp	29	9			<input type="checkbox"/>
Output (2)								
1	SQLError	string		4096	0			<input type="checkbox"/>
2	Name_output	string	nvarchar2	29	0			<input type="checkbox"/>

At the bottom, there are tabs for 'Resumen', 'SQL', and 'Avanzado'.

Puerto SQLError

El puerto SQLError devuelve errores SQL desde la base de datos de los procedimientos almacenados o las consultas SQL.

La siguiente figura muestra el puerto SQLError en una transformación de SQL reutilizable:

The screenshot shows the 'Resumen' (Summary) tab of the SQL Transformation tool. The 'General' section shows the name 'SQLTransformation' and an empty description. The 'Puertos' (Ports) section contains a table with the following data:

	Nombre	Tipo	Tipo nativo	Precisión	Escala	Valor predeterminado	Descripción	Copiar a salida
Input (0)								
Output (1)								
1	SQLError	string		4096	0			<input type="checkbox"/>

At the bottom, there are tabs for 'Resumen', 'SQL', and 'Avanzado'.

Cuando la consulta SQL contiene errores de sintaxis, el puerto `SQLException` contiene el texto de error de la base de datos. Por ejemplo, la siguiente consulta SQL genera un error de fila desde una base de datos Oracle:

```
SELECT Product_ID FROM Employees
```

La tabla `Employees` no contiene `Product_ID`. El servicio de integración de datos genera una fila. El puerto `SQLException` contiene el texto del error en una línea:

```
ORA-0094: "Product_ID": invalid identifier Database driver error... Function Name:
Execute SQL Stmt: SELECT Product_ID from Employees Oracle Fatal Error
```

Puede configurar varias instrucciones de consulta en la consulta SQL o puede llamar a varios procedimientos almacenados. Cuando se configure la transformación de SQL para que continúe ante un error SQL, la transformación de SQL puede devolver filas para una instrucción de consulta, aunque devuelve errores de base de datos para otra instrucción de consulta. La transformación de SQL devuelve cualquier error de base de datos en una fila aparte.

Número de filas afectadas

Habilite el puerto de salida `NumRowsAffected` para devolver el número de filas que las instrucciones de consulta `INSERT`, `UPDATE` o `DELETE` cambian para cada fila de entrada. Puede configurar el puerto de salida `NumRowsAffected` para las consultas SQL.

El servicio de integración de datos devuelve `NumRowsAffected` para cada instrucción de la consulta. De forma predeterminada, `NumRowsAffected` está deshabilitado.

Cuando se habilita `NumRowsAffected` y la consulta SQL no contiene una instrucción `INSERT`, `UPDATE` o `DELETE`, `NumRowsAffected` es cero en cada fila de salida.

Cuando la consulta SQL contiene varias instrucciones, el servicio de integración de datos devuelve `NumRowsAffected` para cada instrucción. `NumRowsAffected` contiene el número de filas que las instrucciones `INSERT`, `UPDATE` y `DELETE` cambian para una fila de entrada.

Por ejemplo, una consulta contiene las siguientes instrucciones:

```
DELETE from Employees WHERE Employee_ID = '101';
SELECT Employee_ID, LastName from Employees WHERE Employee_ID = '103';
INSERT into Employees (Employee_ID, LastName, Address)VALUES ('102', 'Gein', '38 Beach Rd')
```

La instrucción `DELETE` afecta una fila. La instrucción `SELECT` no afecta ninguna fila. La instrucción `INSERT` afecta una fila.

El servicio de integración de datos devuelve una fila desde la instrucción `DELETE`. `NumRowsAffected` es igual a uno. El servicio de integración de datos devuelve una fila desde la instrucción `SELECT` y `NumRowsAffected` es cero. El servicio de integración de datos devuelve una fila desde la instrucción `INSERT` y `NumRowsAffected` es uno.

Propiedades avanzadas de Transformación de SQL

Es posible modificar las propiedades de una transformación de SQL en cualquier momento. El tipo de base de datos predeterminado es Oracle. Si tiene que conectarse a otro tipo de base de datos, cambie el tipo de base de datos antes de añadir puertos a la transformación.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal. Cuando se configura el nivel de seguimiento de la transformación de SQL como Datos detallados, el Servicio de integración de datos escribe cada consulta SQL que prepara en el registro de la asignación.

Tipo de conexión

Describe cómo se conecta el Servicio de integración de datos con la base de datos. El tipo de base de datos es estática. El Servicio de integración de datos se conecta una vez a la base de datos. Seleccione un objeto de conexión a base de datos en la transformación de SQL. Solo lectura.

Tipo de BD

Tipo de base de datos a la que se conecta la transformación de SQL. Elija un tipo de base de datos de la lista. Puede seleccionar Oracle, Microsoft SQL Server, IBM DB2 o ODBC. El tipo de base de datos influye en los tipos de datos que se pueden asignar en la ficha **Puertos**. Cuando se cambia el tipo de base de datos, Developer tool cambia los tipos de datos de los puertos de entrada, de salida y de transferencia.

Continuar si hay errores en una fila

Continúa procesando las instrucciones SQL restantes de una consulta después de que se produzca un error de SQL.

Incluir estadísticas como salida

Añade un puerto de salida NumRowsAffected. El puerto devuelve el número total de filas de la base de datos que se ven afectadas por las instrucciones de consulta INSERT, DELETE y UPDATE para una fila de entrada.

Recuento máximo de filas de salida

Define el número máximo de filas que proporciona la transformación de SQL tras una consulta SELECT. Para configurar un número ilimitado de filas, establezca en cero el valor de Recuento de filas de salida máximo.

Descripción de la consulta

Descripción de la consulta SQL que se define en la transformación.

Modo SQL

Determina si la consulta SQL es una secuencia de comandos externa o si la consulta se define en la transformación. El modo de SQL es Consulta. La transformación de SQL ejecuta una consulta que se define en el Editor de SQL. Solo lectura.

Consulta SQL

Muestra la consulta SQL que se configura en el Editor SQL.

Tiene efectos secundarios

Indica que la transformación de SQL realiza una función además de devolver filas. La transformación de SQL tiene efectos secundarios cuando la consulta SQL actualiza una base de datos. Habilite **Tiene efectos secundarios** cuando la consulta SQL contenga una instrucción como CREATE, DROP, INSERT, UPDATE, GRANT o REVOKE.

La transformación de SQL también tiene efectos secundarios si la transformación devuelve filas NULL para las instrucciones SELECT que no devuelven ningún resultado. Las filas pueden contener valores de puerto de transferencia, información de errores de SQL o el campo NUMRowsAffected.

Deshabilite la propiedad **Tiene efectos secundarios** para permitir la optimización de inserción o la optimización de primera selección. Esta propiedad está habilitada de manera predeterminada.

Devolver solo salida de base de datos

La transformación de SQL no genera filas para las instrucciones SELECT que devuelven 0 resultados, filas para otras instrucciones como INSERT, UPDATE, DELETE o COMMIT o filas nulas.

Habilitar la optimización de inserción

Habilita el Servicio de integración de datos para insertar la lógica desde una transformación de filtro en la asignación en SQL para la transformación de SQL.

Mantener orden de filas

Mantenga el orden de las filas de los datos de entrada en la transformación. Seleccione esta opción si el Servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas.

Cuando el Servicio de integración de datos realiza optimizaciones, puede hacer que se pierda un orden establecido anteriormente en la asignación. Puede establecer el orden en una asignación con un origen de archivo sin formato ordenado, un origen relacional ordenado o una transformación de ordenación. Al configurar una transformación para mantener el orden de las filas, el Servicio de integración de datos tiene en cuenta esta configuración cuando realiza optimizaciones para la asignación. El Servicio de integración de datos realiza optimizaciones para la transformación si puede mantener el orden. El Servicio de integración de datos no realiza optimizaciones para la transformación si el optimización cambia el orden de las filas.

Admite partición

La transformación se puede procesar con varios subprocesos. Desactive esta opción si quiere que el Servicio de integración de datos utilice un subproceso para procesar la transformación. El Servicio de integración de datos puede usar varios subprocesos para procesar el resto de etapas del canal de la asignación.

Deshabilite las particiones para una transformación SQL cuando las consultas SQL requieran que la transformación se procese con un único subproceso. O bien, puede que desee deshabilitar las particiones para una transformación SQL de modo que solo se establezca una conexión con la base de datos.

Consulta de transformación de SQL

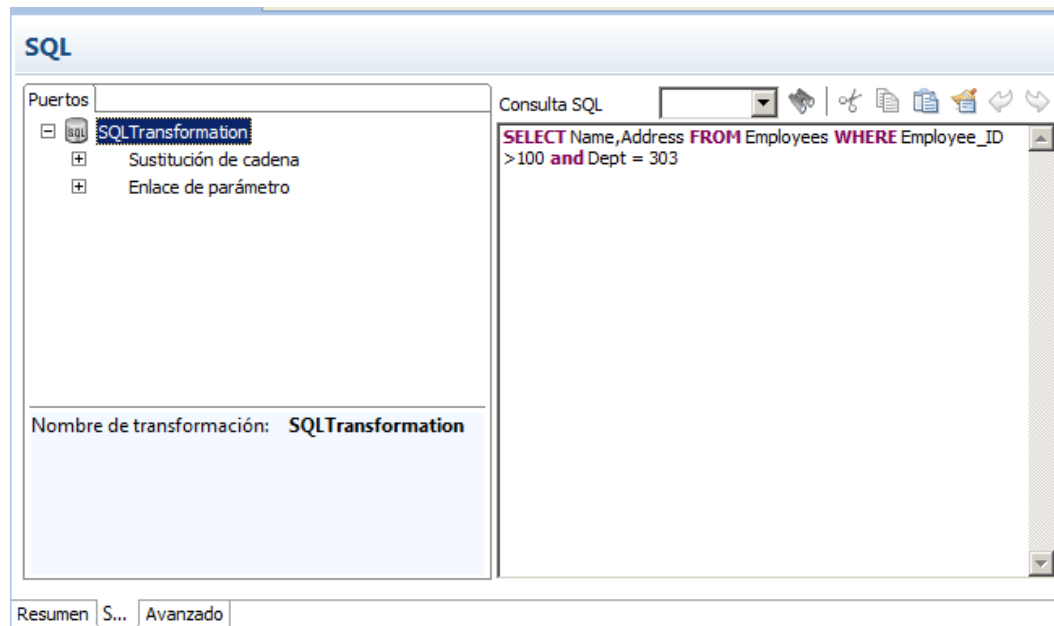
Cree una consulta SQL en el editor de SQL para recuperar filas de una base de datos, o para actualizar la base de datos.

Para crear una consulta, escriba la instrucción de la consulta en el editor de SQL en la vista SQL. Se pueden introducir hasta 32 767 caracteres en una instrucción de consulta SQL. El editor de SQL proporciona una lista de los puertos de transformación a los que se puede hacer referencia en la consulta. Puede hacer doble clic en un nombre de puerto para añadirlo como parámetro de consulta.

Cuando se crea una consulta, el editor de SQL valida los nombres de puerto de la consulta. También comprueba que los puertos utilizados para la sustitución de cadenas son tipos de datos String. El editor de SQL no valida la sintaxis de la consulta SQL.

Puede utilizar constantes en la consulta SQL. Escriba las cadenas entre comillas simples (').

En la siguiente imagen se muestra un ejemplo de consulta SQL:



Puede crear una consulta SQL estática. La instrucción de la consulta no cambia, aunque puede incluir parámetros para cambiar valores. El servicio de integración de datos ejecuta la consulta para cada fila de entrada.

Definir la consulta SQL

Define una consulta SQL que ejecuta las mismas instrucciones de consulta para cada fila de entrada. Puede cambiar las columnas de la consulta o de la tabla según los valores de puerto de entrada en la fila. También puede cambiar los valores de la cláusula WHERE según los valores de puerto de entrada.

Para cambiar los valores de los datos en la cláusula WHERE para cada fila de entrada, configure el enlace de parámetros.

Para cambiar las columnas de la consulta o cambiar la tabla según los valores de puerto de entrada, utilice la sustitución de cadenas.

Enlace de parámetro

Para cambiar los datos de la consulta, configure los parámetros de la consulta y enlace estos con puertos de entrada de la transformación. Cuando enlace un parámetro con un puerto de entrada, identifique el puerto por su nombre en la consulta. El editor de SQL delimita el nombre del puerto mediante signos de interrogación (?). Los datos de la consulta cambian según el valor de los datos en el puerto.

Las siguientes consultas utilizan enlaces de parámetro:

```
DELETE FROM Employee WHERE Dept = ?Dept?
INSERT INTO Employee(Employee_ID, Dept) VALUES (?Employee_ID?, ?Dept?)
UPDATE Employee SET Dept = ?Dept? WHERE Employee_ID > 100
```

La siguiente consulta SQL tiene parámetros de consulta que se enlazan con los puertos de entrada Employee_ID y Dept de una transformación de SQL:

```
SELECT Name, Address FROM Employees WHERE Employee_Num =?Employee_ID? and Dept = ?Dept?
```

El origen puede tener las siguientes filas:

Employee_ID	Dept
100	Products
123	HR
130	Accounting

El servicio de integración de datos genera las siguientes instrucciones de consulta a partir de las filas:

```
SELECT Name, Address FROM Employees WHERE Employee_ID = '100' and DEPT = 'Products'  
SELECT Name, Address FROM Employees WHERE Employee_ID = '123' and DEPT = 'HR'  
SELECT Name, Address FROM Employees WHERE Employee_ID = '130' and DEPT = 'Accounting'
```

Sustitución de cadenas

Utilice variables de cadena para reemplazar componentes de instrucciones de consulta. Por ejemplo, puede utilizar la variable de cadena para reemplazar el nombre de una tabla en una consulta. También puede sustituir los nombres de columna en una instrucción SELECT.

Para sustituir el nombre de una tabla, configure un puerto de entrada para que reciba el nombre de la tabla desde cada fila de entrada. En el editor de SQL, seleccione el puerto desde la lista de puertos **Sustitución de cadenas**. Developer Tool identifica el puerto de entrada por el nombre en la consulta y delimita el nombre con una tilde (~).

La siguiente consulta contiene una variable de cadena , ~Table_Port~:

```
SELECT Emp_ID, Address from ~Table_Port~ where Dept = 'HR'
```

El origen puede pasar los siguientes valores a la columna **Table_Port**:

Table_Port

Employees_USA

Employees_England

Employees_Australia

El servicio de integración de datos sustituye la variable ~Table_Port~ con el valor del nombre de tabla del puerto de entrada:

```
SELECT Emp_ID, Address from Employees_USA where Dept = 'HR'  
SELECT Emp_ID, Address from Employees_England where Dept = 'HR'  
SELECT Emp_ID, Address from Employees_Australia where Dept = 'HR'
```

Cardinalidad entre fila de entrada y fila de salida

Cuando el servicio de integración de datos ejecuta una consulta SELECT, la transformación de SQL devuelve una fila por cada fila que recupera. Cuando la consulta no recupera ningún dato, la transformación de SQL devuelve cero o una fila por cada fila de entrada.

Cómo procesar instrucciones de consultas

Cuando la consulta SELECT se ejecuta correctamente, la transformación de SQL puede recuperar varias filas. Cuando la consulta contiene otras instrucciones, el servicio de integración de datos puede generar una fila que contenga errores de SQL o el número de filas afectadas.

Configuración de puertos

El puerto de salida NumRowsAffected contiene el número de filas que la instrucción UPDATE, INSERT o DELETE cambia para una fila de entrada. La transformación de SQL devuelve el número de filas afectadas para cada instrucción de una consulta. Cuando la transformación de SQL contiene puertos de transferencia, la transformación devuelve los datos de columna al menos una vez para cada fila de origen.

Configuración del recuento máximo de filas

Recuento máximo de filas de salida limita el número de filas que una transformación de SQL devuelve a partir de consultas SELECT.

Filas de error

El servicio de integración de datos devuelve errores de fila cuando detecta errores de conexión o errores de sintaxis. La transformación de SQL devuelve errores al puerto SQLError.

Continuar ante error de SQL

Puede configurar la transformación de SQL para que continúe procesando cuando se produce un error en una instrucción SQL. La transformación de SQL no genera un error de fila.

Cómo procesar instrucciones de consultas

El tipo de consulta SQL determina cuántas filas devuelve la transformación SQL. La transformación SQL puede devolver ninguna, una o varias filas. Cuando la consulta contiene una instrucción SELECT, la transformación SQL devuelve cada columna de la base de datos en un puerto de salida. La transformación devuelve todas las filas elegibles.

La siguiente tabla enumera las filas de salida que genera la transformación SQL para distintos tipos de instrucciones de consulta cuando no se produce ningún error en el modo de consulta:

Instrucción de consulta	Filas de salida
UPDATE, INSERT, sólo DELETE	Una fila por cada instrucción de la consulta.
Una o más instrucciones SELECT.	Número total de filas recuperadas de la base de datos.
Consultas DDL como CREATE, DROP, TRUNCATE	Una fila por cada instrucción de la consulta.

Configuración de puertos

Cuando se habilita Incluir estadísticas como salida, Developer Tool crea el puerto NumRowsAffected. El servicio de integración de datos devuelve, como mínimo, una fila con NumRowsAffected según las instrucciones de la consulta SQL.

La siguiente tabla enumera las filas de salida que genera la transformación de SQL si se habilita NumRowsAffected:

Instrucción de consulta	Filas de salida
UPDATE, INSERT, sólo DELETE	Una fila por cada instrucción con NumRowsAffected para la instrucción.
Una o más instrucciones SELECT.	Número total de filas recuperadas de la base de datos. NumRowsAffected es cero en cada fila.
Consultas DDL como CREATE, DROP, TRUNCATE	Una fila con ningún NumRowsAffected.

Recuento máximo de filas de salida

Se puede limitar el número de filas que la transformación de SQL devuelve para consultas SELECT. Configure la propiedad **Recuento máximo de filas de salida** para limitar el número de filas. Cuando una consulta contiene varias instrucciones SELECT, la transformación de SQL limita el número total de filas de todas las instrucciones SELECT.

Por ejemplo, establezca **Recuento máximo de filas de salida** en 100. La consulta contiene dos instrucciones SELECT:

```
SELECT * FROM table1; SELECT * FROM table2;
```

Si la primera instrucción SELECT devuelve 200 filas, y la segunda instrucción SELECT devuelve 50 filas, la transformación de SQL devuelve 100 filas de la primera instrucción SELECT. La transformación de SQL no devuelve ninguna fila de la segunda instrucción.

Para configurar un número ilimitado de filas, establezca en cero el valor de Recuento máximo de filas de salida.

Filas de error

El servicio de integración de datos devuelve errores de fila cuando detecta un error de conexión o un error de sintaxis. La transformación de SQL devuelve errores de SQL al puerto SQLException.

Cuando se configura un puerto de transferencia o el puerto NumRowsAffected, la transformación de SQL devuelve como mínimo una fila por cada fila de origen. Cuando una consulta no devuelve datos, la transformación de SQL devuelve los datos de transferencia y los valores de NumRowsAffected, pero devuelve valores nulos en los puertos de salida. Se pueden quitar filas con valores nulos haciendo pasar las filas de salida a través de una transformación de filtro.

La siguiente tabla describe las filas que genera la transformación de SQL para las instrucciones de consulta UPDATE, INSERT o DELETE:

Puerto NumRowsAffected o puerto de transferencia configurado	SQLError	Salida de filas
Ningún puerto configurado	No	Una fila con NULL en el puerto SQLError.
Ningún puerto configurado	Sí	Una fila con error en el puerto SQLError.
Cualquiera de los dos puertos configurados	No	Una fila por cada instrucción de consulta con los datos de columna de NumRowsAffected o del puerto de transferencia.
Cualquiera de los dos puertos configurados	Sí	Una fila con el error en el puerto SQLError, el puerto NumRowsAffected o los datos del puerto de transferencia.

La siguiente tabla describe el número de filas de salida que la transformación de SQL genera para las instrucciones SELECT:

Puerto NumRowsAffected o puerto de transferencia configurado	SQLError	Salida de filas
Ningún puerto configurado	No	Una o más filas, según las filas devueltas desde cada instrucción SELECT.
Ningún puerto configurado	Sí	Una fila mayor que la suma de filas de salida para las instrucciones correctas. La última fila contiene el error en el puerto SQLError.
Cualquiera de los dos puertos configurados	No	Una o más filas, según las filas devueltas para cada instrucción SELECT: <ul style="list-style-type: none"> - Si NumRowsAffected está habilitado, cada fila contiene una columna NumRowsAffected con un valor cero. - Si se configura un puerto de transferencia, cada fila contiene los datos de la columna de transferencia. Cuando la consulta devuelve varias filas, los datos de la columna de transferencia se duplican en cada fila.
Cualquiera de los dos puertos configurados	Sí	Una o más filas, según las filas devueltas para cada instrucción SELECT. La última fila contiene el error en el puerto SQLError: <ul style="list-style-type: none"> - Cuando NumRowsAffected está habilitado, cada fila contiene una columna NumRowsAffected con un valor cero. - Si se configura un puerto de transferencia, cada fila contiene los datos de la columna de transferencia. Cuando la consulta devuelve varias filas, los datos de la columna de transferencia se duplican en cada fila.

La siguiente tabla describe el número de filas de salida que genera la transformación de SQL para las consultas DDL como CREATE, DROP o TRUNCATE:

Puerto NumRowsAffected o puerto de transferencia configurado	SQLError	Salida de filas
Ningún puerto configurado	No	Una fila con NULL en el puerto SQLError.
Ningún puerto configurado	Sí	Una fila que contiene el error en el puerto SQLError.

Puerto NumRowsAffected o puerto de transferencia configurado	SQLError	Salida de filas
Cualquiera de los dos puertos configurados	No	Una fila que incluye la columna NumRowsAffected con valor cero y los datos de columna de transferencia.
Cualquiera de los dos puertos configurados	Sí	Una fila con el error en el puerto SQLError, el puerto NumRowsAffected con valor cero y los datos de la columna de transferencia.

Continuar ante error de SQL

Puede optar por omitir un error SQL que se produce en una instrucción de consulta. Habilite **Continuar ante error de SQL si hay errores en una fila**. El servicio de integración de datos continúa ejecutando las instrucciones restantes para la fila.

El servicio de integración de datos no genera un error de fila. Sin embargo, el puerto SQLError contiene la instrucción SQL que falló y mensajes de error.

Por ejemplo, una consulta puede tener las siguientes instrucciones:

```
DELETE FROM Persons WHERE FirstName = 'Ed';
INSERT INTO Persons (LastName, Address) VALUES ('Gein', '38 Beach Rd')
```

Si la instrucción DELETE falla, la transformación de SQL devuelve un mensaje de error desde la base de datos. El servicio de integración de datos continúa procesando la instrucción INSERT.

Deshabilite la opción **Continuar ante error de SQL** para solucionar problemas de la base de datos y asociar los errores con las instrucciones de consulta que han generado los errores.

Optimización del filtro con la transformación de SQL

El servicio de integración de datos puede aplicar una optimización del filtro con una transformación de SQL si la condición de filtro hace referencia solo a los puertos de transferencia y la transformación de SQL no tiene efectos secundarios.

La transformación de SQL tiene efectos secundarios en las siguientes circunstancias:

- La consulta SQL actualiza una base de datos. La consulta SQL contiene una instrucción como CREATE, DROP, INSERT, UPDATE, GRANT o REVOKE.
- La transformación devuelve filas nulas para las instrucciones SELECT que no devuelven ningún resultado. Las filas pueden contener valores de puerto de transferencia, información de errores de SQL o el campo NumRowsAffected.

El servicio de integración de datos puede aplicar los métodos de optimización de primera selección e inserción con la transformación de SQL.

Optimización de primera selección con la transformación de SQL

El servicio de integración de datos puede realizar la optimización de primera selección con una transformación de SQL si la condición de filtro hace referencia solo a los puertos de transferencia y la transformación de SQL no tiene efectos secundarios.

La transformación de SQL tiene efectos secundarios en las siguientes circunstancias:

- La consulta SQL actualiza una base de datos. La consulta SQL contiene una instrucción como CREATE, DROP, INSERT, UPDATE, GRANT o REVOKE.
- La transformación devuelve filas nulas para las instrucciones SELECT que no devuelven ningún resultado. Las filas pueden contener valores de puerto de transferencia, información de errores de SQL o el campo NUMRowsAffected.

Habilitar la optimización de primera selección con la transformación de SQL

Habilite la optimización de primera selección en la transformación de SQL si la transformación de SQL no tiene ningún efecto secundario.

1. Habilite la opción **Devolver la base de datos de solo salida** en **Propiedades avanzadas** de la transformación de SQL.
2. Desactive **Tiene efectos secundarios** en **Propiedades avanzadas** de la transformación.
3. Si la transformación tiene un puerto **NumAffectedRows**, quite el puerto.

Optimización de inserción con la transformación de SQL

Con la optimización de inserción, el servicio de integración de datos inserta la lógica de filtro desde una transformación de filtro de la asignación a la consulta de la transformación de SQL.

Utilice las siguientes reglas y directrices cuando habilite la optimización de inserción con la transformación de SQL:

- La transformación de la consulta SQL solo debe contener instrucciones SELECT.
- La consulta SQL de la transformación debe ser una subconsulta válida.
- La condición de filtro no puede hacer referencia a los campos de Error de SQL o NumRowsAffected.
- Los nombres de los puertos de salida deben coincidir con los nombres de las columnas de la instrucción SQL SELECT. Cuando hace referencia a un puerto de salida en una condición de filtro, el servicio de integración de datos inserta el nombre del puerto correspondiente en la consulta SQL. Puede agregar alias a SQL si las columnas de la consulta no coinciden con los nombres de puertos de salida. Por ejemplo, `SELECT mycolname1 AS portname1, mycolname2 AS portname2`.
- La transformación no puede tener efectos secundarios.

Ejemplo de la optimización de inserción con la transformación SQL

Una transformación de SQL recupera pedidos por el ID del cliente. Una transformación de filtro que aparece tras la transformación de SQL solamente devuelve las filas en las que el importe del pedido es superior a 1.000.

El servicio de integración de datos inserta el siguiente filtro en una instrucción SELECT en la transformación de SQL:

```
orderAmount > 1000
```

Cada instrucción de la consulta SQL se convierte en otra subconsulta de la instrucción SELECT que contiene el filtro.

La siguiente instrucción de consulta muestra la instrucción de consulta original como una subconsulta de la instrucción SELECT:

```
SELECT <customerID>, <orderAmount>, ... FROM (original query statements) ALIAS WHERE  
<orderAmount> > 1000
```

Si la consulta SQL tiene varias instrucciones, cada instrucción está incluida en otra subconsulta. La subconsulta tiene la misma sintaxis, incluyendo la cláusula WHERE.

Los puertos *customerID* y *orderAmount* son los nombres de los puertos de salida en la transformación de SQL. La subconsulta no incluye puertos de transferencia, el error de SQL o los puertos de estadísticas de SQL. Si inserta varios filtros en la transformación de SQL, la cláusula WHERE contiene todos los filtros.

Habilitar la optimización de inserción con la transformación de SQL

Habilitar la optimización de inserción mediante las propiedades de configuración en la ficha **Propiedades avanzadas** de la transformación de SQL.

1. Desactive **Tiene efectos secundarios**.
2. Habilite **Devolver la base de datos de solo salida**.
3. Establezca el **Recuento máximo de filas de salida** a cero.
4. Habilitar la optimización de inserción.

Ejemplo de transformación de SQL con una consulta SQL

Usted es un desarrollador del departamento de RRHH de la empresa Hypostores. Hypostores mantiene la información de las nóminas de los empleados en una base de datos aparte de los datos de empleados de recursos humanos. El departamento de RRHH necesita consultar una vista individual de los empleados y los salarios en varias regiones.

Usted quiere crear una asignación de objeto de datos lógicos que muestre una vista individual de los datos de empleados y de salarios en un objeto de datos lógicos de empleados.

Cree una asignación de objeto de datos lógicos con el origen de datos de empleados. Incluya una transformación de SQL para recuperar el salario y la fecha de alta laboral desde la base de datos de nóminas.

Asignación de objeto de datos lógicos

La asignación de objeto de datos lógicos contiene los siguientes objetos:

Tabla Empleado

Tabla relacional de entrada de la base de datos Recursos Humanos.

Tabla Salario

Una tabla de la base de datos Nómina que contiene el salario y la fecha de alta laboral del empleado. La base de datos es una base de datos Oracle.

transformación de SQL

Transformación que recupera la fecha de alta laboral y el salario para cada fila de empleado. La transformación se conecta a una base de datos Nómina y ejecuta una consulta SQL en la tabla Salario de la base de datos.

Objeto de datos lógicos

Contiene la vista combinada de los datos del empleado y del salario. El objeto de datos lógicos recibe la salida desde la transformación de SQL.

Archivo SQLErrors

El archivo SQLErrors es un archivo sin formato que contiene cualquier error SQL que se haya producido en la base de datos. El servicio de integración de datos escribe, como mínimo, una fila en el archivo SQLErrors por cada fila de entrada. Si no se produce ningún error SQL, la columna SQLError contiene NULL. Revise el archivo SQLErrors para solucionar los errores.

Tabla Salario

La tabla Salario es una tabla relacional de la base de datos Nómina. La tabla contiene datos de empleados que mantiene el departamento de nóminas. La transformación de SQL recupera la fecha de alta laboral y el salario del empleado de la tabla Salario.

La siguiente tabla muestra algunas filas de la tabla Salario:

Employee_Num	HireDate	Salary
10	3-May-97	232000
11	11-Sep-01	444000
12	17-Oct-89	656000
13	13-Ago-07	332100

Tabla Empleado

El origen es la tabla Empleado de la base de datos Recursos Humanos.

La siguiente tabla muestra algunas filas de la tabla Empleado:

EmpID	LastName	FirstName	DeptId	Teléfono
10	Smith	Martha	FIN	(415) 552-1623
11	Jones	Cynthia	ENG	(415) 552-1744
12	Russell	Cissy	SLS	(415) 552-1656
13	Goyal	Girish	FIN	(415) 552-1656

Ejemplo de transformación de SQL

La transformación de SQL recupera la fecha de alta laboral y el salario del empleado desde la tabla Salario de la base de datos Nómina. La tabla Salario es una base de datos Oracle.

Utilice los siguientes pasos para configurar la transformación de SQL:

1. Configure las propiedades de la transformación de SQL.
2. Defina los puertos.
3. Cree la consulta SQL.
4. Configure la conexión de base de datos para la transformación de SQL.

Definir propiedades de transformación de SQL

Configure las propiedades de la transformación de SQL en la vista **Propiedades avanzadas**.

Configure las siguientes propiedades:

Tipo de base de datos

El tipo de base de datos es Oracle. En el momento de definir los puertos, puede elegir tipos de datos de puerto aplicables a Oracle.

Continuar si hay errores en una fila

Deshabilitar. Detener el proceso si se produce un error SQL en la fila.

Incluir estadísticas como salida

Deshabilitar. No cree el puerto de salida NumRowsAffected.

Definir los puertos

Defina puertos de entrada para cada columna de la tabla de origen de empleados. Seleccione **Copiar a salida** para cambiar los puertos de transferencia de las columnas. Cuando se selecciona **Copiar a salida**, Developer Tool crea el puerto de salida correspondiente para cada puerto que se copia.

Cree los siguientes puertos de transferencia de entrada:

Nombre	Tipo	Tipo nativo	Precisión	Escala	Copiar a salida
EmpID	decimal	número(p,2)	4	0	x
LastName	string	varchar2	30	0	x
FirstName	string	varchar2	20	0	x
DeptID	string	varchar2	4	0	x
Teléfono	string	varchar2	16	0	x

La transformación de expresión tiene los siguientes puertos:

Nombre	Tipo	Tipo nativo	Precisión	Escala
EmpID	decimal	number(p,s)	4	0
LastName	string	varchar2	30	0
FirstName	string	varchar2	20	0
DeptID	string	varchar2	4	0

Nombre	Tipo	Tipo nativo	Precisión	Escala
Teléfono	string	varchar2	16	0
HireDate	date/time	marca de tiempo	29	0
Salary	decimal	number(p,s)	8	2

Developer Tool añade el sufijo "_output" a cada puerto de salida que crea cuando se selecciona **Copiar a salida**.

Defina manualmente los puertos de salida para las columnas de fecha de alta laboral y de salario. La transformación de SQL devuelve las columnas de fecha de alta laboral y de salario desde la tabla Salario en los puertos.

Definir la consulta SQL

Cree una consulta SQL para seleccionar la fecha de alta laboral y el salario de cada empleado de la tabla Salario.

Defina la consulta en la vista SQL de la transformación de SQL.

Escriba la consulta siguiente en el editor de SQL:

```
select HIREDATE,SALARY,from Salary where EMPLOYEE_NUM =?EmpID?
```

Hiredate, Salary y Employee_Num son nombres de columna de la tabla Salario.

?EMPID? es un parámetro que contiene el valor del puerto EmpID.

Definir la conexión de la base de datos

En la vista **Tiempo de ejecución**, seleccione un objeto de conexión de base de datos para la base de datos a la que se conecta la transformación de SQL. Seleccione un objeto de conexión de base de datos Oracle.

Salida

Conecte el puerto SQLError y el puerto EmpID_output al archivo sin formato SQLErrors. El puerto SQLError contiene valores nulos salvo que se produzca un error SQL.

Conecte EmpID y los otros puertos de salida al objeto de datos lógicos.

La transformación SQL devuelve una fila que contiene datos de la tabla Empleado e incluye la fecha de alta y el salario de la tabla Salario.

La siguiente tabla muestra algunas filas del objeto de datos lógicos:

EmpID	LastName	FirstName	DeptId	Teléfono	HireDate	Salary
10	Smith	Martha	FIN	(415) 552-1623	19970303 00:00:00	2320.00
11	Jones	Cynthia	ENG	(415) 552-1744	20010911 00:00:00	4440.00
12	Russell	Cissy	SLS	(415) 552-1656	19891017 00:00:00	6560.00
13	Goyal	Girish	FIN	(415) 552-1660	20070813 00:00:00	3210.00

Procedimientos almacenados

Se puede llamar a un procedimiento almacenado desde una transformación de SQL. Puede utilizar un procedimiento almacenado para automatizar tareas en una base de datos relacional. Los procedimientos almacenados aceptan variables definidas por el usuario, instrucciones condicionales y otras características que las instrucciones SQL estándar no admiten.

La transformación de SQL se conecta a una base de datos relacional para ejecutar el procedimiento almacenado. La transformación de SQL puede llamar a los procedimientos almacenados desde Oracle, IBM DB2, Microsoft SQL Server, Sybase y ODBC. Un procedimiento almacenado se conserva en la base de datos y se ejecuta en la base de datos.

Cree una conexión ODBC para llamar a un procedimiento almacenado desde una base de datos Sybase. También debe crear una conexión ODBC para llamar a un procedimiento almacenado desde una base de datos Microsoft SQL Server en sistemas operativos que no sean Windows.

Un procedimiento almacenado es un conjunto precompilado de instrucciones de procedimientos de Transact-SQL, PL-SQL o de otras bases de datos. La sintaxis del procedimiento almacenado varía según la base de datos.

Puede utilizar los procedimientos almacenados para completar las tareas siguientes:

- Comprobar el estado de una base de datos de destino antes de cargar datos en ella.
- Determinar si existe suficiente espacio en una base de datos.
- Realizar un cálculo especializado.
- Recuperar datos por un valor.
- Quitar y volver a crear índices.

Puede utilizar un procedimiento almacenado para realizar una consulta o un cálculo que en otro caso debería incluir en una transformación. Por ejemplo, si dispone de un procedimiento almacenado adecuadamente probado para calcular el impuesto de venta, puede realizar dicho cálculo con el procedimiento almacenado en lugar de la posibilidad de volver a crear el mismo cálculo en una transformación de expresión.

Un procedimiento almacenado puede aceptar una entrada y, a continuación, devolver un conjunto de resultados de filas. Un procedimiento almacenado puede ejecutar una tarea DDL que no requiera ninguna entrada y que después no devuelva ninguna salida.

Puede configurar la transformación de SQL para ejecutar más de un procedimiento almacenado. Para cada procedimiento almacenado que configure, configure puertos de transformación para que coincidan con los parámetros del procedimiento almacenado. Cada procedimiento almacenado puede pasar datos a puertos de salida.

La base de datos que contiene el procedimiento almacenado tiene permisos de usuario. Debe tener permisos para ejecutar el procedimiento almacenado en la base de datos.

Nota: Una función almacenada es similar a un procedimiento almacenado, excepto que una función devuelve un valor individual. La transformación de SQL puede ejecutar funciones almacenadas.

Puertos de transformación de SQL para procedimientos almacenados

Los puertos de entrada y de salida de la transformación de SQL corresponden a los parámetros de entrada y salida en el procedimiento almacenado.

Cuando importe un procedimiento almacenado, la herramienta Developer determina el tipo de base de datos desde la conexión de base de datos. Genera puertos de entrada y salida en la transformación de SQL desde

los parámetros en el procedimiento almacenado. La herramienta Developer determina el tipo de datos nativo de cada puerto desde el parámetro del procedimiento almacenado.

Cuando configura manualmente la transformación de SQL, debe configurar los puertos de entrada y de salida en la transformación. Cuando configure el tipo de base de datos, la transformación de SQL cambia el tipo de datos nativo para cada puerto en función del tipo de base de datos que especifique.

Los siguientes tipos de datos puede pasar entre la transformación de SQL y el procedimiento almacenado:

Parámetros de entrada y salida

La transformación de SQL envía parámetros al procedimiento almacenado y la transformación de SQL recibe los parámetros del procedimiento almacenado en los puertos de entrada y de salida.

Valor de devolución

Si el procedimiento almacenado pasa un valor de devolución, la herramienta Developer crea un puerto Valor de devolución.

Errores de SQL

La transformación de SQL devuelve errores del procedimiento almacenado en el puerto SQLError.

Parámetros de entrada y salida

Cuando se llama a un procedimiento almacenado desde una transformación de SQL, todos los campos referenciados por la instrucción CALL identifican un puerto de entrada o de salida. Cuando importe un procedimiento almacenado, la herramienta Developer genera la instrucción de llamada al procedimiento almacenado. De lo contrario, deberá configurar manualmente la instrucción CALL.

Puede editar la instrucción CALL en la vista **SQL** de la transformación.

La instrucción CALL tiene el siguiente formato:

```
?RETURN_VALUE? = call <nombre proc almacenado>(?Campo1?, ?Campo2?, . . . )
```

Escriba los nombres de los puertos entre signos de interrogación. Los nombres de los puertos no tienen que coincidir con los nombres de los parámetros en el procedimiento almacenado. Los puertos de salida deben estar en el mismo que los parámetros en una consulta SELECT.

Puede utilizar un procedimiento almacenado que contiene parámetros INOUT. La transformación de SQL identifica los parámetros INOUT por el nombre del puerto de entrada. El puerto de salida tiene el prefijo `output_`. El servicio de integración de datos enlaza el puerto de entrada y el puerto de salida al mismo parámetro.

Puede configurar una transformación de SQL para devolver un conjunto de resultados. Cuando el procedimiento almacenado devuelve un conjunto de resultados, la herramienta Developer no puede crear puertos de salida para las columnas en el conjunto de resultados. Cuando importe el procedimiento almacenado, debe especificar manualmente los puertos y configurar la llamada al procedimiento almacenado.

Valor de devolución

Un valor de devolución es un código o cadena de texto que define ese estado del procedimiento almacenado. Cuando el procedimiento almacenado tiene un valor de devolución, la transformación de SQL tiene un puerto **Valor de devolución**.

La mayoría de bases de datos puede pasar un valor de devolución después de ejecutar un procedimiento almacenado. El valor de devolución puede contener un valor entero, o puede contener un valor que se defina en el procedimiento almacenado. Por ejemplo, un procedimiento almacenado podría devolver "Correcto" cuando el procedimiento se realiza correctamente.

Si un procedimiento almacenado devuelve un conjunto de resultados en lugar de un solo valor de devolución, la transformación de SQL recibe el primer valor de devolución del procedimiento.

Conjuntos de resultados de un procedimiento almacenado

Puede configurar una transformación de SQL para recibir un conjunto de resultados de un procedimiento almacenado. Un procedimiento almacenado devuelve varias filas en un conjunto de resultados. La transformación de SQL puede devolver cada fila a la asignación.

El siguiente procedimiento almacenado devuelve un conjunto de resultados:

```
CREATE OR REPLACE FUNCTION fetchEMPinfo
(p_State IN VARCHAR2 )
return types.cursorType
AS
my_cursor types.cursorType;
BEGIN
OPEN my_cursor FOR SELECT EMP_ID, NAME, CITY FROM EMP WHERE STATE = p_State ORDER BY
EMP_ID;
RETURN my_cursor;
END;
```

Cuando importa el procedimiento almacenado, la herramienta Developer crea una instrucción de llamada al procedimiento almacenado con una sintaxis similar a esta:

```
call FETCHEMPINFO (?P_STATE?)
```

El parámetro de entrada es `p_state`. La herramienta Developer no crea los puertos de salida de forma automática. Debe crear manualmente los puertos de salida con los mismos tipos de datos que los parámetros del procedimiento almacenado.

Por ejemplo, el conjunto de resultados `resultSet` contiene las columnas `EMP_ID`, `EMPNAME` y `CITY`. Cree puertos de salida para estas columnas.

También debe actualizar manualmente la llamada SQL con las columnas de salida utilizando la siguiente sintaxis:

```
(?EMP_ID?,?EMPNAME?,?CITY?) = call FETCHEMPINFO (?P_STATE?)
```

Conjuntos de resultados con diferentes bases de datos

Configure los procedimientos almacenados para devolver los conjuntos de resultados utilizando diferentes sintaxis en función del tipo de base de datos.

Oracle

Una función almacenada de Oracle devuelve resultados con un cursor:

```
create or replace function sp_ListEmp return types.cursorType
as
l_cursor types.cursorType;
begin
open l_cursor for select ename, empno from emp order by ename;
return l_cursor;
end;
```

Oracle también acepta cursores como parámetros de entrada. No puede configurar cursores como parámetros de entrada con la transformación de SQL.

Microsoft SQL Server

Un proceso almacenado de Microsoft SQL Server devuelve un procedimiento almacenado del conjunto de resultados con una instrucción SELECT en el cuerpo del procedimiento o con el tipo de devolución declarado de forma explícita como una tabla.

```
CREATE PROCEDURE InOut(
  @inout varchar(100) OUT
)
AS
BEGIN
  set @inout = concat(@inout, '__')
  select * from mytable;
END
```

IBM DB2

Un procedimiento almacenado de IBM DB2 puede devolver un conjunto de resultados con un cursor abierto. El número de conjuntos de resultados que devuelve están declarados en la cláusula RESULT SET. El procedimiento almacenado abre un cursor y lo devuelve. El siguiente ejemplo devuelve 2 cursores abiertos.

```
CREATE PROCEDURE TESTMULTIRS
  (IN i_cmacct CHARACTER(5))
  RESULT SETS 2
  LANGUAGE SQL
  BEGIN

  DECLARE csnum INTEGER;

  --Declare serial cursors to consume less resources
  --You do not need a rollable cursor.

  DECLARE getDeptNo CHAR(50); --Be careful with the estimated length.
  DECLARE getDeptName CHAR(200);
  DECLARE c1 CURSOR WITH RETURN FOR s1;
  SET getDeptNo = 'SELECT DEPTNO FROM DEPT';
  SET getDeptName = 'SELECT DEPTNAME FROM DEPT';

  PREPARE s1 FROM getDeptNo;
  OPEN c1;

  END;
```

Sybase

Un proceso almacenado de Sybase devuelve un procedimiento almacenado del conjunto de resultados con una instrucción SELECT en el cuerpo del procedimiento o con el tipo de devolución declarado de forma explícita como una tabla.

```
CREATE PROCEDURE FETCHEMPINFO
(
  @p_State VARCHAR(5),
  @e_id INT OUTPUT,
  @e_name VARCHAR(50) OUTPUT
)
AS
BEGIN
  SET NOCOUNT ON
  SELECT EMP_ID, NAME FROM EMP WHERE STATE = @p_State ORDER BY EMP_ID
  SET NOCOUNT OFF
  SELECT @e_id AS EMP_ID, @e_name AS NAME
  RETURN
END
GO

--Configure the following variables to execute the procedure.

DECLARE @p_State VARCHAR(5)
DECLARE @EMPID int
```



```

DECLARE @EMPNAME varchar(50)

SET @p_State = 'CA'
exec FETCHEMPINFO @p_State, @e_id = @EMPID, @e_name = @EMPNAME
GO

```

Filas del conjunto de resultados

Algunos de los procedimientos almacenados devuelven parámetros de salida además de las filas del conjunto de resultados. La transformación de SQL devuelve los parámetros de salida en la última fila. No incluye los parámetros de salida de ocurrencia individual en las filas del conjunto de resultados.

Por ejemplo, puede escribir un procedimiento almacenado que recibe un ID de empleado y devuelve el nombre del empleado en el parámetro de salida 1 y el departamento de parámetro de salida 2. El procedimiento almacenado también devuelve una fila por cada día de baja por enfermedad del empleado durante todo el año. La fila contiene la fecha, el número de horas y el motivo de la ausencia.

El conjunto de resultados contiene un número diferente de filas para cada empleado. Cada fila en el conjunto de resultados contiene un nombre de empleado y un departamento vacíos. La transformación de SQL devuelve el nombre del empleado y el departamento después del conjunto de resultados. El nombre del empleado y el departamento aparecen en la última fila.

Ejemplo de procedimiento almacenado

Se puede llamar a un procedimiento almacenado que devuelve datos a la transformación de SQL.

El siguiente procedimiento almacenado recibe un número de empleado y devuelve una fila con el número de empleado y el nombre del empleado:

```

CREATE OR REPLACE PROCEDURE SP_GETNAME
(IN_EMPNO IN NUMBER, OUT_EMPNO NUMBER, OUT_NAME OUT STRING)
AS
BEGIN
SELECT EMP_KEY,EMP_NAME into OUT_EMPNO , OUT_NAME from EMP_TABLE where EMP_KEY=IN_EMPNO;
END;/"

```

Para crear la transformación de SQL, importe el procedimiento almacenado. La herramienta Developer crea los puertos de entrada y los puertos de salida. Los nombres de puerto son los mismos que los nombres de los parámetros en el procedimiento almacenado.

La siguiente figura muestra los puertos para la transformación de SQL:

Name	Type	Native Type	Precision	Scale	Default Value	Description	Copy to Output
Input (1)							
1 IN_EMPNO	decimal	number(p,s)	38	0			<input type="checkbox"/>
Output (3)							
1 SQLError	string		4096	0			<input type="checkbox"/>
2 OUT_EMPNO	decimal	number(p,s)	38	0			<input type="checkbox"/>
3 OUT_NAME	string	varchar2	50	0			<input checked="" type="checkbox"/>

La herramienta Developer crea la siguiente llamada al procedimiento almacenado para recuperar el nombre del empleado:

```
call SP_GETNAME (?IN_EMPNO?, ?OUT_EMPNO?, ?OUT_NAME?)
```

Puede ver la llamada al procedimiento almacenado en el editor de SQL. Los errores SQL aparecen en el puerto SQLError.

Conexión de transformación de SQL

Configure la conexión de transformación de SQL en las propiedades de tiempo de ejecución de la transformación. Puede que necesite configurar una conexión de tiempo de ejecución si no ha especificado una conexión al crear la transformación.

Puede definir un nombre de conexión diferente a la conexión que ha seleccionado para crear la transformación de SQL. Debe seleccionar una conexión del mismo tipo de base de datos que el tipo de base de datos de las propiedades **Avanzadas** de la transformación de SQL.

Puede configurar un parámetro para el nombre de conexión de la transformación de SQL. Debe definir el parámetro en la vista **Parámetros** de la asignación antes de asignarlo a una conexión en tiempo de ejecución.

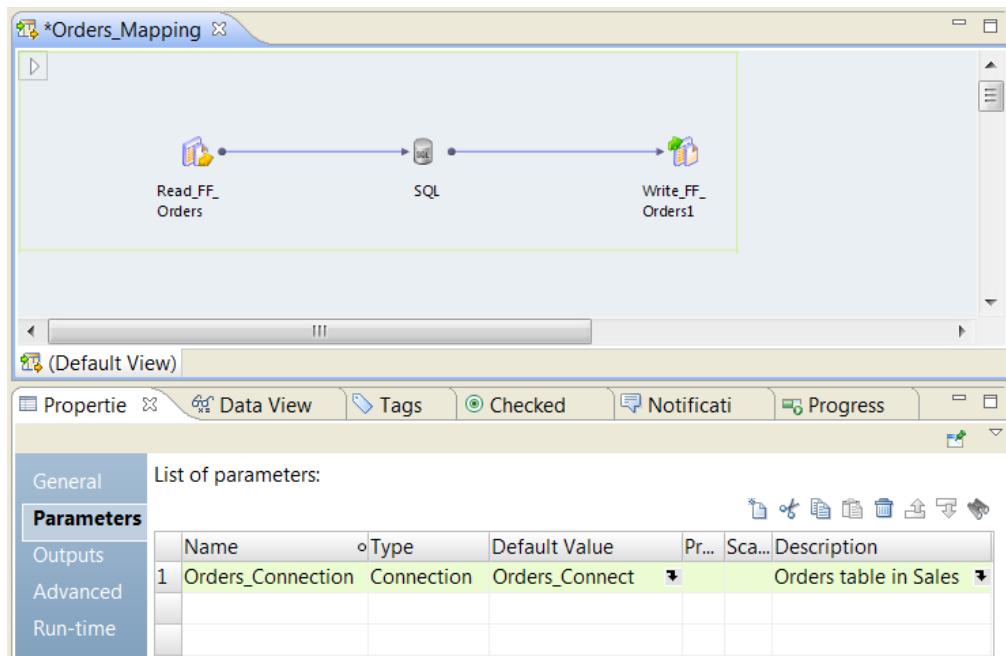
Creación de un parámetro de nombre de conexión

Puede especificar un parámetro definido por el usuario en el nombre de conexión en tiempo de ejecución de una transformación de SQL. Developer tool creará un parámetro de asignación para la conexión en lugar de un parámetro de transformación.

1. Cree una asignación que incluya la transformación de SQL.
2. Haga clic en la ficha **Tiempo de ejecución** de la transformación de SQL.
3. En **Nombre de la conexión**, haga clic en la flecha de selección y elija **Asignar parámetros**.
4. En el cuadro de diálogo **Asignar parámetro**, haga clic en **Nuevo**.
5. En el cuadro de diálogo **Parámetros**, introduzca un nombre de parámetro de conexión y una descripción del parámetro. El tipo de parámetro se establece de forma predeterminada en Conexión.
6. Haga clic en **Aceptar** en el cuadro de diálogo Parámetros y en el cuadro de diálogo **Asignar parámetros**.

Developer tool crea un parámetro de asignación y lo asigna al nombre de la conexión. El nombre de parámetro aparece en las propiedades en tiempo de ejecución.

7. Para ver una lista de los parámetros de asignación, haga clic dentro del editor y, a continuación, en la ficha **Parámetros** de la asignación.



Cómo crear manualmente una transformación de SQL

Puede crear manualmente una transformación de SQL. Cree manualmente la transformación cuando configure una transformación que se ejecuta una consulta SQL. También puede crear manualmente una transformación que llama a un procedimiento almacenado cuando el procedimiento no está disponible para su importación. Cuando crea manualmente la transformación, puede configurar los puertos de entrada y de salida, y el tipo de las instrucciones SQL en el editor de SQL.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nuevo > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de SQL.
4. Haga clic en **Siguiente**.
5. Seleccione **Crear como vacío**.
6. Especifique un nombre para la transformación y especifique la ubicación del repositorio para la transformación.
7. Haga clic en **Finalizar**.
8. Haga clic en la vista **Resumen** para añadir puertos a la transformación.
9. Para añadir un puerto de entrada, haga clic en **Entrada** en el panel **Puertos** para indicar dónde añadir el puerto. Haga clic en el botón **Nuevo** y especifique el nombre de puerto, el tipo nativo y la precisión.

El tipo de base de datos predeterminado es Oracle. La herramienta Developer muestra tipos nativos para las bases de datos Oracle a menos que cambie el tipo de base de datos en la vista **Avanzadas**.

10. Para añadir un puerto de salida, haga clic en **Salida** en el panel **Puertos** antes de añadir el puerto. Haga clic en el botón **Nuevo** y especifique el nombre de puerto, el tipo nativo y la precisión.

El puerto **SQLException** es, de forma predeterminada, el primer puerto de salida.

11. En la vista **Avanzada**, seleccione el tipo de base de datos al que se conecta la transformación de SQL. Configure otras propiedades avanzadas para el control de errores y otras propiedades opcionales.

Cuando selecciona el tipo de base de datos, la herramienta Developer cambia los tipos de datos nativos de los puertos en la vista **Resumen**.

12. Escriba la consulta SQL o la llamada al procedimiento almacenado en la vista **SQL**. Seleccione los puertos para el enlace de parámetros o la sustitución de cadenas en el **Editor de SQL**.

Si el procedimiento almacenado devuelve un conjunto de resultados, debe especificar la llamada al procedimiento almacenado con una sintaxis similar a la siguiente: (?Field1?,?Field2?,?Field3?) = `call Stored_Procedure_Name (?Input_Parm?)`.

TEMAS RELACIONADOS

- [“Definir la consulta SQL” en la página 668](#)

Creación de una transformación de SQL desde un procedimiento almacenado

Puede configurar una transformación de SQL importando un procedimiento almacenado desde una conexión de base de datos.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nueva**.
3. Seleccione la transformación de SQL.
4. Haga clic en **Siguiente**.
5. Seleccione **Crear desde un procedimiento almacenado existente**.
6. Busque y seleccione una conexión de base de datos.
7. Busque y seleccione el procedimiento almacenado para importar.
8. Especifique un nombre y una ubicación para la transformación.
9. Haga clic en **Finalizar**.
Developer Tool crea los puertos y la llamada al procedimiento almacenado.
10. Si el procedimiento almacenado devuelve un conjunto de resultados, debe añadir manualmente los puertos de salida y luego volver a configurar la llamada al procedimiento almacenado.
 - a. En la vista **Resumen**, haga clic en **Salida** en el panel **Puertos**. Haga clic en el botón **Nuevo** y especifique el nombre del puerto de salida, el tipo nativo y la precisión.

- b. En la vista **SQL**, cambie la llamada al procedimiento almacenado para usar la siguiente sintaxis: (? Field1?,?Field2?,?Field3?) = **call** Stored_Procedure_Name (?Input_Parm?)

Puede seleccionar los parámetros de entrada y de salida desde la lista de puertos **Enlace de parámetros** en el editor de SQL.

CAPÍTULO 44

Transformación de estandarizador

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de estandarizador, 678](#)
- [Estrategias de estandarización, 678](#)
- [Propiedades de estandarización, 679](#)
- [Cómo configurar una estrategia de estandarización, 680](#)
- [Propiedades avanzadas de Transformación de estandarizador, 680](#)
- [Transformación de estandarizador en un entorno no nativo, 681](#)

Resumen de la transformación de estandarizador

La transformación de estandarizador es una transformación pasiva que examina cadenas de entrada y crea versiones estandarizadas de las mismas.

La transformación de estandarizador crea columnas que contienen versiones estandarizadas de cadenas de entrada. La transformación puede reemplazar o quitar cadenas en los datos de entrada cuando se crean estas columnas.

Por ejemplo, puede usar la transformación de estandarizador para examinar una columna de datos de direcciones que contenga las cadenas Street, St. y STR. Puede reemplazar todas las instancias de estas cadenas con la cadena St.

Dentro de una transformación de estandarizador puede crear varias estrategias de estandarización. Cada estrategia puede contener varias operaciones de estandarización. La transformación de estandarización proporciona un asistente que puede usarse para crear estrategias.

Estrategias de estandarización

Utilice estrategias de estandarización para crear columnas con versiones estandarizadas de cadenas de entrada.

Cuando se crea una estrategia de estandarización, se añade una o más operaciones. Cada operación implementa una tarea de estandarización específica.

Se pueden añadir los siguientes tipos de operación a una estrategia de estandarización:

Reemplazar coincidencias de tabla de referencia por valores válidos

Reemplaza cadenas que coinciden con valores de tabla de referencia por el valor "Válido" de la tabla de referencia.

Reemplazar coincidencias de tabla de referencia por cadenas personalizadas

Reemplaza cadenas que coinciden con valores de tabla de referencia por una cadena de reemplazo definida por el usuario.

Quitar coincidencias de tabla de referencia

Quita las cadenas que coinciden con valores de tabla de referencia.

Reemplazar cadenas personalizadas

Reemplaza cadenas definidas por el usuario por una cadena de reemplazo definida por el usuario.

Quitar cadenas personalizadas

Quita cadenas definidas por el usuario

Importante: Se puede modificar el orden de las operaciones. El orden de las operaciones puede cambiar la salida de una estrategia, puesto que cada operación lee los resultados de la operación precedente.

Propiedades de estandarización

Para configurar las propiedades de estrategias de estandarización y de operaciones, seleccione la vista **Estrategias** en la transformación de estandarizador.

Propiedades de la estrategia

Las propiedades de la estrategia se aplican a todas las operaciones de la estrategia. Se pueden configurar las siguientes propiedades de estrategia:

Quitar varios espacios

Reemplaza varios espacios consecutivos con un espacio.

Quitar espacios iniciales y finales

Quita espacios al inicio y final de las cadenas de datos.

Delimitadores

Determina los delimitadores que definen los tokens de búsqueda. Por ejemplo, si elige "Punto y coma," la transformación de estandarizador busca en la cadena "naranjas;manzanas" y encuentra las cadenas "naranjas" y "manzanas". La transformación utiliza el delimitador de espacios de forma predeterminada.

Propiedades de la operación

Puede configurar propiedades para los siguientes tipos de operación de estandarización:

Operaciones de tabla de referencia

Las operaciones de tabla de referencia incluyen las siguientes propiedades:

- **Tabla de referencia.** Determina la tabla de referencia que se usa para estandarizar los datos. Haga clic en **Explorar** para seleccionar una tabla de referencia.
- **Distinguir mayúsculas de minúsculas** Determina si el formato de mayúsculas y minúsculas de las cadenas de entrada debe coincidir con el de las entradas de la tabla de referencia.

- **Sustituir por.** Reemplaza las cadenas de entrada que coinciden con entradas de la tabla de referencia con el texto especificado. Solamente se aplica a operaciones de reemplazo.
- **Ámbito.** Especifica la parte de la cadena de entrada que contiene el valor de la tabla de referencia.

Operaciones con cadenas personalizadas

Las operaciones con cadenas personalizadas incluyen las siguientes propiedades:

- **Coincidencia de tokens con.** Define las cadenas de búsqueda que encontrar entre los datos de entrada.
- **Sustituir por.** Sustituye las cadenas de entrada que coinciden con las cadenas de búsqueda especificadas. Solamente se aplica a operaciones de reemplazo.
- **Ámbito.** Especifica la parte de la cadena de entrada que buscar.

Cómo configurar una estrategia de estandarización

Para configurar una estrategia de estandarización, edite los ajustes en la vista **Estrategia** de una transformación de estandarizador.

1. Seleccione la vista **Estrategias**.
2. Haga clic en **Nuevo**.
Se abre el asistente **Nueva estrategia**.
3. Haga clic en el campo **Entradas** para seleccionar los puertos para la estrategia.
4. Edite las propiedades de la estrategia y haga clic en **Siguiente**.
5. Elija una operación y haga clic en **Siguiente**.
6. Configure las propiedades de la operación.
7. Si lo desea, haga clic en **Siguiente** para añadir más operaciones a la estrategia.
8. Después de añadir todas las operaciones a la estrategia, haga clic en **Finalizar**.
9. Si lo desea, añada más estrategias a la transformación.
10. Si lo desea, cambie el orden en que la transformación procesa las estrategias o las operaciones. Seleccione una estrategia u operación y haga clic en **Mover hacia arriba** o **Mover hacia abajo**.

Propiedades avanzadas de Transformación de estandarizador

Configure las propiedades para ayudarle a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de estandarizador.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de estandarizador en un entorno no nativo

El procesamiento de transformaciones de estandarizador en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes y de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 45

Transformación de unión

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de unión, 682](#)
- [Grupos y puertos, 683](#)
- [Propiedades avanzadas de la transformación de unión, 683](#)
- [Cómo procesar una transformación de unión, 684](#)
- [Cómo crear una transformación de unión, 684](#)
- [Transformación de unión en un entorno no nativo, 685](#)

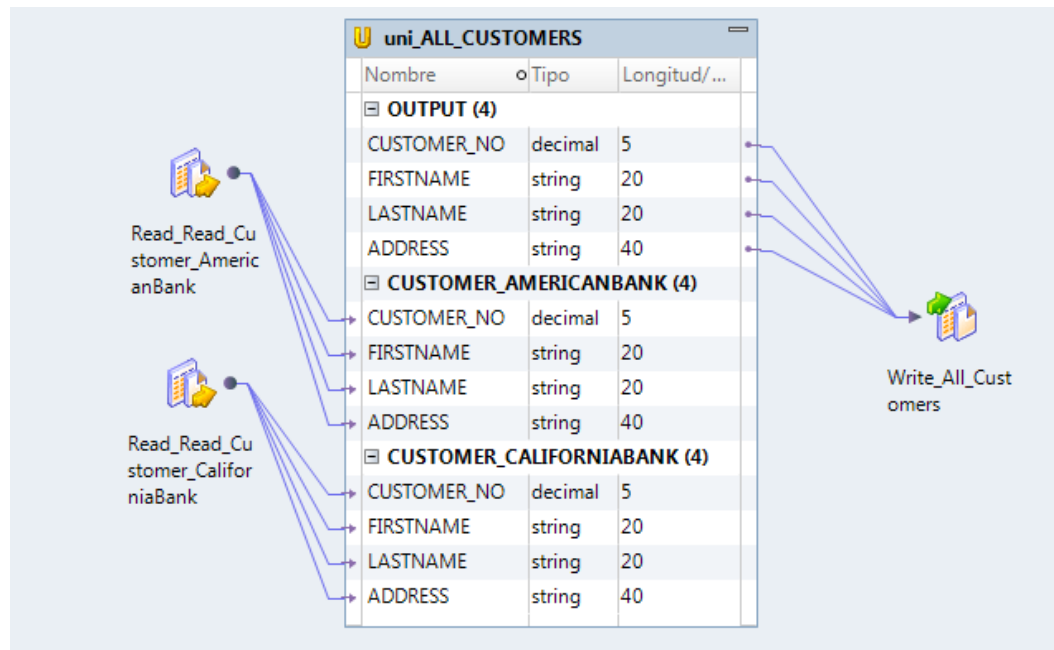
Resumen de la transformación de unión

Utilice la transformación de incorporación para fusionar datos de varios canales o ramificaciones de canales en una ramificación de canal.

La transformación de unión es una transformación activa con varios grupos de entrada y un grupo de salida. Fusiona orígenes con puertos coincidentes y pasa datos a través de un grupo de salida que tiene la misma estructura de puerto que los grupos de entrada. Utilice una transformación de unión en la herramienta Developer para fusionar datos de varios orígenes sin quitar las filas duplicadas.

Por ejemplo, puede combinar los datos de cuentas de clientes de los bancos American Bank y California Bank. Puede crear una asignación que contiene una transformación de unión para fusionar datos de los objetos de origen y escribir en un objeto de destino.

La siguiente figura muestra una asignación con una transformación de unión:



Grupos y puertos

Una transformación de unión tiene varios grupos de entrada y un grupo de salida. Puede crear uno o más grupos de entrada. Developer Tool crea un grupo de salida. No se puede crear, editar o eliminar el grupo de salida. Cada grupo debe tener puertos de coincidencia.

Para crear puertos, puede copiarlos desde una transformación o puede crearlos manualmente. Cuando se crean puertos, Developer Tool crea puertos de entrada en cada grupo de entrada y puertos de salida en el grupo de salida. Developer Tool utiliza los nombres de puerto que se especifican para cada puerto de entrada y de salida. Asimismo, Developer Tool utiliza los mismos metadatos para cada puerto, como el tipo de datos, la precisión y la escala.

Puede conectar los grupos de entrada de distintas ramificaciones en un único canal o desde distintos canales de origen. Cuando se añade una transformación de unión a una asignación, debe comprobar que se conectan los mismos puertos en todos los grupos de entrada. Si se conecta un puerto de un grupo de entrada pero no se conecta el mismo puerto de otro grupo de entrada, el servicio de integración de datos pasa NULL al puerto no conectado.

Propiedades avanzadas de la transformación de unión

Configure las propiedades para ayudarle a determinar cómo el servicio de integración de datos muestra los detalles de registro para la transformación de unión.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Cómo procesar una transformación de unión

Utilice la transformación de unión para fusionar datos de varios canales o ramificaciones de canales en una ramificación de canal. El servicio de integración de datos procesa todos los grupos de entrada en paralelo. Lee simultáneamente los orígenes conectados a la transformación de unión, y pasa bloques de datos a los grupos de entrada de la transformación. La transformación de unión procesa los bloques de datos según el orden en que recibe los bloques desde el servicio de integración de datos. La transformación de unión no bloquea datos de entrada en los grupos de entrada.

Cómo crear una transformación de unión

Puede crear una transformación de unión reutilizable o no reutilizable.

Cómo crear una transformación de unión reutilizable

Cree una transformación de unión reutilizable para utilizarla en varias asignaciones o mapplets.

1. Seleccione un proyecto o una carpeta en la vista **Explorador de objetos**.
2. Haga clic en **Archivo > Nueva > Transformación**.
A continuación, aparece el cuadro de diálogo **Nuevo**.
3. Seleccione la transformación de unión.
4. Haga clic en **Siguiente**.
5. Especifique un nombre para la transformación.
6. Haga clic en **Finalizar**.
La transformación aparece en el editor.
7. Haga clic en el botón **Nuevo** para añadir un puerto a la transformación.
8. Edite el puerto para definir el nombre, el tipo de datos y la precisión.
9. Seleccione la vista **Grupos**.
10. Haga clic en el botón **Nuevo** para añadir un grupo de entrada.
11. Haga clic en la vista **Avanzadas** y edite las propiedades de la transformación.

Cómo crear una transformación de unión no reutilizable

Cree una transformación de unión no reutilizable como parte de una asignación o de un mapplet.

1. En una asignación o en un mapplet, arrastre una transformación de unión desde la paleta Transformación hasta el editor.
La transformación aparece en el editor.
2. En la ficha **General**, edite el nombre y la descripción de la transformación.
3. Seleccione todos los puertos de la transformación de nivel superior y arrástrelos a la transformación de unión. Los puertos aparecen como puertos en un grupo de entrada y en un grupo de salida de la transformación de unión.
4. Haga clic en **Nueva**, en la ficha **Grupos** de la vista **Propiedades** para añadir un grupo de entrada.
Aparece otro grupo de entrada con puertos similares a los del grupo de entrada existente.
5. Seleccione los puertos en el grupo de salida de la transformación de unión y arrástrelos hasta la transformación de nivel inferior en la asignación.

Transformación de unión en un entorno no nativo

El procesamiento de transformaciones de unión en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. Compatible con restricciones en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

Transformación de unión en una asignación de transmisión

Las asignaciones de transmisión tienen restricciones de validación de asignaciones.

Validación de asignaciones

La validación de asignaciones genera un error en la siguiente situación:

- La transformación se ha configurado para fusionar datos procedentes de canales de transmisión y de no transmisión.

Transformación de unión en el motor de Databricks Spark

CAPÍTULO 46

Transformación de estrategia de actualización

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de estrategia de actualización, 686](#)
- [Actualizar transformaciones de estrategia en asignaciones dinámicas, 687](#)
- [Cómo marcar filas dentro de una asignación, 687](#)
- [Cómo especificar opciones de actualización para destinos individuales, 689](#)
- [Transformación de estrategia de actualización en un entorno no nativo, 690](#)

Resumen de la transformación de estrategia de actualización

La transformación de estrategia de actualización es una transformación activa que marca una fila para insertar, actualizar, suprimir o rechazar. Utilice una transformación de estrategia de actualización para controlar las modificaciones que se realicen en las filas existentes de un destino en función de una condición que usted aplique.

Como transformación activa, la transformación de estrategia de actualización puede modificar el número de filas que pasan a través de ella. La transformación de estrategia de actualización comprueba cada fila para ver si cumple una condición en concreto para, a continuación, marcar la fila según proceda. La transformación pasa a la siguiente transformación las filas que marca para insertar, actualizar o suprimir. Puede configurar la transformación para que pase a la siguiente transformación o para que anule las filas marcadas para rechazar.

Por ejemplo, puede usar la transformación de estrategia de actualización para marcar todas las filas de cliente para que se actualicen después de cambiar la dirección de envío de correo. También puede marcar para rechazar todas las filas de empleados que ya no trabajan para la organización.

Puede usar una transformación de estrategia de actualización para escribir resultados en un destino de base de datos relacional cuando la asignación se ejecuta en el motor de Spark. La asignación usa una cadena de conexión JDBC.

Cómo configurar la estrategia de actualización

Para definir una estrategia de actualización, lleve a cabo los siguientes pasos:

1. Para controlar cuántas filas se marcan para insertar, actualizar, suprimir o rechazar en una asignación, añada a la asignación una transformación de estrategia de actualización. Use una transformación de estrategia de actualización para marcar las filas destinadas al mismo fin en diferentes operaciones de bases de datos, o para rechazar filas.
2. Establezca opciones de inserción, actualización y supresión para destinos individuales al configurar la asignación. Destino por destino, puede permitir o prohibir las inserciones y las supresiones de todas las filas marcadas para insertar o suprimir. Puede elegir diferentes formas para gestionar las actualizaciones de todas las filas marcadas para actualizar.
3. Si la asignación se ejecuta en el motor en tiempo de ejecución de Spark, puede elegir Usar fusión de Hive en las propiedades avanzadas de la transformación de estrategia de actualización. Cuando una consulta utiliza una instrucción MERGE en vez de las instrucciones INSERT, UPDATE o DELETE, el procesamiento suele ser más eficaz.

Actualizar transformaciones de estrategia en asignaciones dinámicas

Puede utilizar una transformación de estrategia de actualización en una asignación dinámica. Puede configurar puertos dinámicos en la transformación y hacer referencia a puertos generados.

Puede hacer referencia a un puerto dinámico o a un puerto generado en la transformación de estrategia de actualización. Sin embargo, si el puerto generado no existe en tiempo de ejecución, la asignación fallará.

Si utiliza un puerto dinámico en la transformación de estrategia de actualización, ese puerto no puede tener más de un puerto generado.

Puede parametrizar la expresión de estrategia de actualización. Utilice un parámetro de tipo de expresión y escriba toda la expresión como el valor de parámetro predeterminado.

Cómo marcar filas dentro de una asignación

Añada una transformación de estrategia de actualización a una asignación para marcar filas para su inserción, actualización, supresión o rechazo.

Establezca una expresión de estrategia de actualización para comprobar cada fila y ver si cumple una condición en concreto. A continuación, asigne a cada fila un código numérico para marcar la fila para una operación de base de datos determinada.

En la siguiente tabla se muestran las constantes para cada operación de base de datos y su equivalente numérico:

Operación	Constante	Valor numérico
Insertar	DD_INSERT	0
Actualizar	DD_UPDATE	1
Eliminar	DD_DELETE	2
Rechazar	DD_REJECT	3

El servicio de integración de datos trata cualquier otro valor como una inserción.

Expresiones de estrategia de actualización

Introduzca una expresión de estrategia de actualización en el editor de expresiones.

La expresión de la estrategia de actualización utiliza la función IIF o DECODE del lenguaje de transformación para probar cada fila. Por ejemplo, la siguiente declaración IIF marca una fila para rechazarla si la fecha de entrada es posterior a la fecha de aplicación. De lo contrario, la declaración marca la fila para actualizarla:

```
IIF( ( ENTRY_DATE > APPLY_DATE), DD_REJECT, DD_UPDATE)
```

Puede configurar parámetros en una expresión de estrategia de actualización. Cree parámetros o búselos en el editor de expresiones.

Si la transformación está en una asignación dinámica, los campos generados en la transformación podrían cambiar. Puede parametrizar toda la expresión de estrategia de actualización. Si utiliza un parámetro para definir la expresión, Developer tool no puede validar la expresión. Un parámetro de expresión no puede contener otro parámetro.

Propiedades avanzadas de la transformación de estrategia de actualización

Configure las propiedades avanzadas para ayudar a determinar cómo el servicio de integración de datos procesa los datos de la transformación de estrategia de actualización.

Puede definir las siguientes propiedades avanzadas para la transformación de estrategia de actualización en la ficha Avanzadas:

Reenviar filas rechazadas

Determina si la transformación de estrategia de actualización pasa las filas rechazadas a la siguiente transformación o anula las filas rechazadas. De forma predeterminada, el servicio de integración de datos reenvía las filas rechazadas a la siguiente transformación. El servicio de integración de datos marca las filas para rechazar y las escribe en el archivo de rechazo. Si no selecciona Reenviar filas rechazadas, el servicio de integración de datos anula las filas rechazadas y las escribe en el archivo de registro de asignación.

Usar fusión de Hive

Determina si la transformación de estrategia de actualización usa MERGE de Hive para realizar actualizaciones en destinos de Hive cuando la asignación se ejecuta en el motor de Spark. Cuando una consulta utiliza una instrucción MERGE en vez de las instrucciones INSERT, UPDATE o DELETE, el procesamiento es más eficaz.

La asignación ignora la opción MERGE de Hive en los siguientes casos:

- La asignación se ejecuta en el motor de Databricks Spark.
- La asignación se ejecuta en el motor de Blaze.
- MERGE está restringido por la implementación de Hive en distribuciones de Hadoop particulares.

Las asignaciones que se ejecutan en el motor de Databricks Spark utilizan el comando SQL de Databricks "MERGE INTO" en vez de MERGE de Hive. En el caso de las asignaciones que se ejecutan en el motor de Blaze y en los casos donde MERGE está restringido, el motor en tiempo de ejecución utiliza INSERT, UPDATE y DELETE para realizar la operación.

El registro de la asignación contiene resultados de la operación, incluidos los casos en los que las restricciones hayan afectado a los resultados.

Cuando la actualización afecta a las columnas de partición o de cubo, se omiten las actualizaciones de las columnas.

Nota: Developer tool y el servicio de integración de datos no realizan validaciones con esta restricción. Si la expresión de la estrategia de actualización infringe estas restricciones, la asignación podría tener unos resultados inesperados.

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformaciones de estrategia de agregación y actualización

Si conecta las transformaciones de estrategia de agregación y actualización como partes de la misma canalización, coloque la transformación de estrategia de agregación antes que la de actualización. Así, el servicio de integración de datos lleva a cabo el cálculo agregado para, después, marcar las filas que contengan los resultados de este cálculo con el fin de insertarlo, actualizarlo, eliminarlo o rechazarlo.

Si coloca la transformación de estrategia de actualización antes que la de agregación, debe tener en cuenta cómo la transformación de agregación trata las filas marcadas para diferentes operaciones. En este orden, el servicio de integración de datos marca las filas para llevar a cabo la inserción, actualización, eliminación o rechazo antes de que se realice el cálculo de agregación. Según cómo marque una fila, así tratará la transformación de agregación los valores que contenga y que se utilicen en el cálculo. Por ejemplo, si marca una fila para eliminarla y, a continuación, la utiliza para calcular la suma, el servicio de integración de datos sustrae el valor en esta fila. Si marca una fila para rechazarla y, a continuación, la utiliza para calcular la suma, el servicio de integración de datos no incluye el valor de esta fila. Si marca una fila para insertarla o actualizarla y, a continuación, la utiliza para calcular la suma, el servicio de integración de datos añade a la suma el valor de esta fila.

Cómo especificar opciones de actualización para destinos individuales

Después de crear una transformación de estrategia de actualización para marcar cada una de las filas para que realice una operación de base de datos concreta, establezca las opciones de inserción, actualización y supresión para cada destino de la asignación. Puede anular las inserciones o supresiones de las filas

marcadas para insertar o eliminar. Puede elegir diferentes formas para gestionar las actualizaciones de todas las filas marcadas para actualizar.

Establezca las opciones de la estrategia de actualización en las propiedades avanzadas de un objeto de datos de destino en una asignación. Puede establecer las siguientes opciones de estrategia de actualización:

Insertar

Inserta en un destino todas las filas marcadas para insertar. Esta propiedad está habilitada de manera predeterminada.

Suprimir

Suprime de un destino todas las filas marcadas para suprimir. Esta propiedad está habilitada de manera predeterminada.

Estrategia de actualización

La estrategia de actualización de las filas existentes. Seleccione una de las siguientes estrategias:

- **Actualizar como actualización.** Actualiza todas las filas marcadas para actualizar. Este es el valor predeterminado.
- **Actualizar como inserción.** Inserta todas las filas marcadas para actualizar.
- **Actualizar o insertar.** Actualiza todas las filas marcadas para actualizar si existen en el destino y, a continuación, inserta las que estén marcadas para insertar.

Truncar tabla

Trunca el destino antes de cargar los datos. Esta propiedad está deshabilitada de manera predeterminada.

Transformación de estrategia de actualización en un entorno no nativo

El procesamiento de transformaciones de estrategia de actualización en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible con restricciones.
- Motor de Spark. Compatible con restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible con restricciones.

Transformación de estrategia de actualización en el motor de Blaze

Algunas reglas de procesamiento del motor de Blaze son diferentes de las reglas de procesamiento del servicio de integración de datos.

Restricciones generales

Puede usar la transformación de estrategia de actualización en las distribuciones de Hadoop compatibles con Hive ACID.

Si la transformación de estrategia de actualización recibe varias filas de actualización para el mismo valor de clave principal, la transformación selecciona una fila aleatoria para actualizar el destino.

Si varias transformaciones de estrategia de actualización escriben en instancias diferentes del mismo destino, es posible que no se puedan predecir los datos del destino.

El motor de Blaze ejecuta las operaciones en el siguiente orden: eliminaciones, actualizaciones, inserciones. No procesa filas en el mismo orden en el que las recibe la transformación de estrategia de actualización. Los destinos de Hive siempre realizan operaciones Actualizar como actualización. Los destinos de Hive no admiten las operaciones Actualizar o insertar ni Actualizar como inserción.

Validación de asignaciones y validación de compilación

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación de estrategia de actualización está conectada a más de un destino.
- La transformación de estrategia de actualización no está situada inmediatamente antes que el destino.
- El destino de la estrategia de actualización no es un destino de Hive.
- El destino de la transformación de estrategia de actualización es una tabla de ACID externa.
- El destino no contiene ninguna clave principal.
- La propiedad de destino de Hive para truncar la tabla de destino en tiempo de ejecución está habilitada.
- La propiedad de destino de Hive para crear o reemplazar la tabla de destino en tiempo de ejecución está habilitada.

La asignación genera un error en la siguiente situación:

- El destino no se encuentra en cubo de ORC.
- El destino de Hive se ha modificado y tiene menos filas que la tabla real.

Se producen errores de validación de compilación y la ejecución de la asignación se detiene en las siguientes situaciones:

- La tabla de destino no está habilitada para transacciones.
- La versión de Hive es anterior a 0.14.

Usar tablas de destino de Hive

Para usar una tabla de destino de Hive con una transformación de estrategia de actualización, debe crear la tabla de destino de Hive con la siguiente cláusula en el lenguaje de definición de datos de Hive:

```
TBLPROPERTIES ("transactional"="true").
```

Para usar una transformación de estrategia de actualización con un destino de Hive, compruebe que se hayan configurado las siguientes propiedades en el conjunto de configuración hive-site.xml asociado a la conexión de Hadoop:

```
hive.support.concurrency      true
hive.enforce.bucketing       true
hive.exec.dynamic.partition.mode nonstrict
hive.txn.manager              org.apache.hadoop.hive.ql.lockmgr.DbTxnManager
hive.compactor.initiator.on   true
hive.compactor.worker.threads 1
```

Transformación de estrategia de actualización en el motor de Spark

Algunas reglas de procesamiento del motor de Spark son diferentes de las reglas de procesamiento del servicio de integración de datos.

Restricciones generales de los destinos de Hive

Puede usar la transformación de estrategia de actualización en las distribuciones de Hadoop compatibles con Hive ACID.

La transformación de estrategia de actualización también se puede usar para escribir resultados de asignación de escritura en destinos relacionales compatibles con JDBC.

La transformación de estrategia de actualización no reenvía las filas rechazadas a la siguiente transformación cuando el destino es una tabla de Hive o una tabla compatible con JDBC.

Si la transformación de estrategia de actualización recibe varias filas de actualización para el mismo valor de clave principal, la transformación selecciona una fila aleatoria para actualizar el destino.

Si varias transformaciones de estrategia de actualización escriben en instancias diferentes del mismo destino, es posible que no se puedan predecir los datos del destino.

Si la asignación se ejecuta en el motor de Spark, podrá elegir la opción Usar fusión de Hive. Esta opción presenta las siguientes restricciones:

- Una misma fila para eliminar o actualizar no puede coincidir con varias filas en el destino. Cuando la asignación infringe esta restricción, se genera un error en tiempo de ejecución en la asignación.
- Si configura la expresión de la estrategia de actualización para que se actualicen las columnas de partición o de cubo, la asignación omite la opción MERGE de Hive y no actualiza las columnas.

Nota: Developer tool y el servicio de integración de datos no realizan validaciones con estas restricciones. Si la expresión o la asignación infringen estas restricciones, la asignación se podría ejecutar, pero los resultados no serán los previstos.

Los destinos de Hive siempre realizan operaciones Actualizar como actualización. Los destinos de Hive no admiten las operaciones Actualizar o insertar ni Actualizar como inserción.

Usar datos jerárquicos en expresiones de estrategia de actualización

Puede configurar la expresión para utilizar los tipos primitivos en un campo de datos complejo.

Por ejemplo, tiene una matriz anidada de tipos de datos struct definidos con la siguiente instrucción:

```
addresses array<struct<city:varchar(15), apartment:smallint, pincode:int>>
```

Como la columna "city" es un tipo primitivo, puede utilizarla en la expresión de estrategia de actualización usando la siguiente expresión:

```
IIF( (addresses[0].city = 'NEW YORK'), DD_INSERT, DD_UPDATE)
```

Validación de asignaciones

Los campos de salida de la transformación de estrategia de actualización deben coincidir exactamente con los campos de entrada de destino.

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación de estrategia de actualización está conectada a más de un destino.
- La transformación de estrategia de actualización no está situada inmediatamente antes que el destino.
- El destino de la transformación de estrategia de actualización es una tabla de ACID externa.
- El destino no contiene ninguna clave principal conectada.
- La propiedad para habilitar el truncado de la tabla de destino en tiempo de ejecución está seleccionada.

- Se ha seleccionado una de las siguientes estrategias de destino para la tabla de destino en tiempo de ejecución:
 - Crear o reemplazar la tabla de destino
 - ApplyNewColumns
 - ApplyNewSchema
 - Error

Se produce un error de asignación en las siguientes situaciones cuando el destino es un destino de Hive:

- La tabla de destino no está habilitada para transacciones.
- El destino no se encuentra en cubo de ORC.

Usar tablas de destino de Hive

Para usar una tabla de destino de Hive con una transformación de estrategia de actualización, debe crear la tabla de destino de Hive con la siguiente cláusula en el lenguaje de definición de datos de Hive:

```
TBLPROPERTIES ("transactional"="true").
```

Para usar una transformación de estrategia de actualización con un destino de Hive, compruebe que se hayan configurado las siguientes propiedades en el conjunto de configuración hive-site.xml asociado a la conexión de Hadoop:

```
hive.support.concurrency      true
hive.enforce.bucketing      true
hive.exec.dynamic.partition.mode nonstrict
hive.txn.manager              org.apache.hadoop.hive.ql.lockmgr.DbTxnManager
hive.compactor.initiator.on   true
hive.compactor.worker.threads 1
```

Transformación de estrategia de actualización en el motor de Databricks Spark

Puede utilizar la transformación de estrategia de actualización para realizar las operaciones INSERT y UPDATE, entre otras. También puede realizar operaciones INSERT y UPDATE en columnas seleccionadas de Delta Lake.

Restricciones generales

La transformación de estrategia de actualización solamente se puede utilizar en el motor de Databricks Spark cuando se usa un destino de tabla de Delta Lake.

La transformación de estrategia de actualización solo se puede usar en distribuciones de Databricks que admiten Delta ACID.

Validación de asignaciones

La validación de asignaciones genera un error en las siguientes situaciones:

- La transformación de estrategia de actualización está conectada a más de un destino.
- La transformación de estrategia de actualización no está situada inmediatamente antes que el destino.
- El destino de la transformación de estrategia de actualización es una tabla de ACID externa.
- El destino no contiene ninguna clave principal conectada.
- La propiedad para habilitar el truncado de la tabla de destino en tiempo de ejecución está seleccionada.

- Se ha seleccionado una de las siguientes estrategias de destino para la tabla de destino en tiempo de ejecución:
 - Crear o reemplazar la tabla de destino
 - ApplyNewColumns
 - ApplyNewSchema
 - Error

CAPÍTULO 47

Transformación de consumidor de servicio web

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de consumidor de servicio web, 695](#)
- [Selección WSDL, 698](#)
- [Puertos de transformación de consumidor de servicio web, 699](#)
- [Asignación de entrada de transformación de consumidor de servicio web, 701](#)
- [Asignación de salida de transformación de consumidor de servicio web, 704](#)
- [Propiedades avanzadas de la transformación de consumidor de servicio web, 707](#)
- [Optimizaciones de filtro, 712](#)
- [Crear una transformación de consumidor de servicio web, 714](#)
- [Ejemplo de transformación de consumidor de servicio web, 716](#)

Resumen de la transformación de consumidor de servicio web

La transformación de consumidor de servicio web se conecta a un servicio web como un cliente del servicio web para acceder a los datos o para transformarlos. La transformación de consumidor de servicio web es una transformación de varios grupos.

Un servicio web utiliza estándares abiertos, como SOAP, WSDL y XML. SOAP es el protocolo de comunicación para los servicios web. La solicitud del cliente del servicio web y la respuesta del servicio web son mensajes SOAP. Un WSDL es un esquema XML que describe los protocolos, los formatos y las firmas de las operaciones del servicio web.

Las operaciones del servicio web incluyen solicitudes de información, de actualización de datos o de ejecución de tareas. Por ejemplo, la transformación de consumidor de servicio web envía una solicitud SOAP para ejecutar una operación del servicio web denominada `getCustomerOrders`. La transformación pasa un ID de cliente en la solicitud. El servicio web recupera la información del cliente y del pedido. El servicio web devuelve la información a la transformación en una respuesta SOAP.

La transformación de consumidor de servicio web se conecta a un servicio web mediante una URL de punto final definida en el WSDL, en una conexión de servicios web o en un puerto de entrada de URL de punto final. La seguridad para los servicios web se habilita en una conexión de servicios web.

Mensajes SOAP

La transformación de consumidor de servicio web utiliza el Simple Object Access Protocol (SOAP) para intercambiar información con el proveedor de servicios web y para solicitar servicios web. SOAP define el formato de la solicitud del servicio web y de los mensajes de respuesta.

Cuando se transforman datos con una transformación de consumidor de servicio web, la transformación genera una solicitud SOAP y se conecta al servicio web. La transformación se conecta al servicio web mediante una URL de punto final definida en el objeto WSDL, en una conexión de servicios web o en un puerto de entrada de URL de punto final. La solicitud SOAP contiene la información que necesita el servicio web para ejecutar la operación solicitada. La operación del servicio web devuelve datos a la transformación en una respuesta SOAP. La transformación asigna datos desde la respuesta SOAP y devuelve los datos en puertos de salida.

La transformación del consumidor de servicio web codifica los encabezados del mensaje SOAP en ISO-8859-1.

La transformación puede procesar mensajes SOAP con codificación de documento/literal. El estilo de documento/literal requiere un esquema XML para describir el mensaje SOAP. Los mensajes SOAP se forman desde el XML. Cuando un mensaje SOAP contiene elementos de ocurrencia múltiple, los grupos de elementos constituyen niveles en la jerarquía XML. Los grupos están relacionados cuando un nivel está anidado dentro de otro.

Un mensaje de solicitud SOAP puede contener datos jerárquicos. Por ejemplo, la transformación de consumidor de servicio web envía una solicitud para añadir pedidos de clientes a una base de datos de ventas. La transformación pasa dos grupos de datos en un mensaje de solicitud SOAP. Un grupo contiene un ID y un nombre de cliente y el otro grupo contiene información de los pedidos. La información de los pedidos se puede generar varias veces.

Un mensaje de respuesta SOAP puede contener datos jerárquicos. Por ejemplo, la transformación de consumidor de servicio web genera una solicitud SOAP para los pedidos de clientes. El servicio web devuelve un encabezado de pedido y elementos de detalle de los pedidos de ocurrencia múltiple en la respuesta SOAP.

Archivos WSDL

Un archivo WSDL contiene una descripción de los datos que deben pasarse al servicio web para que el emisor y el receptor entiendan los datos que deben intercambiar. Debe importar un archivo WSDL en el repositorio antes de poder crear una transformación de consumidor de servicio web.

El WSDL describe las operaciones que se van a realizar en los datos y en un enlace en un protocolo o transporte, de modo que el consumidor de servicio web pueda enviar el mensaje de solicitud en el formato correcto. El WSDL describe la dirección de red para conectarse al servicio web.

El WSDL incluye información sobre cómo codificar la solicitud SOAP y los mensajes de respuesta. La codificación SOAP determina el formato del cuerpo del mensaje SOAP. Describe el formato de los mensajes de solicitud y de respuesta que utiliza el servicio web para comunicarse con el consumidor de servicio web. Los desarrolladores de servicios web pueden utilizar una variedad de juegos de herramientas para crear servicios web. Los juegos de herramientas son compatibles con diversas maneras de codificar los mensajes SOAP.

La transformación de consumidor de servicio web es compatible con el estilo de codificación SOAP de documento/literal. Puede utilizar WSDL 1.1 con la transformación de consumidor de servicio web. No se pueden usar anexos WDSL como mensajes MIME, DIME y MTOM.

Operaciones

Un servicio web contiene una operación para cada acción respaldada por el servicio web.

Por ejemplo, un servicio web puede tener una operación llamada `getcustomerid` que recibe un nombre de cliente y que responde con los detalles del cliente. La entrada de operación incluye un elemento para el nombre de cliente. La salida de operación incluye elementos para detalles del cliente según el nombre de cliente.

Cuando se configura una transformación de consumidor de servicio web, se define la manera en la que la transformación asigna datos a la entrada de operación y de la salida de operación. En la transformación se configura la siguiente información:

Asignación de entrada

Defina cómo deben asignarse los puertos de entrada de transformación a los nodos de entrada de operación del servicio web. La entrada de operación define los elementos en la solicitud SOAP para la operación.

Asignación de salida

Defina cómo deben asignarse los nodos de salida de operación del servicio web a los puertos de salida de transformación. La salida de operación define los elementos en una respuesta SOAP para la operación.

Seguridad de servicio web

La seguridad para los servicios web se habilita en una conexión de servicios web. Puede configurar los siguientes tipos de seguridad:

Seguridad de servicio web

El servicio de integración de datos puede incluir un encabezado de seguridad de servicio web cuando envía una solicitud SOAP al proveedor del servicio web. El encabezado de seguridad del servicio web contiene información de autenticación de modo que el proveedor del servicio web puede autenticar el servicio de integración de datos.

La transformación de consumidor de servicio web proporciona el token de nombre de usuario. El servicio de integración de datos crea un encabezado SOAP de seguridad aparte en la solicitud SOAP y pasa la solicitud al proveedor del servicio web.

Puede utilizar los siguientes tipos de seguridad de servicio web en una conexión de servicios web:

- `PasswordText`. El servicio de integración de datos no cambia la contraseña en el encabezado SOAP de seguridad WS.
- `PasswordDigest`. El Servicio de integración de datos combina la contraseña con un valor de seguridad (`nonce`) y una marca de tiempo. El Servicio de integración de datos aplica una dispersión SHA a la contraseña, la codifica como `base64` y utiliza la contraseña codificada en el encabezado SOAP.

Seguridad de la capa de transporte

Seguridad implementada en la parte superior de la capa de transporte (capa TCP) del TCP/IP mediante `Secure Sockets Layer (SSL)`. Los servicios web utilizan `Hypertext Transfer Protocol sobre SSL (HTTPS)` como dirección web para el transporte de mensajes seguro. Las transformaciones de consumidor de servicio web pueden utilizar `TLS 1.2`, `TLS 1.1` o `TLS 1.0`. Puede utilizar el siguiente método de autenticación con la seguridad de la capa de transporte: autenticación HTTP, autenticación de servidor proxy y certificados SSL.

Autenticación SSL

Puede utilizar la autenticación SSL al conectarse mediante el protocolo HTTPS.

Puede utilizar los siguientes tipos de autenticación SSL:

- Autenticación SSL unidireccional
- Autenticación SSL bidireccional

Autenticación HTTP

Puede utilizar la autenticación HTTP al conectarse mediante el protocolo HTTP.

Puede utilizar los siguientes métodos de autenticación HTTP:

- Autenticación básica
- Autenticación implícita
- Autenticación de NT LAN Manager (NTLM)

Selección WSDL

Antes de crear una transformación de consumidor de servicio web, debe importar un archivo WSDL al repositorio de modelos. Solo puede importar archivos WSDL que se hayan definido con un estilo de vinculación SOAP de documento/literal.

El WSDL define la firma de la operación del servicio web que desee ejecutar. Cuando se importa un WSDL, Developer Tool crea un objeto de datos físicos que se puede reutilizar para otras transformaciones.

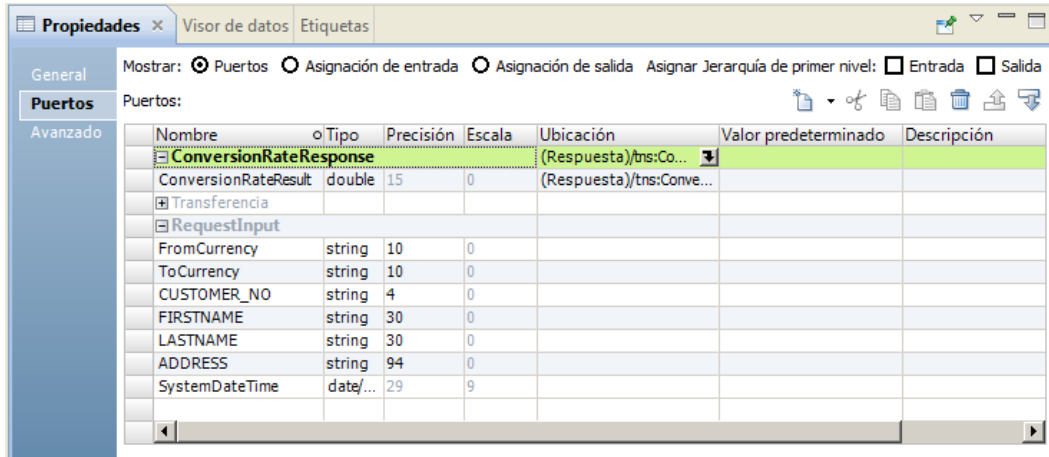
Un WSDL puede definir múltiples operaciones. Cuando cree una transformación de consumidor de servicio web, seleccione la operación que desea ejecutar. Puede ver las jerarquías de entrada y de salida de la operación en la transformación de consumidor de servicio web. Las jerarquías definen la estructura del mensaje de solicitud SOAP y del mensaje de respuesta SOAP.

También puede importar un WSDL con operación de entrada unidireccional. Debe crear puertos de salida ficticios al importar un WSDL con operación de entrada unidireccional.

Puertos de transformación de consumidor de servicio web

Cuando visualice los puertos de transformación, muestre los puertos si no necesita ver la jerarquía de la operación. Cuando se muestran los puertos, puede definir grupos y puertos y asignar nodos de la salida de operación a los puertos de salida.

La siguiente figura muestra los puertos para una transformación de consumidor de servicio web no reutilizable:



Nombre	Tipo	Precisión	Escala	Ubicación	Valor predeterminado	Descripción
ConversionRateResponse				(Respuesta)/tns:Co...		
ConversionRateResult	double	15	0	(Respuesta)/tns:Conve...		
Transferecia						
RequestInput						
FromCurrency	string	10	0			
ToCurrency	string	10	0			
CUSTOMER_NO	string	4	0			
FIRSTNAME	string	30	0			
LASTNAME	string	30	0			
ADDRESS	string	94	0			
SystemDateTime	date/...	29	9			

Una transformación de consumidor de servicio web puede poseer varios grupos de entrada y de salida. Cuando cree puertos, cree grupos y añada los puertos a los grupos. Defina los puertos en una jerarquía de grupo según la estructura de la entrada de operación o la jerarquía de salida de operación. Añada una clave para relacionar un grupo secundario con un grupo principal. Todos los grupos, excepto el grupo situado en el nivel más bajo de la jerarquía, deben tener claves primarias. Todos los grupos de la jerarquía, excepto el grupo raíz, deben tener claves externas.

La transformación posee un grupo de entrada raíz denominado RequestInput. Debe añadir una clave principal al grupo de entrada raíz. La clave debe contener valores string, bigint o integer.

También puede añadir puertos de transferencia adicionales al grupo de entrada raíz. Los puertos de transferencia pasan datos a través de la transformación sin modificar los datos. El puerto de transferencia puede ocurrir una vez en los datos de entrada. Puede añadir el puerto de transferencia a cualquier grupo de salida. Asocie el puerto de salida al puerto de entrada. El valor de entrada que se transfiere mediante una solicitud SOAP se repite en las filas de salida de la respuesta SOAP.

También puede añadir encabezados HTTP, puertos de cookie, un puerto de URL dinámica y puertos para la autenticación de seguridad de servicio web al grupo de entrada raíz. Los datos en el grupo raíz se generan una vez.

Para asignar un nodo de salida de operación a un puerto de salida, haga clic en el campo de la columna **Ubicación** y expanda la jerarquía del cuadro de diálogo **Seleccionar ubicación**. A continuación, elija un nodo de la jerarquía.

Puertos de entrada de encabezado HTTP

Un servicio web puede requerir encabezados HTTP adicionales. Puede crear puertos de entrada en el grupo de entrada raíz para pasar la información de los encabezados adicionales al proveedor del servicio web.

Para añadir un encabezado HTTP y un puerto HTTP, seleccione el grupo de entrada raíz y haga clic en la flecha situada junto al botón **Nuevo**. A continuación, haga clic en **Encabezado HTTP**. Especifique un nombre de encabezado y un nombre de puerto.

Se pueden crear varios encabezados HTTP.

Otros puertos de entrada

Puede añadir puertos de entrada predefinidos a la transformación de consumidor de servicio web.

Puede añadir los siguientes puertos de entrada predefinidos:

Puerto de cookie

Puede configurar la transformación de consumidor de servicio web para que utilice autenticación por cookies. El servidor remoto del servicio web hace un seguimiento de los usuarios del servicio web mediante cookies. Es posible aumentar el rendimiento cuando una asignación llama un servicio web varias veces.

Cuando se proyecta el puerto de cookie a un mensaje de solicitud de servicio web, el proveedor del servicio web devuelve un valor de cookie en el mensaje de respuesta. Puede pasar el valor de cookie a otra transformación situada en un nivel inferior en la asignación o puede guardar el valor de la cookie en un archivo. Cuando se guarda el valor de la cookie en un archivo, puede configurar la cookie como entrada de la transformación de consumidor de servicio web.

Puede proyectar el puerto de salida de la cookie a cualquiera de los grupos de salida de la transformación de consumidor de servicio web.

Puerto de URL de punto final

La transformación de consumidor de servicio web se conecta con un servicio web mediante una URL de punto final. Puede definir la URL de punto final en el archivo WSDL, en una conexión de servicios web o en un puerto de entrada de URL de punto final. Cuando la transformación recibe la URL dinámicamente en un puerto, el servicio de integración de datos reemplaza la URL definida en el archivo WSDL o en la conexión de servicios web.

La transformación de consumidor de servicio web puede tener un valor de puerto URL para cada solicitud de servicio web. Añada un puerto de URL de punto final al grupo de entrada raíz.

Puertos de seguridad WS

Se habilita la seguridad de servicio web en una conexión de servicios web. Cuando se habilita la seguridad de servicio web, debe definir el nombre de usuario y la contraseña en una conexión de servicios web o en puertos de entrada de seguridad WS.

Cuando se añaden puertos de seguridad WS, se pasan el nombre de usuario y la contraseña a través de puertos de entrada en la transformación. Cuando la transformación recibe el nombre de usuario y la contraseña dinámicamente en los puertos, el servicio de integración de datos reemplaza los valores definidos en la conexión de servicios web.

Nota: una conexión de servicios web posee un nombre de usuario y una contraseña para la autenticación de HTTP y de seguridad WS.

Para añadir puertos de entrada predefinidos, haga clic en el grupo de entrada raíz en el área **Puertos**. Haga clic en la flecha situada junto al botón **Nuevo** y luego haga clic en **Otros puertos**. Elija los puertos que desea añadir.

Asignación de entrada de transformación de consumidor de servicio web

Cuando visualice los puertos de transformación, muestre la asignación de entrada para ver la jerarquía de entrada de operación. Cuando se muestra la asignación de entrada, puede definir grupos y puertos de entrada y asignar puertos de entrada a nodos de entrada de operación.

La asignación de entrada incluye las siguientes áreas:

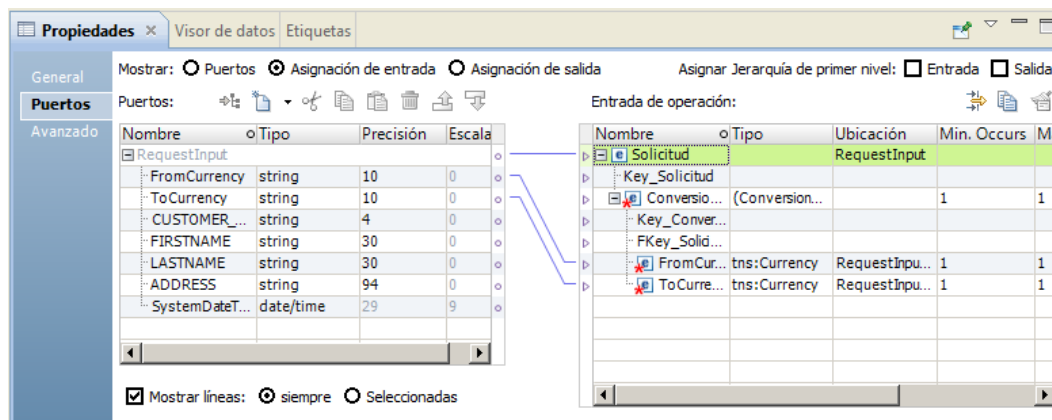
Puertos

Cree los grupos de entrada de transformación y los puertos de entrada en el área **Puertos**.

Entrada de operación

El área **Entrada de operación** muestra los nodos del mensaje de solicitud SOAP que la transformación de consumidor de servicio web envía al servicio web. El objetos de datos WSDL que se utiliza para crear la transformación define la jerarquía de entrada de a operación.

La siguiente figura muestra la asignación de entrada para una transformación de consumidor de servicio web no reutilizable:



Después de crear puertos de entrada, asigne los puertos de entrada del área **Puertos** a los nodos del área **Entrada de operación**. Cuando se asigna un puerto de entrada a un nodo de la entrada de operación, la ubicación del nodo aparece en la columna **Ubicación** del área **Entrada de operación**.

La herramienta del desarrollador asigna nodos del primer nivel de la entrada de operación a puertos de entrada cuando se opta por asignar el primer nivel de la jerarquía de entrada. Asimismo, la herramienta del desarrollador crea los puertos que necesita para realizar la asignación. Si el primer nivel de la jerarquía contiene un nodo principal de ocurrencia múltiple con uno o varios nodos secundarios de ocurrencia múltiple, la herramienta del desarrollador no asigna el primer nivel de la jerarquía.

Puede asignar datos XML de una cadena o un puerto de entrada de texto al mensaje de solicitud SOAP completo. Cuando se asignan datos XML a la solicitud SOAP completa, no se pueden asignar puertos a nodos de la entrada de operación.

Puede optar por ver las líneas que conectan los puertos de entrada a los nodos de la entrada de operación.

TEMAS RELACIONADOS

- [“Resumen de la generación de mensajes SOAP del servicio web” en la página 729](#)

Reglas y directrices para asignar puertos de entrada a nodos

Revise las siguientes reglas cuando asigne puertos de entrada a nodos en la jerarquía de entrada de operación:


- Puede asignar un puerto de entrada a un nodo de la jerarquía. Puede asignar el mismo puerto a un número indeterminado de claves de la jerarquía.
- El puerto de entrada y el nodo deben tener tipos de datos compatibles.
- Puede asignar puertos de un grupo de entrada a varios niveles de jerarquía en la entrada de operación.
- Debe asignar puertos de entrada a las claves en la entrada de operación. Cualquier puerto que se asigne a una clave debe ser del tipo de datos string, integer o bigint. Asigne los datos a las claves en todos los niveles de la entrada de operación situados encima del nivel de jerarquía que esté incluyendo en el mensaje SOAP. Incluya las claves externas para todos los niveles situados por encima e incluyendo el nivel que esté asignando.

Nota: No es necesario que asigne puertos de entrada a claves si únicamente está asignando el nivel más bajo de la jerarquía de entrada de operación.

- Puede asignar varios puertos de entrada con valores string, integer o bigint a una clave en el área **Entrada de operación** para crear una clave compuesta. Cuando se hace clic en el campo **Ubicación** para una clave compuesta, puede volver a ordenar los puertos de entrada o eliminar uno de los puertos.

Personalizar opciones de vista

Puede cambiar la jerarquía de entrada de operación para que las claves se muestren en el área **Entrada de operación**. También puede mostrar construcciones de agrupación que definen cómo ordenar nodos.

Haga clic en el botón **Personalizar vista** () en el área **Entrada de operación**. Habilite una de las siguientes opciones:

Secuencia, Selección y Todo

Muestra una línea que indica si una definición de elemento es secuencia, selección o todo.

Los nodos en un grupo todo deben incluirse en el mensaje SOAP.

Los nodos en un grupo secuencia deben estar en el orden especificado en el WSDL.

En el mensaje SOAP debe aparecer, como mínimo, un nodo en un grupo selección.

Claves

Visualice las claves en el área **Entrada de operación**. El área **Entrada de operación** incluye claves para cada grupo. Puede añadir una clave a un puerto de entrada en el área **Puertos**.

Cómo asignar puertos de entrada a la entrada de operación

Cuando se muestra la asignación de entrada de transformación, puede definir grupos y puertos de entrada y asignar puertos de entrada a nodos de entrada de operación.

1. Abra una transformación de consumidor de servicio web.
2. Para mostrar la asignación de entrada de la transformación, utilice uno de los siguientes métodos:

- Para una transformación reutilizable, haga clic en la vista **Resumen**. Elija mostrar la asignación de entrada.
 - Para una transformación no reutilizable, haga clic en la ficha **Puertos** en la vista **Propiedades**. Elija mostrar la asignación de entrada.
3. Defina una clave primaria para el grupo de entrada raíz.
 4. Para añadir un grupo o un puerto de entrada al área **Puertos**, utilice uno de los siguientes métodos:

Opción	Descripción
Arrastrar un nodo	Arrastre un nodo de grupo o un nodo secundario en el área Entrada de operación hasta una columna vacía del área Puertos . Si el nodo es un nodo de grupo, Developer Tool agrega un grupo sin puertos.
Añadir un grupo o puerto manualmente	Para añadir un grupo, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Grupo . Para añadir un puerto, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Campo .
Arrastrar un puerto desde otra transformación	En el editor, arrastre un puerto desde otra transformación a la transformación de consumidor de servicio web.
Copiar un puerto	Seleccione puertos de otra transformación y cópielos al área Puertos . Para copiar puertos, puede utilizar métodos abreviados de teclado o puede utilizar los botones copiar y pegar de Developer Tool.
Seleccione Asignar primer nivel de la jerarquía .	Seleccione Asignar primer nivel de la jerarquía . Developer Tool asigna nodos que se encuentran en el primer nivel de la entrada de operación a puertos y grupos de entrada. Developer Tool también crea los puertos y grupos de entrada para realizar la asignación.

5. Si crea manualmente un puerto o lo copia desde otra transformación, haga clic en la columna **Ubicación** del área **Entrada de operación** y elija un puerto de la lista.
6. Para asignar puertos de entrada como una clave compuesta, utilice uno de los siguientes métodos:

Opción	Descripción
Arrastrar puertos de entrada	Seleccione dos o más puertos de entrada y arrástrelos hasta una clave en la jerarquía de entrada de operación.
Seleccione puertos de entrada desde el cuadro de diálogo Seleccionar ubicación	Haga clic en la columna Ubicación de una clave en la jerarquía de entrada de operación y, a continuación, seleccione los puertos de entrada.

7. Para borrar las ubicaciones de los nodos, utilice uno de los siguientes métodos:

Opción	Descripción
Haga clic en Borrar .	Seleccione uno o más nodos del área Entrada de operación y haga clic en Borrar .
Eliminar las líneas que conectan puertos con nodos	Seleccione una o más líneas que conectan los puertos de entrada con los nodos en la entrada de operación y pulse Eliminar .

8. En el caso de que el objeto de datos WSDL asociado incluya elementos anyType, elementos, atributos anyAttribute, elementos de tipo derivado o grupos de sustitución, elija objetos que se encuentran en el

área **Entrada de operación**. En la columna **Tipo** de un nodo, haga clic en **Elegir** y luego elija uno o más tipos, elementos o atributos de la lista.

- Para asignar datos XML de una cadena o un puerto de entrada de texto a la solicitud SOAP completa, haga clic con el botón derecho en el puerto y seleccione **Asignar como XML**.

Asignación de salida de transformación de consumidor de servicio web

Cuando visualice los puertos de transformación, muestre la asignación de salida para ver la jerarquía de salida de operación. Cuando se muestra la asignación de salida, puede definir grupos y puertos de salida y asignar nodos de salida de operación a puertos de salida.

La asignación de salida incluye las siguientes áreas:

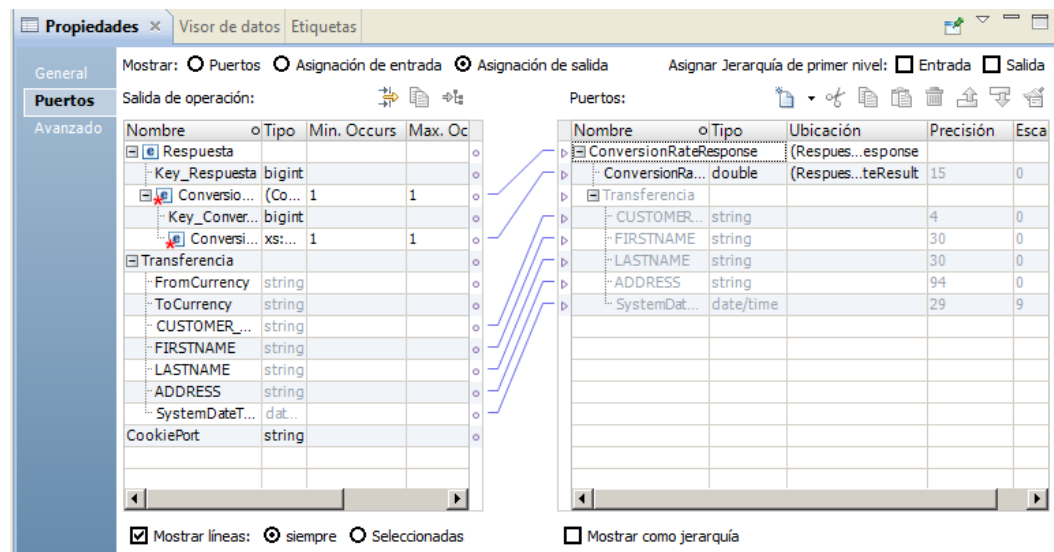
Salida de operación

El área **Salida de operación** muestra los nodos del mensaje de respuesta SOAP que el servicio web devuelve a la transformación de consumidor de servicio web. El objeto de datos WSDL que se utiliza para crear la transformación define la jerarquía de salida de operación.

Puertos

Cree los puertos y los grupos de salida de transformación en el área **Puertos**.

La siguiente figura muestra la asignación de salida para una transformación de consumidor de servicio web no reutilizable:



Después de crear los puertos de salida, asigne los nodos del área **Salida de operación** a los puertos del área **Puertos**. Cuando se asigna un nodo de la salida de operación a un puerto de salida, la ubicación del nodo aparece en la columna **Ubicación** en el área **Puertos**.

La herramienta del desarrollador asigna nodos del primer nivel de la salida de operación a puertos de salida cuando se opta por asignar el primer nivel de la jerarquía de salida. Asimismo, la herramienta del desarrollador crea los puertos que necesita para realizar la asignación. Si el primer nivel de la jerarquía

contiene un nodo principal de ocurrencia múltiple con uno o varios nodos secundarios de ocurrencia múltiple, la herramienta del desarrollador no asigna el primer nivel de la jerarquía.

Puede elegir que se muestren los puertos de salida de una jerarquía. Cada grupo secundario aparece debajo del grupo principal. También puede optar por ver las líneas que conectan los nodos de la salida de operación a los puertos de salida.

Si el objeto de datos WSDL asociado se elimina del repositorio, la herramienta del desarrollador guarda la ubicación de los nodos de operación en la asignación de salida. Cuando se muestra la asignación de salida, el área **Puertos** sigue mostrando la ubicación de los nodos de operación en la columna **Ubicación** de los puertos de salida. Si asocia otro WSDL con la transformación, la herramienta del desarrollador comprueba si cada ubicación es válida. La herramienta del desarrollador borra la ubicación de los nodos de operación del área **Puertos** de la asignación de salida si la ubicación ha dejado de ser válida.

TEMAS RELACIONADOS

- [“Resumen del análisis de mensajes SOAP del servicio web” en la página 720](#)

Reglas y directrices para asignar nodos a puertos de salida

Revise las siguientes reglas cuando asigne nodos de la jerarquía de salida de operación a puertos de salida:

- El nodo de salida de operación y el puerto de salida deben tener tipos de datos compatibles.
- No se puede asignar un nodo a más de un puerto de salida en un grupo.
- Cada puerto de salida debe tener una ubicación válida, a excepción de que el puerto sea un puerto de transferencia.
- Si arrastra un nodo secundario de ocurrencia múltiple a un puerto de salida vacío, debe relacionar el grupo con otros grupos de salida. Cuando se selecciona un grupo, Developer Tool crea claves para relacionar los grupos.
- Cuando se arrastra un elemento de ocurrencia múltiple a un grupo que contiene el elemento principal, puede configurar el número de ocurrencias de elemento secundario que se van a incluir. También puede reemplazar el grupo principal con el grupo secundario de ocurrencia múltiple en la salida de transformación.

Cómo asignar el mensaje SOAP como XML

Puede asignar el mensaje SOAP completo como XML en lugar de devolver los datos a puertos de salida individuales.

Cuando se asigna el mensaje SOAP como XML, el servicio de integración de datos devuelve el mensaje SOAP completo en un puerto. No crea puertos de salida.

Para asignar el mensaje completo, haga clic con el botón derecho en el grupo raíz en el área **Salida de operación**. Seleccione **Asignar como XML**.

Developer Tool crea un puerto de salida de cadenas. La precisión es 65535 bytes.

Personalizar opciones de vista

Puede cambiar la jerarquía de salida de operación para que los puertos de cookies, los puertos de transferencia y las claves se muestren en el área **Salida de operación**. También puede mostrar construcciones de agrupación que definen cómo ordenar nodos.

Haga clic en el botón **Personalizar vista** en el área **Salida de operación**. Habilite una de las siguientes opciones:

Secuencia, Selección y Todo

Muestra una línea que indica si una definición de elemento es secuencia, selección o todo.

Los nodos en un grupo todo deben incluirse en el mensaje SOAP.

Los nodos en un grupo secuencia deben estar en el orden especificado en el WSDL.

En el mensaje SOAP debe aparecer, como mínimo, un nodo en un grupo selección.

Claves

Visualice las claves en el área **Salida de operación**. El área **Salida de operación** incluye claves para cada grupo. Puede añadir una clave a un puerto de salida en el área **Puertos**.

Puertos de transferencia

El área **Salida de operación** muestra los puertos de transferencia. Los puertos de transferencia son puertos que transfieren datos a través de la transformación sin modificar los datos. Puede proyectar puertos de transferencia desde la salida de operación a cualquiera de los grupos de salida de la transformación de consumidor de servicio web. Un puerto de transferencia recibe los datos sólo una vez, por lo que el puerto se encuentra en el nivel raíz de los mensajes SOAP.

Puertos de cookies

Muestra el puerto de cookies. Cuando se configura la autenticación por cookies, el servidor remoto del servicio web hace un seguimiento de los usuarios del consumidor de servicio web mediante cookies. Cuando se proyecta una cookie de servicio web en el mensaje de solicitud, el servicio web devuelve una cookie en el mensaje de respuesta. Puede proyectar la cookie desde la salida de operación a cualquiera de los grupos de salida de la transformación de consumidor de servicio web.

Cómo asignar la salida de operación a puertos de salida

Cuando se muestra la asignación de salida de transformación, puede definir grupos y puertos de salida y asignar nodos de salida de operación a puertos de salida.

1. Abra una transformación de consumidor de servicio web.
2. Para mostrar la asignación de salida de la transformación, utilice uno de los siguientes métodos:
 - Para una transformación reutilizable, haga clic en la vista **Resumen**. Elija mostrar la asignación de salida.
 - Para una transformación no reutilizable, haga clic en la ficha **Puertos** en la vista **Propiedades**. Elija mostrar la asignación de salida.
3. Para añadir un grupo o un puerto de salida al área **Puertos**, utilice uno de los siguientes métodos:

Opción	Descripción
Arrastrar un nodo	Arrastre un nodo de grupo o un nodo secundario en el área Salida de operación hasta una columna vacía del área Puertos . Si el nodo es un nodo de grupo, Developer Tool agrega un grupo sin puertos.
Añadir un grupo o puerto manualmente	Para añadir un grupo, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Grupo . Para añadir un puerto, haga clic en la flecha situada junto al botón Nuevo y luego haga clic en Campo .
Arrastrar un puerto desde otra transformación	En el editor, arrastre un puerto desde otra transformación a la transformación de consumidor de servicio web.

Opción	Descripción
Copiar un puerto	Seleccione puertos de otra transformación y cópielos al área Puertos . Para copiar puertos, puede utilizar métodos abreviados de teclado o puede utilizar los botones copiar y pegar de Developer Tool.
Seleccione Asignar primer nivel de la jerarquía .	Seleccione Asignar primer nivel de la jerarquía . Developer Tool asigna nodos que se encuentran en el primer nivel de la salida de operación a puertos y grupos de salida. Developer Tool también crea los grupos y puertos de salida para realizar la asignación.

- Si crea manualmente un puerto o lo copia desde otra transformación, haga clic en la columna **Ubicación** del área **Puertos** y elija un nodo de la lista.
- Para borrar las ubicaciones de los puertos, utilice uno de los siguientes métodos:

Opción	Descripción
Haga clic en Borrar .	Seleccione uno o más puertos del área Puertos y haga clic en Borrar .
Eliminar las líneas que conectan los nodos con los puertos	Seleccione una o más líneas que conectan los nodos en la salida de operación con los puertos de salida y pulse Eliminar .

- En el caso de que el objeto de datos WSDL asociado incluya elementos anyType, elementos, atributos anyAttribute, elementos de tipo derivado o grupos de sustitución, elija objetos que se encuentran en el área **Salida de operación**. En la columna **Tipo** de un nodo, haga clic en **Elegir** y luego elija uno o más tipos, elementos o atributos de la lista.
- Para asignar el mensaje de respuesta SOAP completo como XML, haga clic con el botón derecho en el grupo raíz en el área **Salida de operación** y seleccione **Asignar como XML**.

Propiedades avanzadas de la transformación de consumidor de servicio web

Las propiedades avanzadas de la transformación de consumidor de servicio web incluyen el nivel de seguimiento, los puertos de fallos genéricos, la conexión de servicios web y los mensajes de solicitud de servicio web simultáneos.

Puede definir las siguientes propiedades avanzadas para la transformación de consumidor de servicio web en la ficha Avanzadas:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Acción de SOAP

Reemplaza el valor de la acción de SOAP definida en el WSDL por un valor constante para la transformación de consumidor de servicio web.

Habilitar control de fallos SOAP genéricos

Devuelve mensajes de fallo que no se han definido en el WSDL. Crea puertos de salida en un grupo de salida GenericFault para controlar los códigos y los mensajes de fallos.

La siguiente tabla describe los puertos de salida de fallo para SOAP 1.1 y SOAP 1.2:

Puerto de salida de fallo para SOAP 1.1	Puerto de salida de fallo para SOAP 1.2	Descripción
Código de fallo	Código*	Devuelve un código de identificación de fallo.
Cadena de fallo	Motivo*	Devuelve una explicación del error en un mensaje.
Detalles del fallo	Detalle	Devuelve información personalizada que el proveedor del servicio web pasa a la transformación de consumidor de servicio web en un mensaje de fallo genérico.
Actor del fallo	Función	Devuelve información sobre el objeto que ha generado el fallo.
-	Nodo	Devuelve el URI del nodo de SOAP que ha generado el fallo.
* Los puertos de salida de código y motivo son jerárquicos.		

Nota: Puede expandir el puerto de salida de fallo de código para extraer el puerto de salida de fallo del subcódigo a un nivel superior.

Habilitar el control de errores HTTP

Devuelve cualquier error HTTP del servicio web. Crea un puerto de salida de errores HTTP en el grupo de salida GenericFault.

Tratar fallo como error

Agrega mensajes de fallos al registro de asignación. Cuando se produce un fallo, el servicio de integración de datos incrementa el recuento de errores para la asignación. Deshabilite esta propiedad para permitir la optimización de primera selección y de inserción. Esta propiedad está habilitada de manera predeterminada.

Conexión

Identifica el objeto de la conexión de servicios web para conectarse con el servicio web. Cree la conexión de servicios web en Developer Tool. Edite la conexión de servicios web en Developer Tool o en Administrator Tool. Cuando configure una conexión de servicios web, configure la URL de punto final, el tipo de seguridad que el servicio web requiera y un tiempo de espera para la conexión.

La transformación de consumidor de servicio web se conecta con un servicio web mediante una URL de punto final. Puede definir la URL de punto final en el archivo WSDL, en una conexión de servicios web o en un puerto de entrada de URL de punto final.

Utilice las siguientes directrices para saber cuándo debe configurar una conexión de servicios web:

- Configure una conexión cuando desee utilizar una URL de punto final que sea diferente de la URL del archivo WSDL y cuando no esté utilizando un puerto de entrada de URL de punto final.
- Configure una conexión cuando el servicio web al que se conecte requiera seguridad de servicio web, autenticación HTTP o certificados SSL.
- Configure una conexión cuando desee cambiar el tiempo de espera predeterminado de la conexión.

Nota: Puede asociar un objeto de datos WSDL del repositorio a una conexión de servicios web. La conexión asociada pasa a ser la conexión predeterminada para cada transformación de consumidor de servicio web que cree a partir de ese WSDL.

Habilitar compresión

Habilita la codificación de solicitudes SOAP con el método de compresión GZIP y habilita la decodificación de respuestas con GZIP o desinflado.

Validación de esquema XML

Valida el mensaje de respuesta SOAP en tiempo de ejecución. Seleccione **Error en XML no válido** o **Sin validación**.

Entrada ordenada

Habilita el servicio de integración de datos para generar una salida sin procesar todos los datos de entrada. Habilite la entrada ordenada cuando los datos de entrada se ordenen según las claves en la jerarquía de entradas de operación.

Optimización de inserción

Habilita la optimización de inserción. Haga clic en el botón **Abrir** en la propiedad **Optimización de inserción** para seleccionar los puertos de filtro que reciben valores de filtro. Para cada puerto de filtro, seleccione el puerto de salida que contiene la columna filtrada en la respuesta de servicio web.

Tiene efectos secundarios

Casilla de verificación que indica que el servicio web realiza cualquier función además de devolver filas. La transformación de consumidor de servicio web tiene un efecto secundario si el servicio web, además de devolver filas, modifica un objeto o interactúa con otros objetos o funciones. El servicio web puede modificar una base de datos, añadir a un total, elevar una excepción, escribir un correo electrónico o llamar a otros servicios web con efectos secundarios. Deshabilite la propiedad **Tiene efectos secundarios** para permitir la optimización de inserción o la optimización de primera selección. Esta propiedad está habilitada de manera predeterminada.

Habilitar simultaneidad

Habilita la transformación del consumidor de servicio web para crear varias conexiones simultáneas a un servicio web de modo que pueda enviar varias solicitudes de servicio web en paralelo. Cuando habilita la transformación de consumidor del servicio web para crear varias conexiones simultáneas en el servicio web, puede establecer el límite del consumo total de memoria y el número de límites de conexiones simultáneas.

En la siguiente tabla se describen las opciones:

Opciones	Descripción
Habilitar simultaneidad	Crea varias conexiones simultáneas en un servicio web.
Límite conexión simultánea	El número de conexiones de servicio web simultáneas. El valor predeterminado es 20.
Límite total de memoria simultánea (en MB)	El límite total de asignación de memoria para todas las conexiones simultáneas. El valor predeterminado es 100 MB.

Control de errores de servicio web

Puede configurar la transformación de consumidor de servicio web para que pase fallos SOAP y errores HTTP hacia abajo en una asignación. Puede incrementar el recuento de errores cuando se produce un fallo. Configure el control de errores de servicio web en las propiedades avanzadas de la transformación.

Un servicio web devuelve un mensaje de respuesta o devuelve un fallo. Un fallo es un error. El servicio web puede generar fallos diferentes según los errores que se producen.

La transformación de consumidor de servicio web puede devolver los siguientes tipos de fallos:

Fallos SOAP

Errores SOAP que define el WSDL. Configure los puertos de error de salida que devuelven los fallos en el mensaje de respuesta del servicio web. Para un enlace SOAP 1.1, el servicio de integración de datos devuelve el mensaje de fallo, el código de fallo, la cadena de fallo y los elementos del actor del fallo para el fallo. Para un enlace SOAP 1.2, el servicio de integración de datos devuelve el mensaje de fallo, el código, el motivo, el nodo y los elementos de función para el fallo.

Fallos SOAP genéricos

El servicio web genera fallos SOAP en tiempo de ejecución. Los elementos de fallo son diferentes para un enlace SOAP 1.1 y un enlace SOAP 1.2. El WSDL no define fallos SOAP genéricos. Los fallos SOAP genéricos incluyen fallos de autenticación y errores de solicitud SOAP.

Errores HTTP

Developer Tool agrega el puerto de salida del fallo HTTP cuando se habilita el control de errores HTTP en la transformación. El servicio de integración de datos devuelve errores HTTP del servicio web en un puerto de cadena individual. Un error HTTP incluye un código de error y un mensaje.

Si la respuesta SOAP del servicio web tiene datos XML no válidos, la transformación de consumidor de servicio web devuelve un error.

Puede especificar si se deben tratar los fallos SOAP como errores. Cuando se habilita Tratar fallo como error y se produce un fallo SOAP, el servicio de integración de datos incrementa el recuento de errores para la asignación. El fallo aparece en el registro de mensajes.

Compresión de mensajes

Cuando se habilita la compresión de mensajes SOAP, la transformación de consumidor de servicio web comprime los mensajes de solicitud de servicio web y recibe mensajes de respuesta de servicio web comprimidos.

La transformación de consumidor de servicio web codifica la solicitud SOAP mediante compresión GZip. La transformación acepta un mensaje de respuesta codificado con GZip o la compresión por desinflado.

Cuando el servicio de integración de datos recibe la respuesta del servicio web, comprueba el encabezado HTTP de codificación de contenido en el mensaje SOAP y descodifica el mensaje.

El valor predeterminado es no realizar codificación de compresión. El servicio web no comprime la respuesta SOAP.

La siguiente tabla muestra los encabezados en la solicitud los mensajes de respuesta cuando la compresión está activada o desactivada:

Compresión	Encabezado
Activada	Encabezado de codificación de contenido: GZip Encabezado de codificación de aceptación: GZip, desinflado
Desactivada	Encabezado de codificación de contenido vacío Encabezado de codificación de aceptación vacío

En ocasiones, un servicio web codifica un mensaje de respuesta con una compresión predeterminada. La transformación de consumidor de servicio web descodifica el mensaje si está codificado mediante GZip o desinflado. La transformación de consumidor de servicio web registra un mensaje en el registro de asignación en el caso de que el servicio web codifique el mensaje de respuesta inesperadamente.

Habilite la compresión en las propiedades avanzadas de la transformación.

Simultaneidad

Puede habilitar la transformación de consumidor del servicio web para crear varias conexiones simultáneas en un servicio web de modo que pueda enviar varias solicitudes de servicio web en paralelo.

Por ejemplo, al comprobar la información bancaria, puede configurar la transformación de consumidor de servicio web en simultaneidad para que se envíen varias filas en paralelo. Si hay 20 filas de entrada, puede enviar 20 solicitudes simultáneas para un procesamiento más rápido.

Cuando habilita la simultaneidad en la transformación de consumidor de servicio web, puede configurar el límite total del consumo de memoria.

Cuando habilita la simultaneidad en la transformación de consumidor de servicio web, puede configurar el número de conexiones de servicio web simultáneas.

Reglas y directrices para la simultaneidad

Utilice las siguientes reglas y directrices cuando utilice la simultaneidad:

- La simultaneidad es compatible con las filas de entrada ordenada como varias conexiones simultáneas para un servicio web. Las filas de salida ordenadas no son compatibles.
- Utilice la simultaneidad del sí el conjunto de datos tiene más de 100 filas.
- Se recomienda no aumentar el número de conexiones de servicio web simultáneas. El número de conexiones de servicio web simultáneas está vinculado al número de conexiones utilizadas por el sistema operativo. Aumentar el número de conexiones es caro.
- Utilice sistemas que tengan procesadores de varios núcleos con un mínimo de 100 MB de memoria RAM para un rendimiento óptimo al utilizar la función de simultaneidad.
- El límite de memoria de simultaneidad representa la memoria consumida por los flujos de trabajo simultáneos durante la invocación de servicios web.
- Cuando habilita la simultaneidad en la transformación de consumidor de servicio web, puede configurar el límite de consumo de memoria. Asegúrese de que el consumo de memoria no sea mayor que la memoria RAM física del servidor.

Mejores prácticas para la simultaneidad

Para un rendimiento óptimo al utilizar la simultaneidad, siga estas mejores prácticas:

- Evitar cambiar los valores predeterminados del límite de memoria simultánea total y el límite de conexión simultánea.
- Evitar utilizar la simultaneidad para conjuntos de datos con menos de 100 filas.
- Evitar puertos de transferencia en la asignación al utilizar la simultaneidad.

Optimizaciones de filtro

Las optimizaciones de filtro aumentan el rendimiento reduciendo el número de filas que pasan por la asignación. El servicio de integración de datos puede aplicar la optimización de primera selección o la de inserción.

Cuando el servicio de integración de datos aplica un método de optimización del filtro, mueve un filtro tan cerca del origen como sea posible en una asignación. Si el servicio de integración de datos no puede mover un filtro antes de una transformación en una asignación, es posible que pueda insertar la lógica de filtro en una transformación.

Habilitar la optimización de primera selección con la transformación del consumidor de servicio web

Habilite la optimización de primera selección para la transformación de consumidor de servicio web si la transformación no tiene efectos secundarios y no trata los fallos como errores.

1. Abra la vista **Propiedades avanzadas** de la transformación del consumidor de servicio web.
2. Desactive **Tratar fallo como error**.
3. Desactive **Tiene efectos secundarios**.

Optimización de inserción con la transformación del consumidor de servicio web

Puede configurar la optimización de inserción con la transformación de consumidor de servicio web cuando la transformación está en una tabla virtual en un servicio de datos SQL.

La asignación llama al servicio web para recuperar un conjunto o subconjunto de datos según las instrucciones en la consulta SQL de usuario final. La consulta SQL de usuario final contiene una condición de filtro opcional.

Con la optimización de inserción, la transformación del consumidor de servicio web recibe el valor de filtro en un puerto de filtro. El puerto de filtro es un puerto de entrada inconexo que identifica como un puerto de filtro al configurar la optimización de inserción. El puerto de filtro tiene un valor predeterminado que garantiza que el servicio web devuelve todas las filas si la consulta del usuario final no contiene ningún filtro. El puerto de filtro no es un puerto de transferencia.

Nota: El campo de filtro debe formar parte del grupo raíz en la solicitud de servicio web.

Cuando configure un puerto de filtro, puede identificar un puerto de salida en la transformación de consumidor del servicio web que recibe los datos de la columna desde la respuesta de servicio web. Por ejemplo, si el puerto de filtro es un puerto de entrada llamado EmployeeID, el puerto de salida de la respuesta

puede ser un puerto llamado EmployeeNum. Developer Tool debe asociar el puerto de entrada de filtro y un puerto de salida para insertar la lógica de filtro desde la lectura de tabla virtual con la solicitud del consumidor de servicio web. Los puertos de entrada para una solicitud de servicio web suelen ser distintos a los puertos de salida de la respuesta de servicio web.

El campo de filtro no puede ser un puerto de transferencia. Cuando configura un puerto de filtro, el valor predeterminado del puerto cambia al valor de la condición de filtro, por lo que el valor del puerto de salida de transferencia cambia. Un filtro basado en el puerto de transferencia de salida devuelve resultados inesperados.

Puede insertar varias expresiones de filtro para la transformación de consumidor de servicio web. Cada condición de filtro debe tener el siguiente formato:

```
<Field> = <Constant>
```

Las condiciones de filtro deben estar unidas por AND. No puede combinar las condiciones con un OR.

Optimización de inserción con el ejemplo de transformación del consumidor de servicio web

Un servicio de datos SQL devuelve pedidos de todos los clientes o devuelve pedidos de un cliente específico según la consulta SQL que reciba del usuario.

El servicio de datos contiene un objeto de datos lógicos con los siguientes componentes:

Tabla de clientes

Una tabla de base de datos Oracle que contiene la información de los clientes.

Transformación de consumidor de servicio web

Una transformación que llama a un servicio web para recuperar los últimos pedidos de los clientes. La transformación del consumidor de servicio web tiene puertos de entrada para customerID y orderNum. La transformación tiene puertos de transferencia que contienen datos de los clientes que recibe desde la tabla de clientes. El puerto orderNum es el puerto de filtro y no está conectado. orderNum tiene el valor predeterminado "*". Cuando el servicio web recibe este valor en la solicitud de servicio web, devuelve todos los pedidos.

Tabla virtual de pedidos

Una tabla virtual que recibe los datos de los clientes y los pedidos desde el servicio web. El usuario final consulta esta tabla. Los pedidos contienen una columna de cliente, una columna orderID y datos de clientes y pedidos.

El usuario final pasa la siguiente consulta SQL al servicio de datos SQL:

```
SELECT * de OrdersID donde cliente = 23 y orderID = 56
```

El servicio de integración de datos divide la consulta para optimizar la asignación. El servicio de integración de datos utiliza la optimización de primera selección y mueve la lógica de filtro, cliente = 23, a la lectura de tabla de clientes. El servicio de integración de datos utiliza la optimización de inserción e inserta la lógica de filtro, orderID = 56, en el puerto del filtro de la transformación del consumidor de servicio web. La transformación del consumidor de servicio web recupera ordersID 56 para clientes 23.

Habilitar la optimización de inserción con la transformación del consumidor de servicio web

Habilite la optimización de inserción para la transformación de consumidor de servicio web si la transformación no tiene efectos secundarios y no trata los fallos como errores.

1. Abra la vista **Propiedades avanzadas** de la transformación del consumidor de servicio web.

2. Desactive **Tratar fallo como error**.
3. Desactive **Tiene efectos secundarios**.
4. Haga clic en el botón **Abrir** de la propiedad **optimización de inserción**.
5. Seleccione el nombre del puerto de filtro en el cuadro de diálogo Entrada optimizada. Puede seleccionar varios puertos de filtro.
6. Haga clic en la columna **Salida**.
7. Para cada puerto de filtro, seleccione el puerto de salida que contiene la columna filtrada en la respuesta de servicio web.
8. Especifique un valor predeterminado para cada puerto de filtro.

Nota: No puede configurar un valor predeterminado para un puerto del consumidor de servicio web a menos que sea un puerto de filtro.

Crear una transformación de consumidor de servicio web

Puede crear una transformación de consumidor de servicio web reutilizable o no reutilizable. Pueden existir transformaciones reutilizables en varias asignaciones. Las transformaciones no reutilizables existen en una única asignación.

Puede crear transformaciones de consumidor del servicio web para un enlace SOAP 1.1 y SOAP 1.2 desde un solo objeto WSDL.

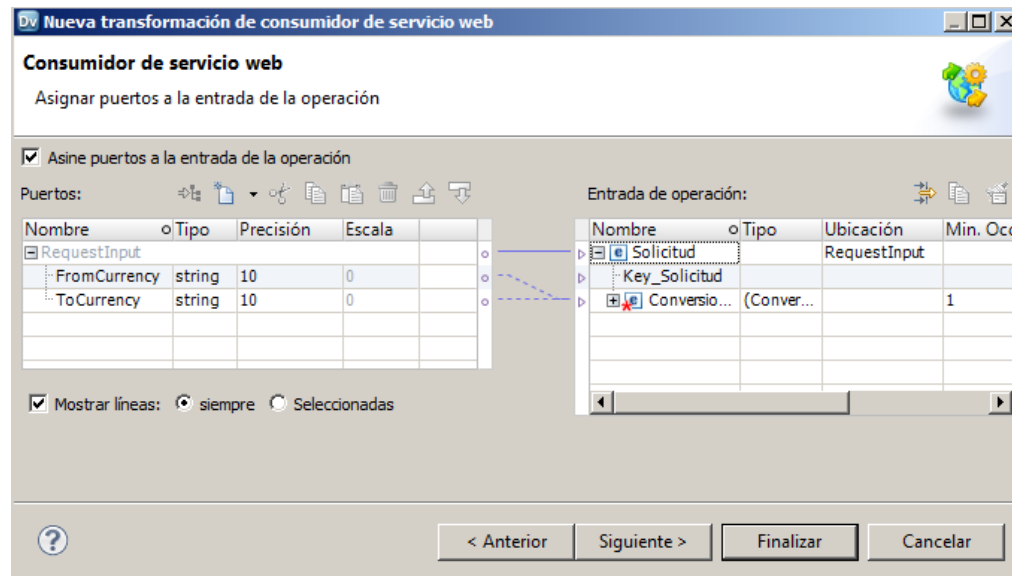
1. Para crear una transformación, utilice uno de los siguientes métodos:

Opción	Descripción
Reutilizable	Seleccione un proyecto o una carpeta en la vista Explorador de objetos . Haga clic en Archivo > Nuevo > Transformación . Seleccione la transformación de consumidor de servicio web y haga clic en Siguiente .
No reutilizable	En una asignación o en un mapplet, arrastre una transformación de consumidor de servicio web desde la paleta Transformación hasta el editor.

A continuación, aparece el cuadro de diálogo **Nueva transformación de consumidor de servicio web**.

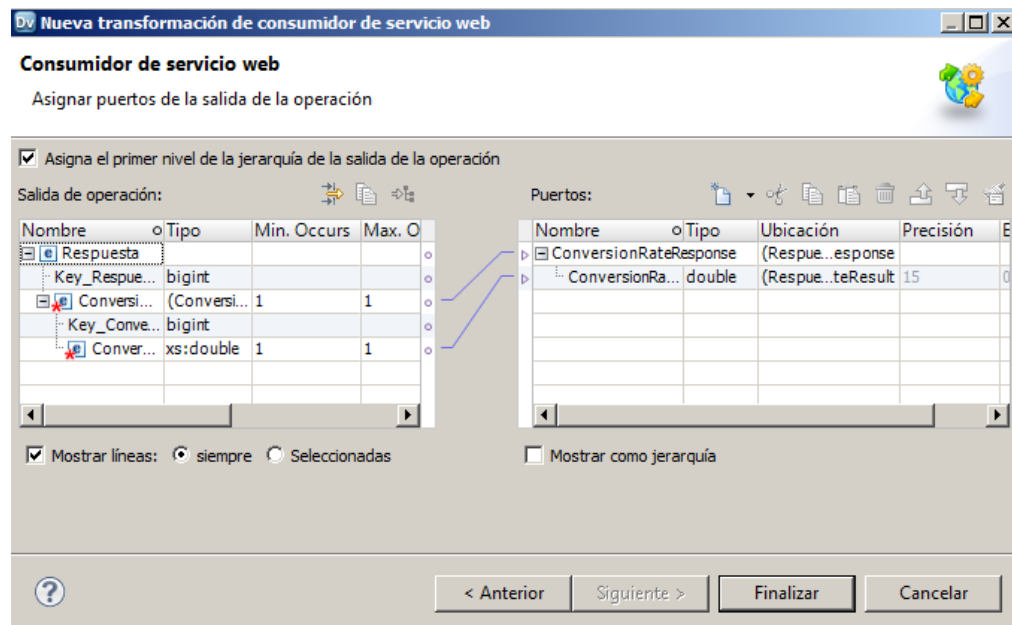
2. Explore y seleccione un objeto de datos WSDL para definir la solicitud de servicio web y los mensajes de respuesta.
Si el WSDL no se encuentra en el repositorio, puede importarlo desde el cuadro de diálogo Nueva transformación de consumidor de servicio web.
3. Explore y seleccione una operación del WSDL.
Puede elegir una operación que tenga un enlace SOAP 1.1 o 1.2.
4. Haga clic en **Siguiente**.

Aparece la pantalla **Asignar puertos a la entrada de la operación**. El área **Puertos** muestra los puertos de entrada y los grupos de entrada de transformación. El área **Entrada de operación** muestra la jerarquía de mensajes de solicitud.



5. Defina los grupos y los puertos de entrada y asigne los puertos de entrada a nodos de entrada de operación.
6. Haga clic en **Siguiente**.

Aparece la pantalla **Asignar puertos de la salida de la operación**. El área **Salida de operación** muestra la jerarquía de mensajes de respuesta. El área **Puertos** muestra los puertos de salida y los grupos de salida de transformación.



7. Defina los grupos y los puertos de salida y asigne nodos de salida de operación a los puertos de salida.
8. Haga clic en **Finalizar**.

9. Haga clic en la vista **Avanzadas** para configurar las propiedades de la transformación y la conexión de servicios web.

TEMAS RELACIONADOS

- [“Cómo asignar puertos de entrada a la entrada de operación” en la página 702](#)
- [“Cómo asignar la salida de operación a puertos de salida” en la página 706](#)
- [“Propiedades avanzadas de la transformación de consumidor de servicio web” en la página 707](#)

Ejemplo de transformación de consumidor de servicio web

Su organización necesita poner la información sobre pedidos de la línea de productos RT100 a disposición de la organización de ventas. El equipo de ventas necesita consultar a diario el resumen de los pedidos y los detalles de estos.

Cree un objeto de datos lógicos que exponga la información diaria sobre pedidos en tablas virtuales. La asignación de lectura contiene una transformación de consumidor de servicio web que devuelve los pedidos más recientes de RT100. La transformación de consumidor de servicio web consume un servicio web que devuelve el resumen diario de pedidos e información detallada de los pedidos para la línea de productos RT100.

Archivo de entrada

El archivo de entrada es un archivo sin formato que contiene el número de la línea de producto.

Cree un objeto de datos físicos para definir el archivo de entrada. El archivo tiene un campo, `Product_Line`. El valor del campo es RT100. Defina la ubicación del objeto de datos físicos en la vista **Propiedades de tiempo de ejecución**.

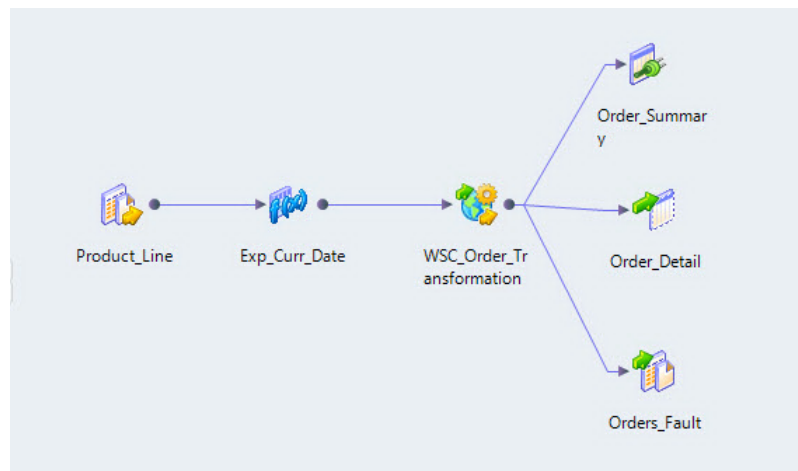
Modelo de objeto de datos lógicos

Un analista empresarial de su organización crea un modelo de datos lógicos que describe las estructuras del resumen de pedidos y de la tabla de detalles de pedidos. El modelo de datos lógicos contiene los objetos de datos lógicos `Order_Summary` y `Order_Detail`.

En analista crea un esquema en una herramienta de modelado que define el modelo de datos lógicos. El modelo de datos lógicos se importa desde el esquema y se crean los objetos de datos `Order_Summary` y `Order_Detail`.

Asignación de objetos de datos lógicos

La asignación de objetos de datos lógicos describe cómo acceder a datos mediante el objeto de datos lógicos.



La asignación de lectura contiene los siguientes objetos:

Product_Line

Archivo sin formato de entrada que contiene el número de línea de producto.

Transformación Exp_Curr_Date

Transformación de expresión que devuelve la fecha actual y una clave primaria para el grupo de entrada del nivel raíz de la transformación de consumidor de servicio web.

Transformación WSC_Order

Transformación de consumidor de servicio web que consume un servicio web para recuperar información sobre pedidos. La transformación pasa la línea de productos y la fecha actual al servicio web en el mensaje de solicitud. La transformación recibe la información sobre los pedidos desde un servicio web en el mensaje de respuesta.

Tabla Order_Summary

Un objeto de datos lógicos que contiene información sobre pedidos como Order_No, Customer_Id, Qty y Order_Date.

Tabla Order_Detail

Un objeto de datos lógicos que contiene información sobre pedidos como Order_No, Product_Id, Qty y Status.

Orders_Fault

Archivo sin formato de salida que recibe mensajes de fallo genéricos.

Transformación de consumidor de servicio web

La transformación de consumidor de servicio web recibe como entrada una línea de productos, una fecha y un número de secuencia. La transformación consume la operación de servicio web Get_Order_Info para recuperar la información del pedido.

Cuando se crea la transformación de consumidor de servicio web, seleccione un objeto de datos WSDL que describa los mensajes de solicitud y de respuesta del servicio web. Un mensaje de servicio web contiene

grupos jerárquicos de elementos XML. Un elemento puede contener otros elementos. Algunos elementos pueden ocurrir varias veces. Cree la transformación desde el objeto WSDL Order_Info que se encuentra en el repositorio.

Configure los puertos de entrada de transformación y asigne los puertos a la jerarquía de entrada de operación. Asigne nodos de la jerarquía de salida de operación a los puertos de salida. Defina la conexión de servicios web y las propiedades de tiempo de ejecución.

Asignación de entrada de transformación

Cuando se muestra la asignación de entrada en la vista **Puertos**, se pueden definir puertos de entrada y pueden asignarse a nodos en la entrada de operación.

El área **Puertos** de transformación tiene un grupo raíz y un grupo Pedido. El grupo raíz es el grupo de entrada Solicitud. Añada un puerto al grupo de entrada Solicitud para representar la clave primaria.

El grupo Pedido tiene los puertos de entrada **Select_Date** y **Select_Product_Line**.

Asigne los puertos de entrada a los nodos **Order_Date** y **Product_Line** en el área **Entrada de operación**.

El área **Entrada de operación** define el mensaje de solicitud que la transformación de consumidor de servicio web pasa al servicio web. De forma predeterminada, los nodos aparecen en el área **Entrada de operación**.

Asignación de salida de transformación

Cuando se muestra la asignación de salida en la vista **Puertos**, se pueden definir los puertos de salida asignando nodos de la salida de operación a los grupos de salida de transformación.

El servicio web devuelve la siguiente jerarquía en un mensaje de respuesta de un servicio web:

```
Response
  Orders
    Order
      Key_Order
      Order_ID
      Order_Date
      Customer_ID
      Total_Qty
      Order_Details
        Order_Detail
          Product_ID
          Description
          Qty
          Status
```

El servicio web devuelve múltiples pedidos. El pedido es un nodo de ocurrencia múltiple en el nivel Orders. Para cada pedido, el servicio web puede devolver múltiples detalles del pedido. Order_Detail es un nodo de ocurrencia múltiple en el nivel Order_Details.

Nota: Developer Tool añade el nodo Key_Order en la interfaz del usuario. Puede asignar la clave a grupos de salida para definir las relaciones entre los grupos. Para este ejemplo, Order_ID es la clave primaria en Order y es la clave externa en Order_Details.

Cree los siguientes grupos de salida en el área **Puertos**:

```
Order
  Order_ID
  Order_Date
  Customer_ID
  Total_Qty

Order_Detail
  Order_ID
  Product_ID
```

Description
Qty
Status

El servicio de integración de datos escribe una fila desde el grupo Order siempre que cambia el valor de Order_ID.

El servicio de integración de datos escribe una fila desde el grupo Order_Detail siempre que cambian los valores de Order_ID y Product_ID.

Propiedades avanzadas de la transformación

Configure las siguientes propiedades avanzadas para la transformación de consumidor de servicio web:

Habilitar control de fallos SOAP genéricos

Agrega puertos de salida que reciben mensajes de fallos SOAP.

Conexión

Seleccione una conexión de servicios web para acceder al servicio web.

Permitir compresión

La transformación de consumidor de servicio web comprime los mensajes web mediante GZIP.

CAPÍTULO 48

Análisis de mensajes SOAP del servicio web

Este capítulo incluye los siguientes temas:

- [Resumen del análisis de mensajes SOAP del servicio web, 720](#)
- [Interfaz de usuario de transformación, 721](#)
- [Configuración de salidas de ocurrencia múltiple, 722](#)
- [Cómo analizar elementos anyType, 724](#)
- [Cómo analizar tipos derivados, 725](#)
- [Análisis de elementos QName, 726](#)
- [Cómo analizar grupos de sustitución, 727](#)
- [Cómo analizar construcciones XML en mensajes SOAP, 727](#)

Resumen del análisis de mensajes SOAP del servicio web

El servicio de integración de datos genera datos de filas cuando analiza un mensaje SOAP en una transformación de servicio web.

La transformación de entrada de servicios web y la transformación de consumidor de servicio web son transformaciones de servicios web que analizan mensajes SOAP.

Para configurar una transformación para que analice un mensaje SOAP, cree puertos de salida en una estructura similar a la de la jerarquía de mensajes SOAP. Asigne los nodos de la jerarquía de mensajes SOAP a los puertos.

Puede configurar grupos normalizados de puertos de salida, grupos desnormalizados y grupos pivotados de puertos. Cuando el mensaje SOAP contiene tipos derivados, elementos anyType o grupos de sustitución, puede configurar grupos de salida diferentes según los tipos que pueden aparecer en la instancia del mensaje SOAP.

TEMAS RELACIONADOS

- [“Asignación de salida de transformación de consumidor de servicio web” en la página 704](#)

Interfaz de usuario de transformación

La transformación de consumidor de servicio web y la transformación de entrada de servicios web ofrecen una interfaz de usuario que puede utilizarse para asignar datos del mensaje SOAP a los puertos de salida de la transformación.

La siguiente figura muestra una asignación entre los nodos de mensaje SOAP 1.1 y los puertos de salida en una transformación de consumidor de servicio web:

The screenshot shows the 'Ports' window in Eclipse IDE. It has two panes. The left pane, titled 'Operation output', shows a tree of nodes from a SOAP message. The right pane, titled 'Ports', shows a table of output ports. Blue arrows indicate the mapping between nodes and ports.

Name	Type	Location	Precision	Scale
Company		(Response)/tns:Company		
Name	string	(Response)/tns:Name 10	10	0
fault1		(Fault)		
faultcode	string	(Fault)/faultcode	10	0
faultstring	string	(Fault)/faultstring	10	0
faultactor	string	(Fault)/faultactor	10	0

Área Operación

El área Operación contiene la jerarquía de mensajes SOAP. Los nodos complejos o los nodos de ocurrencia múltiple definen los niveles de jerarquía en la estructura. Developer Tool añade claves a los niveles que definen las relaciones principales-secundarias entre ellas.

En la figura anterior, la jerarquía de mensajes SOAP tiene los siguientes niveles:

Respuesta o Solicitud

Nivel que representa la raíz del mensaje de respuesta o de solicitud.

Empresa

Nivel superior de los datos de solicitud.

Departamentos

Departamentos de ocurrencia múltiple en el ámbito de la empresa.

Empleados

Empleado es un elemento complejo dentro de un departamento.

Grupo de fallos

Grupo de mensajes de fallo que recibe mensajes de error.

Área Puertos

Puede asignar datos desde niveles del mensaje SOAP a puertos de salida. Cada grupo de puertos de salida puede relacionarse con otros grupos de salida con relaciones de clave externa primaria.

En la figura anterior, la transformación posee grupos de puertos salida que corresponden a los grupos de nodos del mensaje SOAP.

Configuración de salidas de ocurrencia múltiple

Cuando una transformación de entrada o una transformación de consumidor de servicio web devuelve datos de ocurrencia múltiple, puede configurar los puertos de salida en configuraciones diferentes.

Puede configurar datos de salida normalizados, datos de salida pivotados o datos de salida desnormalizados.

Por ejemplo, un mensaje SOAP contiene los elementos complejos Departamentos y Empleados. Cada departamento contiene varios empleados. Departamentos es un elemento principal de Empleados.

El mensaje SOAP contiene la siguiente jerarquía de elementos:

```
Departments
  Department_ID
  Department_Name
  Employees
    Employee_ID
    Employee_Name
```

Salida relacional normalizada

Cuando se crean datos de salida normalizados, los valores de los datos no se repiten en un grupo de salida. Se crea una relación exclusiva entre los niveles de jerarquía del mensaje SOAP y los grupos de salida de los puertos.

Cuando el mensaje SOAP contiene un nivel de jerarquía principal Departamentos y un nivel de jerarquía secundario Empleados, puede crear los siguientes grupos de puertos:

```
Departments
  Department_Key
  Department_ID
  Department_Name

Employees
  Department_Key
  Employee_ID
  Employee_Name
```

Department_Key es una clave generada que relaciona el grupo de salida Empleados con un grupo Departamento.

Claves generadas

Cuando se agrega un grupo de salida, Developer Tool relaciona el grupo de salida con otro grupo de salida con una clave generada. Developer Tool añade una clave bigint al grupo principal y al grupo secundario. En tiempo de ejecución, el servicio de integración de datos crea los valores de clave para las claves generadas.

Ejemplo

La jerarquía de mensajes SOAP tiene los siguientes nodos:

```
Departments
  Dept_Key
  Dept_Num
  Dept_Name

Employees
  Dept_FK
  Employee_Num
  Employee_Name
```

Cuando se crea un grupo de salida de puertos para Departamentos, se asigna el nodo Departamentos a un campo vacío del área Puertos. Developer Tool crea el siguiente grupo de salida:

```
Departments
  Dept_Num
  Dept_Name
```

Cuando se asigna el nodo Empleados a un campo vacío del área Puertos, Developer Tool le pedirá que relacione el grupo Empleados con el grupo Departamentos. Puede relacionar el grupo Empleados con más de un grupo. Developer Tool añade una clave a cada grupo.

Developer Tool crea los siguientes grupos y claves generadas:

```
Departments
  Key_Departments
  Dept_Num
  Dept_Name

Employees
  Key_Departments
  Employee_Num
  Employee_Name
```

Nota: No es necesario que asigne nodos a las claves generadas. El servicio de integración de datos crea los valores de clave en tiempo de ejecución.

Developer Tool puede crear claves generadas en varios niveles de un grupo de salida. El grupo Empleados puede contener los siguientes puertos:

```
Employees
  Key_Employees
  Key_Departments
  Key_Managers
  Employee_Num
  Employee_Name
```

Key_Departments y Key_Managers son las claves generadas que apuntan a los grupos principales. Key_Employees es una clave generada para el grupo Empleados. Key_Employees aparece cuando se relaciona un grupo secundario con el grupo Empleados.

Salida relacional desnormalizada

Se puede desnormalizar una salida relacional. Cuando se desnormalizan los datos de salida, los valores de elemento del grupo principal se repiten para cada elemento secundario.

Para desnormalizar datos de salida, asigne nodos del nivel de jerarquía principal al grupo secundario de los puertos de salida.

El siguiente ejemplo muestra Department_ID y Department_Name en el grupo de salida Empleados:

```
Employees
  Department_ID
  Department_Name
```

```
Employee_ID
Employee_Name
```

Department_ID y Department_Name se repiten para cada empleado del departamento:

Department_ID	Department_Name	Employee_ID	Employee_Name
100	Contabilidad	56500	Kathy Jones
100	Contabilidad	56501	Tom Lyons
100	Contabilidad	56509	Bob Smith

Salida relacional pivotada

Puede incluir un número específico de elementos de ocurrencia múltiple en un grupo de salida.

Para pivotar elementos de ocurrencia múltiple, asigne el elemento secundario de ocurrencia múltiple al grupo principal de los puertos de salida. Developer Tool le insta a definir el número de elementos secundarios que incluir en el principal.

El siguiente ejemplo muestra dos instancias de Employee_ID en el grupo principal Departamentos:

```
Departments
  Department_ID
  Department_Name
  Employee_ID1
  Employee_ID2
```

Cómo analizar elementos anyType

El elemento anyType representa una elección de todos los tipos globales en un WSDL o esquema. Cuando se asignan nodos a puertos en Developer Tool, se eligen los tipos que deben aparecer en el mensaje SOAP para el elemento anyType. Debe reemplazar un elemento anyType en el mensaje SOAP con un tipo complejo o xs:string. Cree grupos de puertos para cada tipo que elija.

Debe elegir un tipo para asignar datos a puertos de salida. Si el WSDL o el esquema no contiene un tipo de global, Developer Tool reemplaza el elemento anyType con xs:string.

Para elegir un tipo de elemento complejo en el área Operación, haga clic en **Elegir** en la columna **Tipo** del elemento anyType. Aparece una lista de tipos complejos disponibles y xs:string.

Cuando se reemplaza un elemento anyType con tipos derivados, el servicio de integración de datos llena uno a uno los elementos para un tipo. El mensaje SOAP no contiene datos para el tipo base y el tipo derivado al mismo tiempo.

Ejemplo de tipos derivados

El WSDL contiene un elemento anyType. Se reemplaza el elemento con AddressType y un tipo derivado llamado USAddressType. La jerarquía de mensajes SOAP tiene los siguientes grupos:

```
Address:AddressType (base type)
  Address: AddressType
    Street
    City

Address:USAddressType (derived type)
  Street
  City
  State
```

ZipCode

El mensaje SOAP contiene los siguientes datos:

```
<address xsi:type="AddressType">
  <street>1002 Mission St.</street>
  <city>san jose</city>
</address>

<address xsi:type="USAddressType">
  <street>234 Fremont Blvd</street>
  <city>Fremont</city>
  <zip>94556</zip>
  <state>CA</state>
</address>
```

El servicio de integración de datos devuelve una fila para xsi: AddressType:

Calle	Ciudad
1002 Mission St.	San José

El servicio de integración de datos devuelve una fila para el tipo derivado xsi: USAddressType:

Calle	Ciudad	Estado	Código postal
234 Fremont Blvd.	Sunnyvale	CA	94556

El servicio de integración de datos no llena AddressType si el tipo es xsi: USAddressType.

Cómo analizar tipos derivados

Puede analizar mensajes SOAP que contienen tipos derivados. Cuando se definen los puertos que reciben los datos desde el mensaje SOAP, elija los tipos que pueden aparecer en el mismo. Los elementos en los tipos que se eligen determinarán los puertos que necesita crear.

Por ejemplo, el WSDL puede contener un AddressType y un tipo derivado llamado USAddressType. Puede crear los siguientes grupos en el área Operación de Developer Tool:

```
Address
  Address: AddressType
  Street
  City

Address:USAddressType
  Street
  City
  State
  ZipCode
```

El mensaje SOAP puede contener los siguientes datos:

```
<address>
  <street>1002 Mission St.</street>
  <city>san jose</city>
</address>

<address xsi:type="USAddressType">
  <street>234 Fremont Blvd</street>
  <city>Fremont</city>
  <zip>94556</zip>
```

```

<state>CA</state>
</address>

<address xsi:type="USAddressType">
<street>100 Cardinal Way</street>
<city>Redwood City</city>
<zip>94536</zip>
<state>CA</state>
</address>

<address>
<street>100 El Camino Real</street>
<city>Sunnyvale</city>
</address>

```

El servicio de integración de datos devuelve las siguientes filas para el tipo base, Address:

Calle	Ciudad
1002 Mission St.	San José
234 Fremont Blvd	Sunnyvale
100 Cardinal Way	Redwood City
100 El Camino Real	Sunnyvale

El servicio de integración de datos devuelve las siguientes filas para el tipo base, USAddress:

Calle	Ciudad	Estado	Código postal
234 Fremont Blvd.	Sunnyvale	CA	94556
100 Cardinal Way	Redwood City	CA	94536

El servicio de integración de datos devuelve todas las direcciones en el tipo base. El servicio de integración de datos devuelve direcciones de EE.UU. en el tipo derivado. El tipo derivado incluye los elementos Calle y Ciudad que USAddressType hereda del tipo base.

Análisis de elementos QName

Cuando el servicio de integración de datos analiza elementos QName en el mensaje SOAP, actualiza valores QName que pertenecen al espacio de nombres del esquema para utilizar el prefijo de espacio de nombres definido en el esquema. De lo contrario, el servicio de integración de datos no actualiza el valor del elemento.

Por ejemplo, el esquema tiene el prefijo de espacio de nombres `tns` definido para el espacio de nombres `"http://usuario/prueba"`. El mensaje SOAP tiene el prefijo de espacio de nombres `mytns` definido para el mismo espacio de nombres. Cuando el servicio de integración de datos analiza el valor QName `mytns:myelement`, el valor cambia a `tns:myElement`.

Cuando el servicio de integración de datos genera elementos QName en el mensaje SOAP, no actualiza el valor del elemento.

Cómo analizar grupos de sustitución

Un grupo de sustitución reemplaza un elemento por otro del mismo grupo. Los grupos de sustitución son similares a los tipos derivados, excepto que cada definición de elemento incluye un nombre de grupo de sustitución.

Puede configurar un grupo de salida de puertos que recibe elementos de un tipo específico en un grupo de sustitución. Puede crear un grupo de salida de puertos diferente que recibe un elemento desde otro tipo en el grupo de sustitución.

Cómo analizar construcciones XML en mensajes SOAP

Un mensaje SOAP puede contener construcciones XML como elementos de selección, de lista y de unión.

Con ciertas limitaciones, las transformaciones de servicio web pueden analizar mensajes SOAP con estas construcciones.

Elemento de selección

Un elemento de selección restringe un elemento secundario a uno de los elementos en la declaración <selección>.

El siguiente texto muestra un elemento persona que es un empleado o un contratista:

```
<xs:element name="person">
  <xs:complexType>
    <xs:choice>
      <xs:element name="employee" type="employee"/>
      <xs:element name="contractor" type="contractor"/>
    </xs:choice>
  </xs:complexType>
</xs:element>
```

Puede asignar elementos de selección mediante los siguientes métodos:

- Crear puertos de salida para cada elemento de selección en un grupo de salida. Algunos elementos tendrán valores nulos en la fila de salida.
- Crear un grupo de salida para cada selección. Para el ejemplo anterior, cree un grupo de empleados y un grupo de contratistas. El servicio de integración de datos genera una fila basada en el elemento que aparece en el mensaje SOAP.

Elemento de lista

Una lista es un elemento XML que puede contener varios valores de tipo simple, como "Lunes Martes Miércoles".

El servicio de integración de datos puede devolver una lista como un valor de cadena. Cuando el mensaje SOAP contiene una lista, no se pueden asignar elementos de la lista a filas de salida individuales. Puede configurar una transformación de expresión para separar los elementos de la lista si necesita separarlas en una asignación.

Elemento de unión

El elemento de unión es un tipo simple que es una unión de más de un tipo.

El siguiente texto muestra un elemento Tamaño que es la unión de dos tipos simples, size_no y size_string:

```
<xs:element name="Size">
  <xs:simpleType>
    <xs:union memberTypes="size_no size_string" />
  </xs:simpleType>
</xs:element>
```

Para asignar el tamaño a un puerto de salida, cree un puerto para el tamaño. Configure el puerto de salida como una cadena. Puede configurar otra transformación en la asignación para convertir los datos a otro tipo.

CAPÍTULO 49

Generación de mensajes SOAP del servicio web

Este capítulo incluye los siguientes temas:

- [Resumen de la generación de mensajes SOAP del servicio web, 729](#)
- [Interfaz de usuario de transformación, 730](#)
- [Relaciones entre puerto y nivel de jerarquía , 731](#)
- [Claves, 732](#)
- [Asignar puertos, 734](#)
- [Cómo pivotar puertos de ocurrencia múltiple , 736](#)
- [Asignación de datos desnormalizados, 737](#)
- [Tipos derivados y sustitución de elementos, 739](#)
- [Cómo generar construcciones XML en mensajes SOAP, 740](#)

Resumen de la generación de mensajes SOAP del servicio web

El servicio de integración de datos genera datos XML a partir de grupos de datos de entrada cuando genera un mensaje SOAP. Cuando se crea una transformación de consumidor de servicio web, una transformación de salida de servicio web o una transformación de fallo, se configuran los puertos de entrada que deben asignarse a la jerarquía de mensajes SOAP.

Para configurar una transformación para que genere un mensaje SOAP, cree grupos de puertos de entrada y asigne cada grupo a un grupo de la jerarquía de mensajes SOAP. Un WSDL o esquema define la estructura del mensaje SOAP.

Puede configurar grupos de datos en el mensaje SOAP a partir de datos de entrada desnormalizados. También puede pivotar datos de entrada de ocurrencia múltiple en nodos de ocurrencia múltiple en los mensajes SOAP.

Puede asignar datos a tipos derivados, elementos anyType o grupos de sustitución en un mensaje SOAP. Debe elegir qué tipos pueden aparecer en el mensaje SOAP cuando se define una transformación. Los tipos que elija determinarán los puertos de entrada que necesita crear.

Cuando se visualiza la jerarquía de mensajes SOAP en Developer Tool, la jerarquía contiene claves. Las claves no aparecen en el mensaje SOAP. El servicio de integración de datos utiliza claves para definir

relaciones entre grupos principales y secundarios en el mensaje SOAP. Para configurar valores clave, asigne datos de entrada a claves en el mensaje SOAP.

TEMAS RELACIONADOS

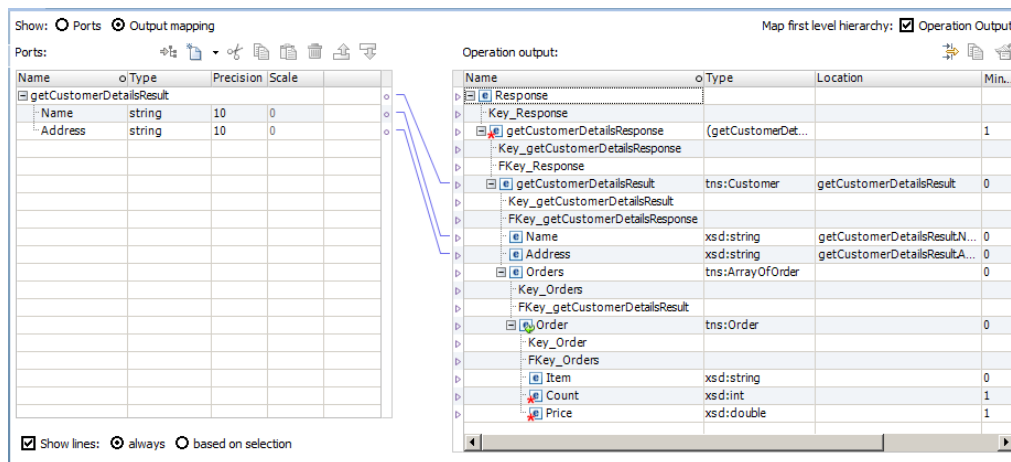
- [“Asignación de entrada de transformación de consumidor de servicio web” en la página 701](#)

Interfaz de usuario de transformación

La transformación de salida, la transformación de fallo y la transformación de consumidor de servicio web de un servicio web contienen una interfaz de usuario que permite configurar el mensaje SOAP.

Para configurar una transformación para que genere un mensaje SOAP, cree puertos de entrada en una estructura similar a la de la jerarquía de mensajes SOAP. El WSDL o esquema determina la estructura de la jerarquía. Asigne cada puerto de entrada a un nodo en el mensaje SOAP.

La siguiente figura muestra una asignación entre los puertos de entrada y los nodos de mensaje SOAP en una transformación de salida de servicio web:



Área Puertos de entrada

Cree grupos de puertos de entrada en el área **Puertos de entrada**. Incluya puertos de entrada para cada nivel en la jerarquía de mensajes SOAP que desee asignar.

Debe crear un grupo de entradas Respuesta o Solicitud, así como los grupos secundarios que reciben los datos.

Cuando cree grupos de puerto de entrada, defina una clave primaria en cada grupo principal. Defina una clave externa en cada grupo secundario. La clave externa relaciona el grupo con un grupo principal.

No es necesario que defina claves para el nivel Respuesta o el nivel raíz WSDL salvo que esté pasando datos en el nivel raíz WSDL. Por ejemplo, el nivel raíz puede contener encabezados HTTP.

Puede crear grupos de puertos similares a los siguientes grupos para clientes y pedidos:

```

Response
  Response_Key

  Customer_Details_Root
  
```

```

Key_Cust_Det
FK_Response_Key

Customer
  Customer_ID
  FK_Cust_Det
  Name
  Address

Orders
  Order_Num
  FK_Cust_ID

Order_Items
  Order_Num
  Item
  Count
  Price

```

Área Operación

El área **Operación** muestra los elementos en la jerarquía de mensajes SOAP definidos por el WSDL o el esquema. El mensaje SOAP no tiene que contener todos los elementos del WSDL o del esquema. El mensaje contiene los datos que se asignan de los puertos de entrada.

Los nodos de ocurrencia múltiple y los nodos complejos definen los niveles de jerarquía en la estructura de mensajes SOAP. Developer Tool añade claves a los niveles para crear relaciones jerárquicas entre ellos. Todos los niveles de la jerarquía, excepto los niveles hoja, poseen una clave primaria. Cada nivel secundario tiene una clave externa que apunta a un nivel principal. Las claves que aparecen en la jerarquía de mensajes SOAP no aparecen en la instancia de mensaje SOAP. El servicio de integración de datos necesita que las claves tengan valores para relacionar los niveles de los datos cuando genera el mensaje SOAP.

La columna **Ubicación** contiene el nombre de grupo y el puerto de entrada que contiene los datos para un elemento en el mensaje SOAP. La columna **Ubicación** está vacía hasta que se asigna un puerto de entrada al nodo.

En la figura precedente, el mensaje SOAP contiene una instancia individual de detalles y pedidos de cliente. El grupo Pedidos contiene un elemento de ocurrencia múltiple llamado Pedido. La jerarquía de mensajes SOAP tiene los siguientes niveles que se relacionan mediante una clave:

```

Response
  GetCustomerDetailsResponse
    GetCustomerDetailsResult
      Orders
        Order

```

El nivel Respuesta representa la raíz del mensaje de respuesta. El servicio de integración de datos requiere que este nivel anexe encabezados al mensaje SOAP.

El nivel GetCustomerDetailsResponse es la raíz del mensaje.

Relaciones entre puerto y nivel de jerarquía

Cuando se asignan puertos de entrada a la jerarquía de mensajes SOAP, se mantiene una relación entre un grupo de entrada y un nivel de la jerarquía de mensajes SOAP. Por ejemplo, puede tener dos grupos de entrada, Departamento y Empleado.

El grupo de entrada Departamento recibe las siguientes filas:

Dept_num	Nombre	Ubicación
101	HR	Nueva York
102	Producto	California

El grupo de entrada Empleado recibe las siguientes filas:

Dept_num	Empleado
101	Alice
101	Bob
102	Carol
102	Dave

Asigne el número de departamento en el grupo Empleado como clave externa que establezca la relación entre el grupo Departamento y el grupo Empleado. El número de departamento aparece en el nivel de jerarquía de departamento, pero no el nivel de empleado.

El mensaje SOAP contiene la siguiente estructura XML:

```
<department>
  <dept_num>101</dept_num>
  <name>HR</name>
  <location>New York</location>

  <employee>
    <name>Alice</name>
  </employee>

  <employee>
    <name>Bob</name>
  </employee>
</department>

<department>
  <dept_num>102</dept_num>
  <name>Product</name>
  <location>California</location>

  <employee>
    <name>Carol</name>
  </employee>

  <employee>
    <name>Dave</name>
  </employee>
</department>
```

Claves

Una jerarquía de mensajes SOAP incluye claves. El servicio de integración de datos necesita claves para construir la jerarquía XML en el mensaje SOAP.

Debe asignar datos de puertos de entrada a las claves en la jerarquía de mensajes SOAP. Asigne datos a las claves en cada nivel al que suministre datos. Cuando tenga un nodo de ocurrencia múltiple, tendrá que relacionar el nodo con un nodo principal.

Las claves aparecen en el mensaje SOAP sin tipos. Cualquier puerto que se asigna a una clave debe ser del tipo de datos string, integer o bigint. La clave primaria del grupo principal y la clave externa de cada grupo

secundario deben tener el mismo tipo de datos, la misma precisión y la misma escala. Puede asignar las claves generadas a las claves de mensajes SOAP.

Puede asignar un puerto a un nodo y a una clave que estén en el mismo nivel de jerarquía. Por ejemplo, puede asignar Employee_ID a un nodo en el mensaje SOAP y, además, puede asignarlo a una clave en el nivel Empleado.

Cuando los nodos de dos grupos de la jerarquía posean una relación principal-secundario, realice las siguientes tareas:

- Asigne un puerto a la clave primaria en el grupo del nodo principal.
- Asigne un puerto a la clave externa en el grupo del nodo secundario.

También puede asignar claves primarias a puertos de entrada para eliminar filas con una clave primaria con el valor null o con claves primarias duplicadas.

Puede crear una clave compuesta en un mensaje SOAP mediante la asignación de varios puertos a la misma clave. Utilice claves compuestas cuando tenga que desnormalizar datos y mantener claves únicas para ciertas combinaciones de valores de ocurrencia múltiple. Puede crear claves compuestas que contengan valores de cadena, de bigint o de entero.

Nota: Puede incluir una transformación de expresión en la asignación de operación para generar valores de clave.

Ejemplo de clave compuesta

Configure una clave de división-departamento única a partir de los siguientes grupos de puertos:

```
Company
  Company_Num
  Company_Name

  Division
    Company_Num
    Division_Num
    Division_Name

    Department
      Division_Num
      Dept_Num
      Dept_Name
      Location
```

Dept_Num es exclusivo dentro de una división, pero no lo es para todas las divisiones de una empresa.

Puede configurar un grupo Departamento que contenga la información sobre la división y el departamento. Configure el número de división y el número de departamento como parte de la clave compuesta:

```
Department
  Division_Num + Dept_Num (key)
  Dept_Name
  Location
```

El orden en el que se asignan los puertos determina el valor de la clave.

Asignar puertos

Después de crear puertos de entrada, asigne cada uno de ellos a la jerarquía de mensajes SOAP. La ubicación del puerto aparece junto al nodo en el área **Operación**.

Puede asignar puertos a los siguientes tipos de nodos:

Nodo atómico

Un elemento o atributo simple que no tiene elementos o atributos secundarios y que no es divisible.

Nodo atómico de ocurrencia múltiple

Un elemento o atributo simple que ocurre varias veces en la misma ubicación de la jerarquía.

Nodo complejo

Un elemento que contiene otros elementos.

Si el nodo principal no tiene una ubicación, el nodo principal recibe como ubicación el nombre del grupo de entrada. Cuando el nodo principal tiene una ubicación, cada nodo del nivel jerárquico debe tener una ubicación de salida desde la misma ubicación.

Se puede asignar un nombre de grupo de entrada a un nodo principal en un nivel de jerarquía. Developer Tool actualiza el campo de ubicación para el nodo principal en la jerarquía. Developer Tool no actualiza los nodos secundarios que pertenecen al grupo en la jerarquía. Cuando se asignan puertos de entrada a los nodos secundarios, la ubicación de cada puerto de entrada debe ser la misma que la del nodo principal.

Después de asignar un grupo de entrada a un nivel de jerarquía, se puede modificar más adelante. Puede hacer clic en **Borrar** o puede eliminar las líneas entre las áreas Puertos y Operación. Para eliminar las líneas, arrastre el puntero de las líneas para seleccionarlas. Haga clic en **Eliminar**.

Asignación de un puerto

Cuando se asigna un puerto a un nodo en el mensaje SOAP, Developer Tool proporciona resultados diferentes según el tipo de nodo al que se asigna el puerto.

En la siguiente tabla se describen los resultados que se obtienen cuando se asigna un puerto individual a nodos de destino diferentes en el área **Operación**:

Nodo de destino	Resultados
Nodo atómico	Cuando se asigna un puerto individual a un nodo y el nodo principal no tiene ninguna ubicación, el nodo recibe la ubicación del puerto. La ubicación del nodo principal recibe la ubicación del grupo de entrada para el puerto individual. Cuando se asigna un puerto individual a un nodo y el nodo principal ya posee una ubicación, puede cambiar la ubicación del nodo principal y borrar la ubicación de los demás nodos secundarios situados en el mismo nivel. La ubicación del nivel de jerarquía cambia al nombre de grupo del puerto.
Nodo atómico de ocurrencia múltiple o la clave primaria del nodo atómico de ocurrencia múltiple	Cuando se asigna un puerto individual al nodo atómico de ocurrencia múltiple, Developer Tool establece la ubicación para el nodo atómico en el grupo del puerto seleccionado.
Nodo complejo	Cuando se asigna un puerto individual a un nodo complejo, Developer Tool establece la ubicación del nodo complejo como ubicación del grupo que contiene el puerto. Developer Tool le solicita el nodo atómico de ocurrencia múltiple al que asignar el puerto. Si todos los nodos atómicos de ocurrencia individual poseen una ubicación, no podrá asignar el nodo complejo.

Asignación de un grupo

Cuando se asigna un grupo de entrada a un nodo en el mensaje SOAP, Developer Tool proporciona distintos resultados según el tipo de nodo al que se asigna el puerto.

En la siguiente tabla se describen los resultados que se obtienen cuando se asigna un grupo a un nodo en el área **Operación**:

Nodo de destino	Resultados
Nodo atómico	No se puede asignar un grupo a un nodo atómico.
Nodo atómico de ocurrencia múltiple	Se le insta a que elija un puerto en el grupo de entrada para actualizar la ubicación para el nodo y la clave primaria.
Nodo complejo de ocurrencia múltiple	Developer Tool define la ubicación para el nodo complejo como la ubicación del grupo.

Asignación de varios puertos

Cuando se asignan varios puertos a un nodo en el mensaje SOAP, Developer Tool proporciona resultados distintos según el tipo de nodo al que se hayan asignado los puertos. Puede asignar varios puertos al mismo tiempo si los asigna desde el mismo grupo.

En la siguiente tabla se describen los resultados para el nodo cuando se asignan varios puertos a nodos:

Nodo Destino	Resultados
Nodo atómico individual	Cuando se asignan varios puertos a un nodo individual, se actualiza la ubicación para más de un nodo atómico individual en el área Operación . Si la jerarquía no tiene suficientes nodos en el nivel para actualizarlos, Developer Tool solo asigna puertos para los nodos disponibles.
Nodo atómico de ocurrencia múltiple	Cuando se asignan varios puertos a un nodo atómico de ocurrencia múltiple, se pivotan los puertos en múltiples ocurrencias del nodo. Developer Tool crea instancias del nodo según el número de puertos que se asignan. Aparece un mensaje que describe el número de puertos que se han proyectado.
Nodo complejo de ocurrencia múltiple	Cuando se asignan varios puertos a un nodo complejo, debe seleccionar los nodos atómicos de ocurrencia individual que deben actualizarse. Se pivotan los puertos en múltiples ocurrencias del nodo. Developer Tool crea instancias del nodo según el número de puertos que se asignan.

Cómo pivotar puertos de ocurrencia múltiple

Puede asignar varios puertos de entrada a un nodo de ocurrencia múltiple en el mensaje SOAP. Developer Tool pivota los datos de entrada en múltiples nodos en el mensaje SOAP.

Para cambiar el número de elementos que deben pivotarse, seleccione **Reemplazar pivote existente** en el cuadro de diálogo **Opciones de asignación**.

Si elimina una de las instancias de puerto pivotado del área **Puertos**, Developer Tool elimina todas las instancias del área **Operación**.

Ejemplo de pivote

Un grupo de entrada puede tener las siguientes filas:

Núm.	Nombre	Ubicación	emp_name1	emp_name2	emp_name3
101	HR	Nueva York	Alice	Tom	Bob
102	Producto	California	Carol	Tim	Dave

Cada fila contiene un número de departamento y tres nombres de empleado.

Empleado es un nodo de ocurrencia múltiple en la jerarquía de mensajes SOAP. Puede asignar todas las instancias de Empleado de la fila de entrada a la jerarquía de mensajes SOAP. Seleccione todas las instancias de Empleado. Haga clic en **Asignar**. El cuadro de diálogo **Opciones de asignación** le pedirá que seleccione un nodo de la lista.

Developer Tool cambia el nodo Empleado para incluir los múltiples nodos de nombre en la jerarquía de mensajes SOAP:

```
Department
  num
  name
  location
  Employee (unbounded)
```



```
emp_name1  
emp_name2  
emp_name3
```

El mensaje SOAP devuelve la siguiente jerarquía:

```
<department>  
  <num>101</num>  
  <name>HR</name>  
  <location>New York</location>  
  <employee>  
    <emp_name>Alice</name>  
  </employee>  
  <employee>  
    <emp_name>Tom</name>  
  </employee>  
  <employee>  
    <emp_name>Bob</name>  
  </employee>  
</department>  
  
<department>  
  <num>102</num>  
  <name>Product</name>  
  <location>California</location>  
  <employee>  
    <emp_name>Carol</name>  
  </employee>  
  <employee>  
    <emp_name>Tim</name>  
  </employee>  
  <employee>  
    <emp_name>Dave</name>  
  </employee>  
</department>
```

Asignación de datos desnormalizados

Puede asignar datos desnormalizados y pasarlos a nodos normalizados en un mensaje SOAP.

Cuando se asignan datos desnormalizados, se pasan datos de un grupo de entrada a varios nodos en la jerarquía de mensajes SOAP. Puede crear relaciones de grupo en el mensaje SOAP similares a los siguientes tipos de relaciones:

Relación de nodos lineal

El nodo A es el nodo principal del nodo B. El nodo B es un nodo principal del nodo C. El nodo C es el nodo principal del nodo D.

Relación de nodos jerárquica

El nodo A es un nodo principal del nodo B. El nodo A también es el nodo principal del nodo C. Los nodos B y C no están relacionados.

La siguiente tabla muestra filas de entrada que contienen datos desnormalizados de división y de departamento:

División	Dept_Num	Dept_Name	Teléfono	Employee_Num	Employee_Name
01	100	Contabilidad	3580	2110	Amir
01	100	Contabilidad	3580	2113	Robert
01	101	Ingeniería	3582	2114	Stan

División	Dept_Num	Dept_Name	Teléfono	Employee_Num	Employee_Name
01	101	Ingeniería	3582	2115	Jim
02	102	Instalaciones	3583	2116	José

Los datos de entrada contienen números y nombres únicos de empleados. Los datos de división y de departamento se repiten para cada empleado de un mismo departamento y división.

Relación de grupos lineal

Cuando se configuran puertos, puede configurar un grupo aparte para División, Departamento y Empleado. División es un grupo principal de Departamento, y Departamento es el grupo principal de Empleado. Puede configurar grupos en la siguiente estructura lineal:

```

Division
  Division_Key
  Division_Num
  Division Name

  Department
    Department_Key
    Division_FKey
    Dept_Num
    Dept_Name
    Phone

  Employee
    Department_Fkey
    Employee_Num
    Employee_Name

```

El mensaje SOAP contiene instancias únicas de División y de Departamento, aunque Division_Num y Dept_Num se repiten en los datos de entrada. Defina Division_Num como clave primaria en el grupo División. Defina Dept_Num como clave primaria en el grupo Departamento.

Relación de grupos jerárquica

Puede crear una jerarquía de grupo que contenga el grupo principal División y los grupos secundarios Departamento y Empleado. Departamento y Empleado no mantienen una relación clave primaria-clave externa. Departamento y Empleado son grupos secundarios de División. Puede configurar los grupos en la siguiente estructura:

```

Division
  Division_Key
  Division_Num
  Division_Name

  Department
    Division_FKey
    Dept_Num
    Dept_Name

  Employee
    Division_FKey
    Employee_Num
    Employee_Name

```

Tipos derivados y sustitución de elementos

Puede asignar puertos de entrada a tipos derivados complejos, elementos anyType y grupos de sustitución en un mensaje SOAP. El mensaje puede incluir elementos para el tipo base y los tipos derivados.

En una relación de tipos, el tipo base es el tipo desde el cual se deriva otro tipo. Un tipo derivado hereda elementos desde el tipo base. Un tipo complejo extendido es un tipo derivado que hereda elementos desde un tipo base e incluye elementos adicionales. Un tipo complejo restringido es un tipo derivado que restringe algunos elementos del tipo base.

Cómo generar tipos derivados

Cuando un WSDL o esquema incluye tipos derivados, debe elegir los tipos que desea incluir en el mensaje SOAP.

Por ejemplo, el WSDL define un tipo base AddressType. El WSDL también contiene un USAddressType y un UKAddressType que son los AddressTypes derivados.

Cada tipo contiene los siguientes elementos:

- AddressType: calle, ciudad
- USAddressType (extiende AddressType): estado, zipCode
- UKAddressType (extiende AddressType): postalCode, país

Cuando se elige un USAddressType en el área Operación, Developer Tool crea un grupo para el elemento USAddressType en el mensaje SOAP. El grupo incluye la calle y la ciudad desde la dirección base y el estado y el zipCode para el USAddress. Los tipos derivados que extienden tipos base siempre incluyen los elementos desde el tipo base.

Si elige todos los tipos derivados disponibles para el mensaje SOAP, Developer Tool crea grupos similares a los siguientes en la jerarquía SOAP:

```
Address
  Address: Address
    Street
    City

Address:USAddressType
  Street
  City
  State
  ZipCode

Address: UKAddressType
  Street
  City
  PostalCode
  Country
```

Necesitará definir grupos de puertos de entrada para Address, USAddress y UKAddress.

Cómo generar elementos anyType y atributos

Ciertos elementos de esquema y atributos permiten que haya cualquier tipo de datos en un mensaje SOAP.

El elemento anyType representa una selección de todos los tipos conocidos globalmente. Antes de asignar un puerto a un elemento anyType en un mensaje SOAP, elija un tipo complejo disponible o xs:string. Si el WSDL o el esquema no contienen un tipo complejo, Developer Tool reemplaza el tipo de elemento anyType con xs:string.

Para elegir un tipo de elemento complejo en el área Operación, haga clic en **Elegir** en la columna **Tipo** del elemento anyType. Aparece una lista de tipos complejos disponibles y xs:string.

Los siguientes elementos y atributos permiten cualquier tipo de datos:

elemento anyType

Permite que un elemento sea de cualquier tipo de datos en el archivo XML asociado.

elemento anySimpleType

Permite que un elemento sea cualquier simpleType en el archivo XML asociado.

CUALQUIER elemento de contenido

Permite que un elemento sea cualquier elemento global definido en el esquema.

atributo anyAttribute

Permite que un elemento sea cualquier atributo ya definido en el esquema.

Cómo generar grupos de sustitución

Utilice grupos de sustitución para sustituir un elemento por otro en un mensaje SOAP. Los grupos de sustitución funcionan de forma similar a los tipos derivados, excepto que las definiciones de elemento incluyen un nombre de grupo de sustitución.

Por ejemplo, puede tener un tipo base Dirección y los tipos derivados USAddress y UKAddress:

```
xs:element name="Address" type="xs:string"/>
<xs:element name="USAddress" substitutionGroup="Address"/>
<xs:element name="UKAddress" substitutionGroup="Address"/>
```

Cuando se configura la jerarquía de mensajes SOAP, puede elegir qué elemento sustituir por Dirección en el mensaje SOAP.

Cómo generar construcciones XML en mensajes SOAP

Un WSDL o un esquema puede contener elementos de selección, de lista o de unión. Las transformaciones de servicio web pueden generar mensajes SOAP que contengan estos elementos.

Elemento de selección

Un elemento de selección restringe un elemento secundario a uno de los elementos en la declaración <selección>.

Para asignar puertos a un mensaje SOAP que contenga elementos de selección, cree un grupo de entrada que incluya todos los elementos en la construcción de selección. Por ejemplo, una descripción de un artículo puede ser una dimensión o un peso:

```
item: description, choice {dimension, weight}
```

Cuando la descripción es una dimensión, la descripción es un tipo complejo que contiene la longitud, la anchura y la altura.

Cuando la descripción es un peso, la descripción es un tipo de cadena simple.

Los datos de entrada tienen las siguientes columnas y filas:

descripción	longitud	anchura	altura	peso
caja	20 cm	18 cm	15 cm	NULL
café	NULL	NULL	NULL	500 g

El mensaje SOAP contiene un grupo de artículos que contiene dimensiones o descripciones de pesos:

```
Item
  Description
  Dimension
    Length
    Width
    Height
  Weight
```

Los valores NULL en los datos de entrada se convierten en elementos ausentes en la salida XML.

El mensaje SOAP contiene los siguientes datos:

```
<item>
  <desc>box</desc>
  <dimension>
    <length>20cm</length>
    <width>18cm</width>
    <height>15cm</height>
  </dimension>
</item>

<item>
  <desc>coffee</desc>
  <weight>500g</weight>
</item>
```

Elemento de lista

Una lista es un elemento XML que puede contener varios valores de tipo simple en el mismo elemento o atributo. El servicio de integración de datos puede procesar una lista en los datos de entrada si la lista se representa como una cadena de datos consolidados.

Si cada artículo de la lista es un elemento individual como ClassDates1, ClassDates2 y ClassDates3, el servicio de integración de datos no puede procesar los elementos como una lista. Puede utilizar una transformación de expresión para combinarlos en una cadena si necesita devolver una lista en un mensaje SOAP.

Las siguientes filas de entrada contienen un elemento de lista llamado ClassDates que contiene los días de la semana:

CourseID	Nombre	ClassDates
Matemáticas 1	Álgebra básica	Lun Mié Vie
Historia 1	Historia del mundo	Mar Jue

El servicio de integración de datos puede devolver un mensaje SOAP con la siguiente estructura XML:

```
<class>
  <courseId>Math 1</courseId>
  <name>Beginning Algebra</name>
  <classDates>Mon Wed Fri</classDates>
</class>
<class>
  <courseId>History 1</courseId>
  <name>World History</name>
  <classDates>Tue Thu</classDates>
</class>
```

Elemento de unión

El elemento de unión es un tipo simple que una unión de más de un tipo. Cuando un mensaje SOAP contiene un elemento de unión, debe asignar un puerto de entrada individual que contenga los datos en una cadena.

Por ejemplo, el mensaje SOAP contiene un elemento llamado tamaño. Tamaño es una unión de entero y cadena:

```
<xs:element name="size">
  <xs:simpleType>
    <xs:union memberTypes="size_no size_string" />
  </xs:simpleType>
</xs:element>
```

Las cadenas de entrada contienen artículos con una descripción y un tamaño. Un artículo puede tener un tamaño numérico, como 42. Asimismo, un artículo puede tener un tamaño que es un valor de cadena, como grande, mediano o pequeño.

La siguiente tabla muestra filas de entrada con un tamaño numérico y tamaño de cadena:

Desc	Tamaño
zapatos	42
camisa	grande

Cree un puerto para el tamaño del artículo. Asigne el puerto como una cadena. El mensaje SOAP contiene los siguientes elementos:

```
<item>
  <desc>shoes</desc>
  <size>42</size>
</item>

<item>
  <desc>shirt</desc>
  <size>large</size>
</item>
```

CAPÍTULO 50

Transformación de media ponderada

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de media ponderada, 743](#)
- [Cómo configurar una transformación de media ponderada, 743](#)
- [Ejemplo de puntuaciones de coincidencia ponderadas, 744](#)
- [Propiedades avanzadas de Transformación de media ponderada, 744](#)
- [Transformación de media ponderada en un entorno no nativo, 745](#)

Resumen de la transformación de media ponderada

La transformación de media ponderada es una transformación pasiva que lee puntuaciones de coincidencia de varias coincidencias de operaciones y produce una sola puntuación de coincidencia.

Puede aplicar una ponderación numérica a cada puntuación que entra en la transformación de media ponderada. Una ponderación es un valor entre cero y 1. Puede editar la ponderación aplicada a cada puntuación de entrada para ampliar o reducir su contribución a la puntuación de salida. Aplique ponderaciones que reflejen la importancia relativa de cada columna de datos en el análisis de duplicados.

Utilice la transformación de media ponderada cuando añada transformaciones de comparación a una asignación o mapplet.

Nota: También puede asignar ponderaciones en una transformación de coincidencia. Utilice la transformación de coincidencia para configurar estrategias de búsqueda de coincidencias y asignar ponderaciones en una transformación individual. Puede incrustar un mapplet de coincidencia en una transformación de coincidencia.

Cómo configurar una transformación de media ponderada

Utilice la transformación de media ponderada para ajustar el total de puntuación de coincidencia que una asignación genera para una serie de operaciones de análisis de coincidencias. Edite las ponderaciones relativas de cada puerto de entrada para reflejar las prioridades de las comparaciones de datos que ha

definido para el conjunto de datos de origen. Cada puerto de entrada en la transformación de media ponderada representa una puntuación de coincidencia de salida de una estrategia de transformación de comparación.

Los pasos siguientes describen el proceso para configurar una transformación de media ponderada no reutilizable en un mapplet o asignación que utiliza transformaciones de comparación.

1. Abra un mapplet o una asignación de análisis de coincidencias y añada una transformación de media ponderada con dirección descendente en relación con las transformaciones de comparación.
2. Conecte la salida de puntuación de una transformación de comparación a un puerto de entrada de media ponderada.
Repita este paso para otras transformaciones de comparación en el mapplet o asignación.
3. Seleccione la ficha **Puertos** en la transformación de media ponderada.
4. Haga doble clic en el campo **Ponderación** de cada entrada y especifique un valor de ponderación entre 0,001 y 1. El valor de ponderación debe reflejar la importancia relativa de la puntuación de entrada cuando se compara con otras entradas en la transformación.
5. Guarde el mapplet o asignación.

Ejemplo de puntuaciones de coincidencia ponderadas

Cree una asignación de análisis de coincidencias para determinar el número de nombres de cliente duplicados en una base de datos de clientes. Puede añadir dos transformaciones de comparación para generar puntuaciones de coincidencia para las columnas `Código postal` y `Apellido` en el conjunto de datos.

Muchos registros tienen códigos postales coincidentes, aunque un número de registros mucho menor tiene apellidos coincidentes. Cuando se promedian estas puntuaciones de coincidencia se debe enfatizar la importancia de las coincidencias más destacadas.

Para enfatizar la importancia de puntuaciones de coincidencia de apellido, puede aplicar una mayor ponderación la puntuación de coincidencia de `Apellido`.

Por ejemplo, establezca el valor de **Ponderación** de la entrada de puntuación de `Apellido` en 0,8, y establezca el valor de **Ponderación** de la entrada de puntuación de `Código postal` en 0,4.

Propiedades avanzadas de Transformación de media ponderada

Configure las propiedades para ayudarle a determinar cómo el servicio de integración de datos procesa los datos para la Transformación de media ponderada.

Puede configurar niveles de seguimiento de registros.

Configure la siguiente propiedad en la ficha **Avanzadas**:

Nivel de seguimiento

Cantidad de detalles que aparecen en el registro de esta transformación. Puede elegir simplificado, normal, inicialización detallada o datos detallados. El valor predeterminado es Normal.

Transformación de media ponderada en un entorno no nativo

El procesamiento de transformaciones de media ponderada en un entorno no nativo depende del motor que ejecuta la transformación.

Tenga en cuenta la compatibilidad de los siguientes motores en tiempo de ejecución no nativos:

- Motor de Blaze. Compatible sin restricciones.
- Motor de Spark. Compatible sin restricciones en asignaciones por lotes. No compatible en asignaciones de transmisión.
- Motor de Databricks Spark. Compatible sin restricciones.

CAPÍTULO 51

Transformación de ventana

Utilice la transformación de ventana cuando desee acumular datos transmitidos en grupos de datos y, a continuación, procesar los conjuntos de datos. La transformación de ventana es una transformación pasiva.

Cuando se lee desde orígenes ilimitados, es posible que desee acumular los datos en grupos de datos delimitados para su posterior procesamiento. Para introducir intervalos delimitados en datos sin límites, utilice una transformación de ventana.

Cuando configure una transformación de ventana, defina el tipo de ventana y los límites de datos por tiempo. Para especificar los límites de datos, configure el tamaño de la ventana y el intervalo de deslizamiento de la ventana. El tamaño de la ventana define el intervalo de tiempo para el cual los datos se acumulan como un grupo de datos. El intervalo de deslizamiento define el intervalo de tiempo después del cual el grupo de datos acumulados se procesa aún más. El retraso de marca de agua define el tiempo de umbral para que un evento retrasado se acumule en un grupo de datos.

La transformación de ventana solo se puede ejecutar para las asignaciones de transmisión en el motor de Spark.

Para obtener más información acerca de la transformación de ventana, consulte la *Guía del usuario de Data Engineering Streaming*.

CAPÍTULO 52

Transformación de escritura

Este capítulo incluye los siguientes temas:

- [Resumen de la transformación de escritura, 747](#)
- [Propiedades de la transformación de escritura , 747](#)
- [Crear una transformación de escritura, 754](#)

Resumen de la transformación de escritura

Una transformación de escritura es una transformación pasiva. La asignación utiliza la transformación de escritura para escribir datos en un destino. Las transformaciones de escritura no se pueden reutilizar.

Las transformaciones de escritura se pueden crear a partir de un objeto de datos físicos, un objeto de datos lógicos o un parámetro. Si desea crear una transformación de escritura a partir de un objeto de datos físico que haya importado desde un origen de adaptador de PowerExchange, el editor de asignaciones podría solicitarle que especifique una operación de escritura antes de poder crear una transformación de escritura desde el objeto de datos.

Las propiedades de una transformación de escritura que se pueden configurar dependen del tipo de objeto de datos que haya utilizado para crear la transformación.

Las transformaciones de escritura pueden representar destinos dinámicos. Puede configurar una transformación de escritura de forma que actualice dinámicamente sus puertos, metadatos y otras propiedades. Para obtener más información sobre cómo configurar destinos dinámicos, consulte la *Guía de asignación de Informatica Developer*.

Propiedades de la transformación de escritura

Después de crear una transformación de escritura, puede configurar sus propiedades.

Configure las propiedades de la transformación de escritura en las fichas de la vista **Propiedades** de la transformación. Las fichas que se pueden configurar dependen del tipo de destino que la transformación de escritura representa.

En la siguiente tabla, se describe cada ficha de propiedad y se indica el tipo de destino para el que se utiliza cada ficha:

Ficha de propiedad	Descripción	Tipo de destino
General	Sirve para especificar las propiedades y el comportamiento de la transformación. En los orígenes de objetos de datos relacionales y personalizados, sincronice los puertos de entrada de la transformación con el origen.	Todo
Objeto de datos	Sirve para especificar el origen de datos de la transformación. En los orígenes de objetos de datos relacionales y personalizados, puede obtener las columnas de objetos de datos del origen de datos en tiempo de ejecución.	Archivo sin formato Relacional Objeto de datos personalizados
Formato	Sirve para introducir la configuración de un origen de datos de archivo sin formato.	Archivo sin formato
Puertos	Sirve para establecer la definición de puertos en función del objeto de datos asociado o del flujo de asignación.	Todo
Tiempo de ejecución	Contiene las propiedades que utiliza el servicio de integración de datos al escribir datos en el destino en tiempo de ejecución (por ejemplo, dónde enviar los archivos de rechazo). En un destino de archivo sin formato, rechace los nombres y los directorios del archivo.	Archivo sin formato Relacional
Parámetros de objeto de datos	Muestra los parámetros del objeto de datos. Permite configurar los valores de los parámetros de la asignación o enlazar los parámetros a los parámetros de una asignación.	Archivo sin formato Objeto de datos personalizados Objetos de datos lógicos
Vinculación en tiempo de ejecución	Permite crear nuevos vínculos en tiempo de ejecución o ver las propiedades de los vínculos.	Todo
Avanzado	Sirve para establecer el nivel de seguimiento y el orden de fila. En un destino relacional o de Hive, establezca la estrategia de esquema de destino.	Archivo sin formato Relacional Objetos de datos personalizados Objetos de datos lógicos

Propiedades generales

Puede configurar el nombre y la descripción de la transformación de escritura. También puede configurar las siguientes propiedades:

Cuando cambien los metadatos de columna

Disponible para destinos relacionales y personalizados. Seleccione una de las siguientes opciones:

- Sincronizar puertos de entrada. Developer tool actualiza los puertos de entrada de la transformación de escritura con los cambios de metadatos que el repositorio de modelos almacena en relación con el objeto de datos.
- No sincronizar. Developer tool no refleja los cambios de metadatos en el objeto de datos.

El valor predeterminado es la opción Sincronizar puertos de entrada.

Objeto de datos físicos

Disponible para destinos de archivo sin formato y personalizados. Es el objeto que se ha utilizado para crear la transformación.

Puede seleccionar el nombre del objeto de datos y configurar sus propiedades.

Propiedades del objeto de datos

En la ficha Objeto de datos, puede especificar o cambiar el destino de la transformación de escritura, así como convertir en dinámicos los destinos de objetos de datos relacionales y personalizados.

Puede configurar las siguientes propiedades:

Especificar por

Para especificar los metadatos y columnas de destino de la transformación de escritura, seleccione una de las siguientes opciones:

- **Valor.** La transformación de escritura utiliza el objeto de datos asociado para especificar los metadatos y columnas de destino.
- **Parámetro.** La transformación de escritura utiliza un parámetro para especificar los metadatos y columnas de destino.

La opción predeterminada es Valor.

Objeto de datos

Si ha creado la transformación de escritura a partir de un objeto de datos existente, el campo mostrará el nombre de ese objeto. Haga clic en **Examinar** para cambiar el objeto de datos y asociarlo a la transformación de escritura.

Parámetro

Seleccione o cree un parámetro para asociarlo a la transformación de escritura.

En tiempo de ejecución, obtener las columnas de objetos de datos del origen de datos

Si habilita esta opción, el servicio de integración de datos obtiene los cambios de definición de datos y metadatos de las tablas de destino para la transformación de escritura.

Para obtener una vista previa del modo en que el servicio de integración de datos obtiene los cambios de definición de datos y metadatos, vea la asignación con los parámetros resueltos.

Propiedades de puertos

En la ficha Puertos, puede configurar las siguientes propiedades:

Columnas definidas por

Seleccione una de las siguientes opciones para definir las columnas de la transformación de escritura:

- **Objeto de datos asociado.** Utilice nombres de columna, metadatos y otras propiedades del objeto de datos en la ficha Objeto de datos.
- **Flujo de asignación.** La asignación obtiene los nombres de columna, metadatos y otras propiedades de los objetos de nivel superior en la asignación.

El valor predeterminado es la opción Objeto de datos asociado.

Propiedades de recursos de columna

Disponible para destinos de archivo sin formato y personalizados. El recurso de cada columna es el objeto de datos del que la columna obtiene el nombre, los metadatos y otras propiedades. Seleccione el nombre del recurso para cambiar las propiedades de recurso.

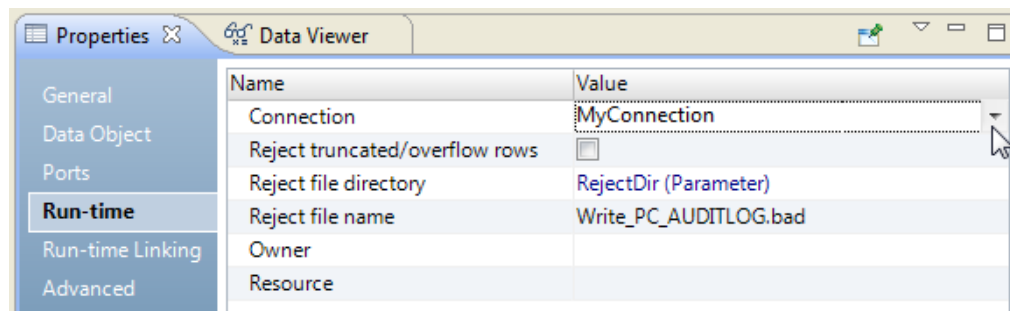
Propiedades de tiempo de ejecución

Puede configurar las siguientes propiedades de transacción de escritura en la ficha Tiempo de ejecución:

Conexión

Disponible en los destinos relacionales. Conexión que utiliza la transformación. Haga clic a la derecha del campo para cambiar la conexión.

La siguiente imagen muestra la ubicación del botón desplegable donde debe hacer clic:



Rechazar filas truncadas/con desbordamiento

Disponible para destinos relacionales y personalizados.

Developer tool permite convertir datos pasándolos de puerto a puerto. A veces, una conversión hace que los datos numéricos se desborden o que las cadenas se trunquen en las columnas que contienen caracteres. Por ejemplo, pasar datos de un puerto Decimal (28, 2) a otro Decimal (19, 2) provocará un desbordamiento numérico. De modo similar, si se pasan datos de un puerto String(28) a otro String(10), el Servicio de integración de datos truncará las cadenas a 10 caracteres.

Si una conversión provoca un desbordamiento, el Servicio de integración de datos omite la fila de forma predeterminada. El Servicio de integración de datos no escribe los datos en el archivo de rechazo. En el caso de las cadenas, el Servicio de integración de datos las trunca y las traslada a la siguiente transformación.

Seleccione esta opción para incluir en el archivo de rechazo de la sesión todos los datos truncados y desbordados entre la última transformación y el destino. El Servicio de integración de datos envía todas las filas truncadas y cualquier fila con desbordamiento al archivo de rechazo de la sesión o a los registros de errores de fila, según cuál sea la configuración de la sesión.

Directorio de archivos de rechazo

Directorio donde se encuentra el archivo de rechazo. El valor predeterminado es el parámetro del sistema RejectDir.

Nombre de archivo de rechazo

El nombre del archivo de rechazo. El valor predeterminado es <output_file_name>.bad.

Si hay varias particiones que escriben en el destino del archivo sin formato, cada partición escribe en un archivo de rechazo independiente llamado <output_file_name><partition_number>.bad.

Propiedades de la vinculación en tiempo de ejecución

Cree y configure un vínculo en tiempo de ejecución en la ficha **Vinculación en tiempo de ejecución**. Un vínculo en tiempo de ejecución es un vínculo de grupo a grupo entre transformaciones que usan un parámetro, una directiva de vínculo o ambos, a fin de saber qué puertos vincular en tiempo de ejecución. Los vínculos en tiempo de ejecución aparecen como líneas gruesas en el editor de asignación.

Cree y configure vínculos en tiempo de ejecución a una transformación de escritura en los siguientes casos:

- El objeto de datos de destino de la transformación de escritura utiliza un parámetro.
- Los puertos de la transformación de nivel superior pueden cambiar en tiempo de ejecución.

Nota: No cree un vínculo en tiempo de ejecución a una transformación de escritura si define las columnas de destino según el flujo de asignación.

Puede realizar las siguientes tareas en la ficha **Vinculación en tiempo de ejecución**:

Crear un vínculo de tiempo de ejecución.

En el área **Vínculos**, haga clic en el botón **Nuevo** y, en el cuadro de diálogo Nuevo vínculo, seleccione la transformación desde la que quiera vincular puertos a la transformación de escritura en tiempo de ejecución.

Configurar las propiedades del vínculo en tiempo de ejecución.

En el área **Propiedades del vínculo**, configure las siguientes propiedades de vínculo en tiempo de ejecución:

Parámetro

Seleccione esta opción cuando los nombres de puerto pueden cambiar entre las ejecuciones de asignación y conoce los valores de nombre de puerto. Utilice un parámetro de tipo Conjunto de vínculos de entrada para conectar puertos por sus valores de nombre entre ejecuciones de asignación. La sintaxis del parámetro de asignación Conjunto de vínculos de entrada consta de pares de puertos separados por comas: `Afield1->Bfield2, Afield3->Bfield4` .

Directiva de vínculo

Seleccione esta opción cuando quiera vincular automáticamente puertos con el mismo nombre. Por ejemplo, si los dos objetos de asignación contienen un puerto llamado SALARY, el Servicio de integración de datos los vinculará. En los nombres de puerto, los prefijos y sufijos pueden ignorarse.

En tiempo de ejecución, el Servicio de integración de datos establece y resuelve los vínculos entre los puertos en el siguiente orden:

- Vínculos creados manualmente en el editor de asignación.
- Vínculos basados en el parámetro configurado para un vínculo en tiempo de ejecución.
- Vínculos basados en la directiva de vínculo configurada para un vínculo en tiempo de ejecución.

Para obtener más información sobre los vínculos en tiempo de ejecución, consulte la *Guía de asignación de Informática Developer*.

Propiedades avanzadas

Configure propiedades avanzadas para determinar la manera en que el servicio de integración de datos procesa los datos de la transformación de escritura.

Configure las siguientes propiedades en la ficha **Avanzadas**:

Nivel de seguimiento

Controle la cantidad de detalle incluido en el archivo de registro de asignación.

Tipo de carga de destino

Tipo de la carga de destino. Seleccione Normal o Masiva. Puede establecer el tipo de carga de destino de los recursos relacionales o de los objetos de datos personalizados.

Si selecciona Normal, el servicio de integración de datos carga los destinos normalmente. Puede elegir Masiva cuando cargue los destinos a DB2, Sybase, Oracle o Microsoft SQL Server. Si especifica Masiva para otros tipos de base de datos, el servicio de integración de datos regresa a la opción de carga normal. La carga masiva puede aumentar el rendimiento de las asignaciones, aunque limita la capacidad de recuperación ya que no se realiza ningún registro de base de datos. Al escribir en un destino Oracle con carga masiva, puede optimizar el rendimiento deshabilitando las restricciones en la base de datos Oracle.

Seleccione el modo Normal si la asignación contiene una transformación de estrategia de actualización. Si elige Normal y el nombre del destino de Microsoft SQL Server incluye espacios, configure el siguiente SQL de entorno en el objeto de conexión:

```
SET QUOTED_IDENTIFIER ON
```

Reemplazo de actualización

Reemplaza la instrucción UPDATE predeterminada del destino.

Eliminar

Elimina todas las filas marcadas para su eliminación.

Esta propiedad está habilitada de manera predeterminada.

Nota: El motor de Databricks Spark omite esta propiedad cuando escribe en un destino de Delta Lake. Para eliminar filas, utilice una transformación de estrategia de actualización.

Insertar

Inserta todas las filas marcadas para su inserción.

Esta propiedad está habilitada de manera predeterminada.

Nota: El motor de Databricks Spark omite esta propiedad cuando escribe en un destino de Delta Lake.

Estrategia de esquema de destino

Tipo de estrategia de esquema de destino de la tabla de destino relacional o de Hive.

Puede seleccionar una de las siguientes estrategias de esquema de destino:

- **RETAIN: Mantener esquema de destino existente.** El servicio de integración de datos conserva el esquema de destino actual.
- **CREATE: Crear o reemplazar la tabla en tiempo de ejecución.** El servicio de integración de datos elimina la tabla de destino en tiempo de ejecución y la reemplaza por una tabla basada en la tabla de destino que indique.
- **Asignar parámetro.** Puede asignar un parámetro que represente el valor de la estrategia de esquema de destino y, a continuación, cambiar el parámetro en tiempo de ejecución.

Consulta DDL para crear o reemplazar

Crea o reemplaza la tabla de destino en tiempo de ejecución según una consulta DDL que haya definido. Aplicable cuando se selecciona la opción de estrategia de esquema de destino **CREATE: Crear o reemplazar la tabla en tiempo de ejecución.**

Truncar tabla de destino

Trunca el destino antes de cargar los datos.

Esta propiedad está habilitada de manera predeterminada.

Truncar la partición de destino

Trunca un destino de Hive particionado interno o externo antes de que cargue datos. Debe elegir **Truncar tabla de destino** para poder seleccionar esta opción.

De manera predeterminada, la opción está deshabilitada.

Estrategia de actualización

La estrategia de actualización de las filas existentes. Puede seleccionar una de las siguientes estrategias:

- Actualizar como actualización. El servicio de integración de datos actualiza todas las filas marcadas para la actualización.
- Actualizar como inserción. El servicio de integración de datos inserta todas las filas marcadas para la actualización. También debe seleccionar la opción **Insertar** destino.

Nota: Si el destino es un Delta Lake de Databricks, no es necesario que seleccione **Insertar**.

- Actualizar o insertar. El servicio de integración de datos actualiza las filas marcadas para la actualización si existen en el destino y luego inserta todas las filas restantes marcadas para la inserción. También debe seleccionar la opción **Insertar** destino.

Nota: Si el destino es un Delta Lake de Databricks, no es necesario que seleccione **Insertar**.

PreSQL

Comando SQL que el servicio de integración de datos ejecuta en las bases de datos de destino antes de leer el origen.

Developer tool no valida el SQL.

PostSQL

Comando SQL que el servicio de integración de datos ejecuta en la base de datos de destino después de escribir en el destino.

Developer tool no valida el SQL.

Mantener orden de filas

Mantenga el orden de las filas de los datos de entrada en el destino. Seleccione esta opción si el servicio de integración de datos no debe realizar ninguna optimización que pueda cambiar el orden de las filas.

Cuando el servicio de integración de datos realiza optimizaciones, puede hacer que se pierda el orden de las filas establecido anteriormente en la asignación. Puede establecer el orden de filas en una asignación con un origen de archivo sin formato ordenado, un origen relacional ordenado o una transformación de ordenación. Al configurar un destino para mantener el orden de las filas, el servicio de integración de datos no realiza optimizaciones en el destino.

Restricciones

Instrucciones SQL para las restricciones de integridad referencial en el nivel de tabla. Solo se aplica a los destinos relacionales.

Crear una transformación de escritura

Al crear una transformación de escritura, debe elegir uno de los siguientes métodos en función del recurso a partir del que se crea la transformación:

Crear la transformación de un objeto de datos en el repositorio de modelos.

Utilice este método si desea basar los metadatos de la transformación de escritura en un objeto de datos.

Para crear una transformación de escritura en el editor de asignación, utilice el asistente para **nueva transformación de escritura**, o bien arrastre un objeto de datos desde la vista del **Explorador de objetos** y seleccione Escritura como el tipo de transformación.

Crear la transformación desde el flujo de la asignación de objetos.

Utilice este método para basar los puertos de la transformación de escritura en el flujo de metadatos de las transformaciones de asignación de nivel superior.

Crear la transformación utilizando un parámetro.

Utilice este método si desea que la transformación de escritura herede los puertos del objeto al que el parámetro representa.

Crear la transformación utilizando otra transformación de la asignación.

Utilice este método si desea que la transformación de escritura tenga los mismos puertos que la transformación de origen.

Crear una transformación de escritura a partir de un objeto de datos

Puede crear y configurar una transformación de escritura en el editor de asignación.

Conviene usar este método si desea configurar propiedades específicas de la transformación. Por ejemplo, puede crear la transformación y, a continuación, configurarla para obtener metadatos de columnas del origen de datos en tiempo de ejecución.

1. Haga clic con el botón derecho en el editor de asignación y seleccione **Añadir transformación**.

Se abre el cuadro de diálogo **Añadir transformación**.

2. Seleccione **Escritura** y haga clic en **Aceptar**.

Se abre el asistente para **nueva transformación de escritura**.

3. Seleccione **Objeto de datos físicos** u **Objeto de datos lógicos** y haga clic en **Examinar**.

Se abre el cuadro de diálogo **Seleccionar objeto de datos**.

4. Seleccione un origen de datos y, a continuación, haga clic en **Aceptar**.

5. Para definir las columnas de la transformación de escritura según el objeto de datos asociado, seleccione **Objeto de datos asociado**.

Los puertos de la transformación de escritura son los mismos que las columnas del objeto de datos asociado.

6. Para que las columnas del objeto de destino se actualicen dinámicamente en tiempo de ejecución con los cambios producidos en el archivo de destino, seleccione **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos**.

El Servicio de integración de datos actualizará los metadatos de columna de la transformación de escritura cuando se ejecute la asignación.

7. Haga clic en **Finalizar**.

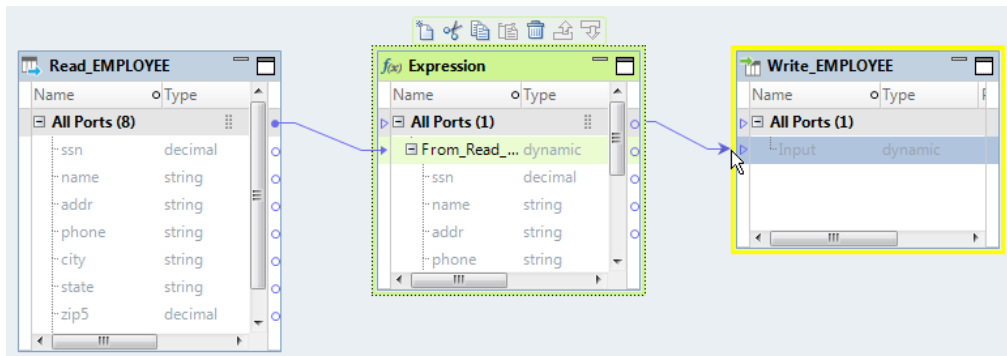
Crear una transformación de escritura a partir de un flujo de asignación

Puede crear y configurar una transformación de escritura en el editor de asignación.

Conviene usar este método si desea configurar propiedades específicas de la transformación. Por ejemplo, puede crear la transformación y, a continuación, configurarla para definir las columnas según el flujo de asignación u obtener los metadatos de columnas del origen de datos en tiempo de ejecución.

1. Haga clic con el botón derecho en el editor de asignación y seleccione **Añadir transformación**.
Se abre el cuadro de diálogo **Añadir transformación**.
2. Seleccione **Escritura** y haga clic en **Aceptar**.
Se abre el asistente para **nueva transformación de escritura**.
3. Seleccione **Objeto de datos físicos** u **Objeto de datos lógicos** y haga clic en **Examinar**.
Se abre el cuadro de diálogo **Seleccionar objeto de datos**.
4. Seleccione un origen de datos y haga clic en **Aceptar**.
5. Seleccione **Flujo de asignación** y haga clic en **Finalizar**.
6. Arrastre puertos de Todos los puertos del nivel superior al puerto **Entrada** de la transformación de escritura.

La siguiente imagen muestra cómo conectar puertos de nivel superior al puerto **Entrada** de la transformación de escritura:



La transformación de escritura recibe las definiciones de columna de los objetos de asignación de nivel superior.

7. Para que las columnas del objeto de destino se actualicen dinámicamente en tiempo de ejecución con los cambios producidos en el archivo de destino, seleccione **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos**.

El Servicio de integración de datos actualizará los metadatos de columna de la transformación de escritura cuando se ejecute la asignación.

Crear una transformación de escritura a partir de un parámetro

Un parámetro es un valor constante que se puede cambiar entre ejecuciones de asignaciones.

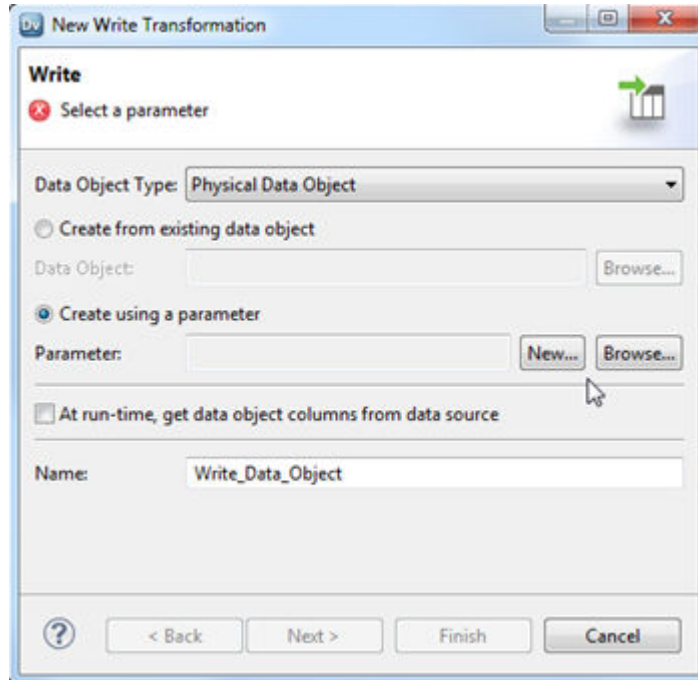
En una asignación dinámica, puede utilizar parámetros para cambiar los orígenes y los destinos. También puede utilizarlos en reglas de entrada, en reglas de selección, en vínculos en tiempo de ejecución y en propiedades de la transformación. Al cambiar el valor del parámetro, el Servicio de integración de datos crea o vuelve a crear el destino según el valor especificado en dicho parámetro.

1. Haga clic con el botón derecho en el editor y seleccione **Añadir transformación**.

2. Seleccione Escritura en la lista de las transformaciones y, a continuación, haga clic en **Aceptar**.

Sugerencia: Escriba la primera letra de la transformación para filtrar la lista.

Se abre el cuadro de diálogo Nueva transformación de escritura.



3. Seleccione **Crear usando un parámetro**.
4. Examine para seleccionar el parámetro, o bien haga clic en **Nuevo** para crear un parámetro y, luego, examine para seleccionar un objeto de datos a partir del cual crear el parámetro.
5. Opcionalmente, seleccione la opción **En tiempo de ejecución, obtener columnas de objetos de datos del origen de datos** para actualizar las columnas de la transformación en tiempo de ejecución.
6. Opcionalmente, haga clic en **Siguiente** para elegir cómo definir las columnas de transformación.
7. Haga clic en **Finalizar**.

Crear una transformación de escritura a partir de una transformación existente

Puede crear una transformación de escritura con los mismos puertos en una transformación que existe en la asignación.

Puede crear un destino para archivos complejos, archivos sin formato o recursos relacionales. Si uno de los puertos de la transformación existente contiene un puerto complejo, debe crear un destino de archivo complejo y vincular los puertos por el nombre o vincular los puertos en tiempo de ejecución en función de la directiva de vinculación. Developer tool puede crear un destino de archivo complejo Avro, Parquet, ORC o JSON.

1. Abra una asignación en el editor.
2. Haga clic con el botón derecho en la transformación en el editor de asignación y seleccione **Crear destino**.

Se abre la ventana **Crear destino**.

3. Elija el tipo de objeto de datos como Archivo complejo, Archivo sin formato, Relacional u Otro.

4. Si elige **Otro**, seleccione el tipo de objeto de destino requerido en la lista.
5. Elija un tipo de vínculo.

Se puede elegir entre los siguientes tipos de vínculos:

Vincular los puertos por nombre

Los puertos de la transformación de escritura corresponden a los del origen y tienen los mismos nombres.

Vincular el puerto dinámico según el flujo de asignaciones

La transformación de escritura contiene puertos dinámicos basados en objetos de nivel superior en el flujo de la asignación.

Vincular los puertos en tiempo de ejecución según la directiva de vinculación

Los puertos se crean en el destino en tiempo de ejecución según la directiva de vinculación que se configura en la ficha Vinculación en tiempo de ejecución de la transformación de escritura.

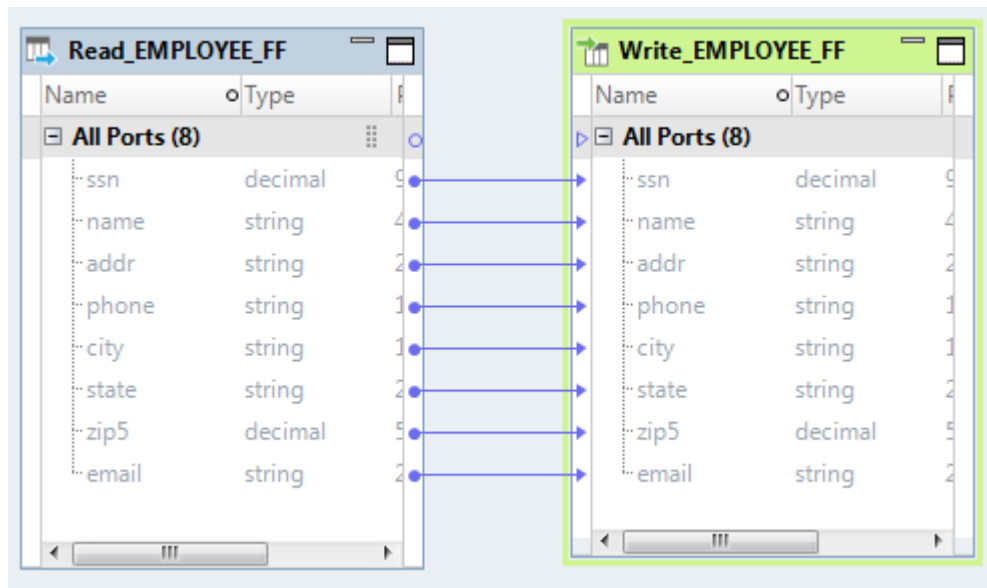
Para obtener más información acerca de puertos dinámicos y configuración de vinculación en tiempo de ejecución, consulte la *Guía de asignaciones del desarrollador de Informatica*.

6. Dé un nombre al nuevo objeto de datos.
7. Si lo desea, haga clic en **Examinar** para seleccionar una ubicación para el objeto de datos.
8. Si ha elegido crear un destino de archivo complejo, seleccione el formato de archivo complejo como Avro o Parquet en el cuadro desplegable **Formato de recurso**.
9. Haga clic en **Finalizar**.

Developer tool realiza las siguientes tareas:

- Agrega una transformación de escritura a la asignación.

La siguiente imagen muestra una transformación de escritura creada a partir de una transformación de lectura:



- Vincula puertos.
- Crea un objeto de datos físico.
Puede configurar las propiedades del objeto de datos físico. Por ejemplo, debe especificar una conexión de HDFS para el objeto de datos de archivo complejo.

APÉNDICE A

Delimitadores de transformación

- [Resumen de delimitadores de transformación, 758](#)

Resumen de delimitadores de transformación

Los delimitadores de transformación especifican las divisiones entre cadenas de datos.

La siguiente tabla enumera los delimitadores que utilizan las transformaciones para analizar y escribir cadenas de datos:

Nombre de delimitador	Símbolo de delimitador
símbolo "at"	@
coma	,
guión	-
comilla doble	"
barra diagonal	/
punto final	.
hash	#
barra vertical	
punto y coma	;
comilla simple	'
espacio	[Barra espaciadora]
tabulación	[tecla Tab]
subrayado	-

INDICE

A

- Acción de SOAP
 - cómo reemplazar una transformación de consumidor de servicio web [707](#)
- Actualizar o insertar (propiedad)
 - descripción [466](#)
- administración de ventanas
 - propiedades [295](#)
- almacén persistente de datos de índice de identidad [487](#)
- almacenamiento de cifrado
 - Transformación de enmascaramiento de datos [233](#)
- análisis de coincidencia de campos
 - análisis de campos definido [479](#)
 - flujo de proceso [504](#)
- análisis de coincidencia de identidades
 - almacén persistente de datos de índice [487](#)
 - análisis de identidades definido [479](#)
 - códigos de estado de persistencia [496](#), [497](#)
 - conjuntos de datos principales [487](#)
 - descripciones de los estados de persistencia [496](#), [497](#)
 - flujo de proceso [517](#)
 - propiedades de salida para el análisis de datos de índice [526](#)
 - reglas y directrices para datos de índice persistentes [487](#)
- análisis de coincidencia de origen dual [479](#)
- análisis de coincidencia de un solo origen [479](#)
- análisis de mensajes SOAP
 - descripción [720](#)
 - elemento de unión [728](#)
 - elementos QName [726](#)
 - salida desnormalizada [723](#)
 - salida normalizada [722](#)
 - salida pivotada [724](#)
 - tipos derivados [725](#)
- Analizador
 - descripción [240](#)
- anyType
 - asignar puertos [739](#)
- archivo de entrada de ejemplo
 - Transformación del Procesador de datos [242](#)
- archivo de origen de ejemplo
 - definir en la transformación del Procesador de datos [258](#)
- archivo de rechazo
 - estrategias de actualización [688](#)
- Archivo WSDL
 - elemento de enlace [696](#)
 - elemento de operación [696](#)
 - elemento de puerto [696](#)
 - elemento de servicio [696](#)
- archivos de memoria caché
 - directorio [74](#)
 - resumen [74](#)
- asignación de entrada
 - Transformación de consumidor de servicio web [701](#)
 - Transformación de consumidor de servicio web REST [610](#)

- asignación de salida
 - Transformación de consumidor de servicio web [704](#)
 - Transformación de consumidor de servicio web REST [612](#)
- asignaciones
 - cómo marcar filas para actualizar [687](#)
 - cómo utilizar transformaciones de enrutador [624](#)
- asignaciones con fallo
 - transformaciones de Java [349](#)
- asignaciones dinámicas
 - selector de puerto [292](#), [434](#)
 - condición de búsqueda [432](#)
 - reglas de selección [292](#), [434](#)
 - transformación de agregación [139](#)
 - transformación de enrutador [619](#)
 - transformación de estrategia de actualización [687](#)
 - transformación de filtro [307](#)
 - Transformación de macro [474](#)
 - transformación de ordenación [639](#)
 - transformación de rango [576](#)
 - transformaciones de búsqueda [426](#)
 - transformaciones de combinación [375](#)
- Asignador
 - descripción [240](#)
- asignar puertos
 - problemas de calidad [165](#)
- atributos anyAttribute
 - Transformación de consumidor de servicio web [701](#), [704](#)
- atributos de puerto
 - cómo propagar [65](#)
- autenticación por cookies
 - Transformación de consumidor de servicio web [700](#)
 - Transformación de consumidor de servicio web REST [609](#)

B

- Biblioteca
 - crear en la transformación del Procesador de datos [260](#)
- búsqueda
 - no almacenada en caché [453](#)
- búsquedas conectadas
 - descripción [416](#)
 - resumen [417](#)
- búsquedas no conectadas
 - descripción [416](#)
 - resumen [418](#)

C

- cadena
 - cómo crear rangos [576](#)
- cálculos
 - uso de variables con [48](#)
- campos de ocurrencia múltiple
 - transformación de normalizador [539](#)

- caracteres de cadena de origen
 - Transformación de enmascaramiento de datos [223](#)
- caracteres de reemplazo de cadena de resultado
 - Transformación de enmascaramiento de datos [224](#)
- ciclo
 - propiedad de transformación de generador de secuencia [629](#)
- Ciclo (propiedad)
 - propiedad de transformación de generador de secuencia [629](#)
- cifrado para preservar el formato [220](#)
- clase EDataType
 - expresiones de Java [364](#)
- clase JExpression
 - expresiones de Java [366](#), [367](#)
- clase JExprParaMetadata
 - expresiones de Java [365](#)
- classpath
 - propiedad de asignación [327](#)
- clave de cifrado de almacenamiento
 - Transformación de enmascaramiento de datos [233](#)
- claves
 - crear con la transformación de generador de secuencia [627](#)
 - jerarquía de mensajes SOAP [732](#)
- claves compuestas
 - crear con la transformación de generador de secuencia [627](#)
 - Transformación de consumidor de servicio web [701](#)
 - Transformación de consumidor de servicio web REST [610](#)
- claves de ordenación
 - configurar [640](#)
- claves externas
 - crear con la transformación de generador de secuencia [627](#)
- claves generadas
 - grupos de salida de servicio web [722](#)
- claves principales
 - crear con la transformación de generador de secuencia [627](#)
- clústeres
 - Transformación de excepción de registros duplicados [274](#)
- código de respuesta
 - Transformación de consumidor de servicio web REST [609](#)
- Código de respuesta HTTP
 - Transformación de consumidor de servicio web REST [609](#)
- código estático
 - transformaciones de Java [331](#)
- código Java
 - cómo encontrar errores [339](#)
 - en transformaciones de Java [328](#)
- columna de dependientes
 - enmascaramiento de datos [218](#)
- Columna Ubicación
 - transformación de servicio web [731](#)
- cómo compilar
 - transformaciones de Java [338](#)
- cómo compilar código Java
 - ficha Código completo [334](#)
- cómo crear
 - transformaciones de Java [337](#), [338](#)
- cómo crear rangos
 - grupos de datos [578](#)
 - valores de cadena [576](#)
- cómo diseñar
 - transformaciones de Java [323](#)
- cómo encontrar
 - errores en código Java [339](#)
- Cómo reenviar filas rechazadas
 - configuración [688](#)
 - opción [688](#)
- componente de inicio
 - Transformación del Procesador de datos [244](#)
- compresión SOAP
 - Transformación de consumidor de servicio web [710](#)
- condición de búsqueda
 - descripción [424](#)
 - especificar por parámetro [435](#)
 - Transformación de enmascaramiento de datos [217](#)
 - usar puertos generados [432](#)
- condición de combinación
 - resumen [378](#)
- condición del filtro de grupo
 - Transformación de enrutador [620](#)
- condiciones
 - Transformación de combinación [378](#)
 - Transformación de enrutador [620](#)
- conexión
 - servicios web [707](#)
 - servicios web REST [614](#)
- conexión HTTP
 - servicios web REST [614](#)
- conexiones de servicios web
 - resumen [707](#)
- configuración de codificación
 - Transformación del Procesador de datos [246](#)
- configuración de control de salida
 - Transformación del Procesador de datos [248](#)
- configuración del puerto de salida
 - descripción [300](#)
- conjuntos de resultados
 - ejemplo de procedimientos almacenados [671](#)
 - transformación de SQL [671](#)
 - ubicación de los parámetros de salida [673](#)
- consideraciones
 - transformaciones de Java [323](#)
- consolidación automática
 - Transformación de excepción de registros duplicados [275](#)
- consolidación manual
 - Transformación de excepción de registros duplicados [275](#)
- constantes
 - reemplazar valores nulos con [52](#)
- consulta de búsqueda
 - consulta predeterminada [419](#)
 - descripción [419](#)
 - directrices de reemplazo [421](#)
 - ORDER BY [419](#)
 - palabras reservadas [421](#)
 - reemplazar [419](#), [421](#)
- consulta SQL
 - transformación de SQL [658](#)
- consultas
 - reemplazar la búsqueda [419](#)
 - transformación de búsqueda [419](#)
- Continuar ante error de SQL
 - transformación de SQL [660](#), [663](#)
- crear
 - Objeto de datos de secuencia [632](#)
 - Transformación de generador de secuencia [635](#)
- criterios de selección
 - selector de puerto [291](#), [433](#)
- cuadro de diálogo Editar grupo de salida
 - transformación de normalizador [552](#)

D

- datos
 - almacenamiento temporal [47](#)
- Datos del clúster
 - grupo de salida [278](#)

- datos pivotados
 - mensajes SOAP [736](#)
- defineJExpression
 - método API de expresión de Java [365](#)
- dependencias
 - implícitas [66](#)
 - ruta de acceso al enlace [65](#)
- deseñofuque
 - valores de fecha [226](#)
 - valores numéricos [225](#)
- destinos
 - relacionales [686](#)
- diccionario
 - enmascaramiento de expresiones repetible [212](#)
 - máscara de datos de sustitución [215](#)
- dígito de inicio
 - Números de la seguridad social [230](#)
- directorio de archivo de memoria caché
 - Transformación de asociación [155](#)
 - transformación de consolidación [192](#)
 - Transformación de generador de claves [399](#)
- Directorio de archivo de memoria caché
 - Transformación de excepción de registros duplicados [279](#)
- directorio de la memoria caché
 - Transformación de enmascaramiento de datos [233](#)

E

- editor de expresiones
 - probar expresiones [289](#)
- Editor de expresiones
 - cómo validar expresiones [45](#)
 - Descripción [44](#)
- efectos secundarios
 - Transformación de consumidor de servicio web [712](#)
 - transformación de SQL [664](#)
- ejemplo
 - expresión dinámica [299](#)
 - claves de partición y ordenación [296](#)
- elemento de unión
 - cómo analizar mensajes SOAP [728](#)
- elementos
 - unión [742](#)
- elementos any
 - Transformación de consumidor de servicio web [701](#), [704](#)
- elementos anyType
 - cómo analizar [724](#)
 - Transformación de consumidor de servicio web [701](#), [704](#)
- elementos de lista
 - cómo analizar mensajes SOAP [727](#)
 - descripción [741](#)
- elementos de selección
 - cómo analizar mensajes SOAP [727](#)
 - cómo se visualizan en una transformación de consumidor de servicio web [705](#)
 - cómo se visualizan en una transformación de consumidor de servicio web REST [613](#)
 - descripción [740](#)
 - visualizar en una transformación de consumidor de servicio web [702](#)
- elementos de tipo derivado
 - Transformación de consumidor de servicio web [701](#), [704](#)
- elementos de unión
 - descripción [742](#)
- elementos QName
 - análisis de mensajes SOAP [726](#)

- encabezado HTTP
 - cómo añadir a una transformación de consumidor de servicio web [700](#)
 - cómo añadir a una transformación de consumidor de servicio web REST [608](#)
- enlace
 - Elemento de archivo WSDL [696](#)
- enlace de parámetro
 - transformación de SQL [658](#)
- enmascaramiento aleatorio
 - enmascaramiento de valores de cadena [210](#)
 - enmascaramiento de valores de fecha [211](#)
 - valores numéricos [209](#)
- enmascaramiento de cifrado
 - descripción [220](#)
- enmascaramiento de claves
 - descripción [213](#)
 - enmascaramiento de valores de cadena [213](#)
 - enmascaramiento de valores de fecha y hora [214](#)
 - enmascaramiento de valores numéricos [214](#)
 - valores numéricos [213](#)
- enmascaramiento de expresiones
 - descripción [211](#)
 - ejemplo de enmascaramiento repetible [212](#)
 - enmascaramiento repetible [211](#)
 - reglas y directrices [213](#)
- enmascaramiento de formato especial
 - direcciones de correo electrónico [227](#)
 - direcciones IP [228](#)
 - Direcciones URL [230](#)
 - Números de la seguridad social [229](#), [230](#)
 - números de SIN repetibles [230](#)
 - números de tarjeta de crédito [226](#)
 - números de teléfono [229](#)
- enmascaramiento dependiente
 - descripción [218](#)
- entrada desnormalizada
 - puertos de servicio web [737](#)
- errores
 - cómo incrementar el umbral en transformaciones de Java [351](#)
 - control [54](#)
- errores ajenos al código de usuario
 - en transformaciones de Java [340](#)
- errores de código de usuario
 - transformaciones de Java [340](#)
- errores de compilación
 - cómo identificar el origen de errores en transformaciones de Java [340](#)
- esquemas
 - transformación de jerárquica a relacional [316](#)
 - Transformación de relacional a jerárquica [598](#)
 - Transformación del Procesador de datos [245](#)
- estrategia de coincidencia
 - Transformación de excepción de registros duplicados [281](#)
- evento de advertencia
 - Transformación del Procesador de datos [253](#)
- evento de error
 - Transformación del Procesador de datos [253](#)
- evento de error fatal
 - Transformación del Procesador de datos [253](#)
- evento de error opcional
 - Transformación del Procesador de datos [253](#)
- evento de notificación
 - Transformación del Procesador de datos [253](#)
- eventos
 - Transformación del Procesador de datos [253](#)
- Excel
 - configurar puertos de transformación [69](#)

- Excel (*continuado*)
 - copiar a Developer tool [69](#)
 - copiar puertos [68](#)
 - editar una transformación [70](#)
 - Copiar reglas y directrices [71](#)
- expresión dinámica
 - crear [301](#)
- expresiones
 - cómo añadir a un puerto [45](#)
 - cómo añadir comentarios [45](#)
 - cómo validar [45](#)
 - en transformaciones [43](#)
 - simplificación [47](#)
 - transformaciones de Java [358](#)
- Expresiones
 - cómo introducir [44](#)
- expresiones de Java
 - clase EDataType [364](#)
 - clase JExpression [366](#), [367](#)
 - clase JExprParaMetadata [365](#)
 - cómo configurar [359](#)
 - cómo configurar funciones [360](#)
 - cómo crear [360](#)
 - cómo crear en el cuadro de diálogo Definir función [360](#)
 - cómo generar [359](#)
 - cómo invocar [352](#)
 - cómo invocar con la interfaz avanzada [363](#)
 - cómo invocar con la interfaz simple [362](#)
 - ejemplo de interfaz avanzada [366](#)
 - ejemplo de interfaz simple [362](#)
 - generar código Java [360](#)
 - interfaz avanzada [363](#)
 - interfaz simple [361](#)
 - método API invokeJExpression [362](#)
 - reglas y directrices [362](#), [364](#)
 - reglas y directrices para invocar [352](#)
 - transformaciones de Java [358](#)
- expresiones dinámicas
 - ejemplo [299](#)
 - configuración del puerto de salida [300](#)
 - resumen [299](#)

F

- fallos SOAP genéricos
 - Transformación de consumidor de servicio web [710](#)
- ficha Agrupar por
 - descripción [143](#)
 - parámetros de lista de puertos [144](#), [579](#)
- ficha Al final
 - transformaciones de Java [333](#)
- ficha Aplicaciones auxiliares
 - transformaciones de Java [331](#), [332](#)
- ficha Código completo
 - errores de compilación de Java [339](#)
 - transformaciones de Java [334](#)
- ficha En entrada
 - transformaciones de Java [333](#)
- ficha Funciones
 - transformaciones de Java [334](#)
- ficha Importaciones
 - transformaciones de Java [331](#), [332](#)
- ficha Objeto de datos
 - descripción [588](#)
 - descripciones de campo [427](#)
 - parametrizar mediante un nuevo objeto de datos [430](#)

- ficha Ordenar
 - transformación de ordenación [640](#)
- filas
 - cómo marcar para actualizar [687](#)
- filas de entrada
 - cómo obtener el tipo de fila para [350](#)
- filtrar filas de origen
 - transformación de búsqueda [422](#)
- filtro de origen de búsqueda
 - limitación de búsquedas [422](#)
- formato de máscara
 - enmascaramiento de valores de cadena [222](#)
 - formatos de máscara especiales [226](#)
- función ABORT
 - uso [52](#)
- función ERROR
 - uso [52](#)
- función IIF
 - reemplazar claves que faltan con la transformación de generador de secuencia [628](#)

G

- GCID
 - descripción [539](#)
 - ejemplo de transformación de normalizador [551](#)
- generar código Java
 - expresiones de Java [360](#)
- generar fila de salida
 - transformaciones de Java [350](#)
- generar tabla de registros incorrectos
 - Transformación de excepción de registros incorrectos [164](#)
- Grupo de salida estándar
 - Transformación de excepción de registros duplicados [278](#)
- grupo de secuencia
 - cómo se visualizan en una transformación de consumidor de servicio web REST [613](#)
- grupo definido por el usuario
 - Transformación de enrutador [620](#)
- grupo todo
 - cómo se visualizan en una transformación de consumidor de servicio web REST [613](#)
 - visualizar en una transformación de consumidor de servicio web [702](#)
- grupos
 - cómo añadir a la transformación de enrutador [623](#)
 - definido por el usuario [620](#)
 - Transformación de enrutador [620](#)
- grupos de salida
 - Transformación de excepción de registros incorrectos [157](#)
 - transformación de normalizador [551](#)
- grupos de sustitución
 - cómo analizar mensajes SOAP [727](#)
 - servicios web [740](#)
 - Transformación de consumidor de servicio web [701](#), [704](#)
- grupos en análisis de coincidencia [480](#)
- grupos predeterminados
 - Transformación de enrutador [620](#)
- GZip
 - cómo comprimir mensajes SOAP [710](#)

I

- incrementar
 - configurar el intervalo de secuencia [630](#)

- Incrementar por (propiedad)
 - propiedad de transformación de generador de secuencia [629](#)
- información del diccionario
 - Transformación de enmascaramiento de datos [217](#)
- inicialización
 - variables [50](#)
- Insertar o actualizar (propiedad)
 - descripción [465](#)
- instrucción macro
 - descripción [473](#)
- interfaz avanzada
 - clase EDataType [364](#)
 - clase JExpression [366](#), [367](#)
 - clase JExprParaMetadata [365](#)
 - cómo invocar expresiones de Java [363](#)
 - ejemplo [366](#)
 - expresiones de Java [363](#)
- interfaz simple
 - ejemplo [362](#)
 - expresiones de Java [361](#)
 - Métodos API de transformación de Java [361](#)
- intervalo
 - enmascaramiento de valores numéricos [225](#)
- intervalo de confirmación de almacenamiento
 - Transformación de enmascaramiento de datos [233](#)
- invoke
 - método API de expresión de Java [369](#)
- invokeJExpression
 - método API [362](#)

J

- JDK
 - transformación de Java [319](#)
- jerarquía de entrada
 - Transformación de normalizador [541](#)
- jerarquía SOAP
 - relación con puertos de entrada [731](#)
- JRE
 - transformación de Java [319](#)

L

- longitud variable
 - en la transformación de ordenación [644](#)

M

- mantener orden de filas
 - Transformación de generador de secuencia [631](#)
- Mapplet
 - referencia [245](#)
- máscara de sustitución
 - descripción [215](#)
 - propiedades de enmascaramiento [216](#)
- memoria caché
 - transformación de búsqueda [425](#)
- memoria caché de búsqueda
 - resumen [425](#)
- memoria caché de búsqueda compartida
 - directrices [456](#)
 - partición, directrices [456](#)
- memoria caché de búsqueda dinámica
 - casos de uso [459](#)
 - descripción [458](#)

- memoria caché de búsqueda dinámica (*continuado*)
 - reemplazo de SQL de búsqueda [464](#)
 - uso de orígenes de archivo sin formato [458](#)
- memoria caché de búsqueda estática
 - resumen [453](#)
- memoria caché de búsqueda persistente
 - resumen [454](#)
- memorias caché
 - archivos de memoria caché [72](#)
 - datos [73](#)
 - directorio para archivos [74](#)
 - índice [73](#)
 - memoria caché de búsqueda dinámica [458](#)
 - memoria caché de búsqueda estática [453](#)
 - optimización de tamaño [78](#)
 - particiones [77](#)
 - resumen [72](#)
 - tamaño [75](#)
 - tamaño aumentado por el Servicio de integración de datos [77](#)
 - tipos [73](#)
 - transformación de búsqueda [451](#)
 - transformaciones [72](#)
- memorias caché de búsqueda
 - definición [451](#)
 - dinámica [458](#)
 - estática [453](#)
 - persistente [454](#)
 - resumen [451](#)
- memorias caché de datos
 - transformaciones [73](#)
- memorias caché de índice
 - transformaciones [73](#)
- mensaje SOAP
 - claves [732](#)
- mensajes de solicitud de servicio web simultáneos
 - cómo habilitar una transformación de consumidor de servicio web [707](#)
- mensajes SOAP
 - asignar elementos de lista [741](#)
 - asignar elementos de selección [740](#)
 - asignar nodos de ocurrencia múltiple [722](#)
 - asignar puertos [734](#)
 - asignar puertos a elementos de unión [742](#)
 - asignar varios puertos de entrada [736](#)
 - cómo analizar elementos anyType [724](#)
 - cómo analizar elementos de lista [727](#)
 - cómo analizar elementos de selección [727](#)
 - cómo analizar grupos de sustitución [727](#)
 - cómo pivotar datos [736](#)
 - resumen [696](#)
- Método API de expresión de Java
 - getResultDataType [368](#)
- método defineJExpression
 - transformaciones de Java [348](#)
- método failSession
 - transformaciones de Java [349](#)
- método generateRow
 - transformaciones de Java [350](#)
- método getBytes
 - transformaciones de Java [368](#)
- método getDouble
 - transformaciones de Java [368](#)
- método getInRowType
 - transformaciones de Java [350](#)
- método getInt
 - transformaciones de Java [368](#)
- método getLong
 - transformaciones de Java [368](#)

- Método getMetada
 - transformaciones de Java [351](#)
- método getResultDataType
 - Transformaciones de Java [368](#)
- método getResultMetadata
 - transformaciones de Java [369](#)
- método getStringBuffer
 - transformaciones de Java [369](#)
- método incrementErrorCount
 - transformaciones de Java [351](#)
- método invokeJExpression
 - transformaciones de Java [352](#)
- método isNull
 - transformaciones de Java [353](#)
- método isResultNull
 - Transformaciones de Java [370](#)
- método logError
 - transformaciones de Java [354](#)
- método logInfo
 - transformaciones de Java [354](#)
- método resetNotification
 - transformaciones de Java [355](#)
- método setNull
 - transformaciones de Java [355](#)
- Método storeMetada
 - transformaciones de Java [356](#)
- métodos
 - API de transformación de Java [347](#)
- métodos API
 - transformación de Java [347](#)
 - transformaciones de Java [347](#)
- métodos API de expresión de Java
 - defineJExpression [365](#)
 - getDouble [368](#)
 - getInt [368](#)
 - getLong [368](#)
 - getResultMetadata [369](#)
 - getStringBuffer [369](#)
 - invoke [369](#)
 - isResultNull [370](#)
- Métodos API de expresión de Java
 - getBytes [368](#)
- métodos definidos por el usuario
 - transformaciones de Java [331](#)

N

- niveles de seguimiento
 - Propiedad de transformación del generador de secuencia [631](#)
 - Transformación de asociación [155](#)
 - transformación de consolidación [192](#)
 - Transformación de excepción de registros duplicados [279](#)
 - Transformación de generador de claves [399](#)
- nombre de usuario de seguridad WS
 - puerto dinámico [700](#)
- Número de filas afectadas
 - transformación de SQL [655](#)
- número de valores almacenados en la memoria caché
 - valor de propiedad Generador de secuencia [629](#)
- Números de la seguridad social
 - enmascaramiento de código de área [229](#)
 - enmascaramiento de datos repetibles [229](#)
- Números de SIN
 - enmascaramiento de datos repetibles [230](#)
 - enmascaramiento de números de la seguridad social [230](#)
- NumRowsAffected
 - salida de filas [661](#)

O

- objeto de datos
 - parametrizado duplicado [430](#)
 - parametrizar [588](#)
- objeto de datos de lectura
 - parametrizar [589](#)
- Objeto de datos de secuencia
 - crear [632](#)
 - propiedades [632](#)
- Objeto de XMap
 - crear en la transformación del Procesador de datos [259](#)
- objeto Regla de validación
 - crear en la transformación del Procesador de datos [260](#)
- objetos de datos físicos
 - sincronización [587](#)
- objetos de datos relacionales
 - crear una transformación de lectura dinámica [592](#)
- objetos de datos relacionales dinámicos
 - crear transformaciones de lectura [592](#)
- opción de parámetro
 - transformación de combinación [381](#)
- operación
 - Elemento de archivo WSDL [696](#)
- optimización de inserción
 - habilitar en la transformación de SQL [665](#)
 - Transformación de consumidor de servicio web [712](#)
 - transformación de SQL [664](#)
- optimización de origen
 - restricciones [591](#)
- optimización de primera selección
 - Transformación de consumidor de servicio web [712](#)
 - transformación de SQL [664](#)
- ordenación
 - longitud variable [644](#)
- ORDER BY
 - consulta de búsqueda [419](#)
 - reemplazar [419](#)
- origen de búsqueda
 - parametrizar [430](#)
- Origen de ejemplo
 - crear en la transformación del Procesador de datos [260](#)

P

- palabras reservadas
 - consulta de búsqueda [421](#)
- paquetes Java
 - cómo importar [331](#)
- parámetro de expresión
 - condición de búsqueda [435](#)
 - de una condición de combinación [381](#)
- parámetros
 - condición de filtro [308](#)
 - transformación de enrutador [623](#)
 - Transformación de macro [474](#)
- parámetros de asignación
 - en el reemplazo de SQL de búsqueda [419](#)
- parámetros de lista de puertos
 - ficha Agrupar por [144](#), [579](#)
- pautas de codificación
 - Transformación del Procesador de datos [248](#)
- problemas de calidad
 - asignar puertos a [165](#)
- procedimientos almacenados
 - ejemplo [673](#)
 - ejemplos de conjuntos de resultados [671](#)

- procedimientos almacenados (*continuado*)
 - escribir en variables [49](#)
 - parámetros [670](#)
 - puerto de valor de devolución [669](#)
 - puertos de entrada y salida [669](#)
 - transformación de SQL [669](#)
 - valor de devolución [670](#)
- Procesador de datos, vista de eventos
 - visualización del registro de eventos [256](#)
- procesamiento al nanosegundo
 - transformaciones de Java [325](#)
- procesamiento de alta precisión
 - transformaciones de Java [325](#)
- propiedad de ordenación
 - transformación de consolidación [192](#)
 - Transformación de excepción de registros duplicados [279](#)
- propiedades
 - transformación de escritura
 - propiedades [747](#)
 - transformación de lectura
 - propiedades [584](#)
- propiedades avanzadas
 - Transformación de asociación [155](#)
 - transformación de consolidación [192](#)
 - Transformación de consumidor de servicio web [707](#)
 - Transformación de consumidor de servicio web REST [614](#)
 - Transformación de excepción de registros duplicados [279](#)
 - Transformación de generador de claves [399](#)
 - transformaciones de Java [325](#)
- Propiedades avanzadas
 - Transformación de excepción de registros incorrectos [165](#)
 - transformación de SQL [655](#)
- propiedades de administración de ventana
 - marco [295](#)
 - ordenación [295](#), [296](#)
 - partición [295](#), [296](#)
 - reglas y directrices [298](#)
- puerto de entrada de archivo
 - Transformación del Procesador de datos [242](#)
- puerto de entrada del búfer
 - Transformación del Procesador de datos [242](#)
- puerto de filtro
 - Transformación de consumidor de servicio web [712](#)
- puerto de parámetros de servicio
 - Transformación del Procesador de datos [242](#)
- puerto de rango
 - Transformación de rango [578](#)
- puerto de salida de archivo
 - Transformación del Procesador de datos [243](#)
- puerto de salida del búfer
 - Transformación del Procesador de datos [243](#)
- Puerto de salida estándar
 - descripción [162](#)
- puerto NEXTVAL
 - Generador de secuencia [627](#)
- puerto SQLException
 - transformación de SQL [654](#)
- puertos
 - asignar a mensajes SOAP [734](#)
 - atributos propagados por transformación [66](#)
 - cómo crear [61](#)
 - cómo vincular [62](#)
 - cómo vincular automáticamente [63](#)
 - cómo vincular manualmente [63](#)
 - cómo vincular por nombre [63](#)
 - cómo vincular por posición [64](#)
 - configurar [62](#)
 - copiar desde Excel [68](#)

- puertos (*continuado*)
 - entrada de servicio web desnormalizado [737](#)
 - orden de evaluación [49](#)
 - puertos de variables [47](#)
 - reglas y directrices de vinculación [64](#)
 - resumen de valores predeterminados [50](#)
 - Transformación de enrutador [624](#)
 - Transformación de excepción de registros duplicados [277](#)
 - Transformación de excepción de registros incorrectos [161](#)
 - transformación de generador de secuencia [626](#)
 - transformaciones de Java [324](#)
- puertos de entrada
 - transformaciones de Java [324](#)
 - valores predeterminados [50](#)
- puertos de entrada de SQL
 - transformación de SQL [652](#)
- puertos de parámetros de servicio
 - Transformación del Procesador de datos [243](#)
- puertos de problema de calidad
 - En blanco o nulo [157](#)
- puertos de problemas de calidad
 - descripción [161](#)
- puertos de salida
 - control de errores [50](#)
 - transformación de Java [324](#)
 - transformaciones de Java [324](#)
 - valores predeterminados [50](#)
- puertos de transferencia
 - transformación de expresión [288](#)
 - transformación de SQL [653](#)
 - valores predeterminados [50](#)
- puertos de transformación
 - resumen [61](#)
- puertos de variable
 - resumen [47](#)
 - transformación de expresión [288](#)
- puertos dinámicos
 - condición de búsqueda [427](#)
 - condición de unión [381](#)
- puntuaciones de coincidencia nulas
 - Transformación de coincidencia [482](#)
- puntuaciones de enlace
 - Transformación de coincidencia [484](#)
- puntuaciones del controlador
 - Transformación de coincidencia [484](#)

R

- Recuento máximo de filas de salida
 - transformación de SQL [660](#), [661](#)
- recuentos de errores
 - cómo incrementar para transformaciones Java [351](#)
- reemplazar acción de SOAP
 - Transformación de consumidor de servicio web [707](#)
- reemplazo de búsqueda
 - directrices [421](#)
- reemplazo de SQL de búsqueda
 - memoria caché dinámica [464](#)
- registro
 - definición [254](#)
- registro de eventos
 - visualización [256](#)
- registro de eventos de tiempo de ejecución
 - Transformación del Procesador de datos [255](#)
- registro de eventos en tiempo de diseño
 - descripción y ubicación [255](#)

- registro de eventos, tiempo de diseño
 - descripción y ubicación [255](#)
- registro de usuario
 - Transformación del Procesador de datos [256](#)
- registro, eventos de tiempo de diseño
 - descripción y ubicación [255](#)
- registros
 - transformación de normalizador [540](#)
 - transformaciones de Java [354](#)
- registros rechazados
 - Transformación de excepción de registros incorrectos [162](#)
- reglas
 - valores predeterminados [56](#)
- reglas de enmascaramiento
 - caracteres de cadena de origen [223](#)
 - caracteres de reemplazo de cadena de resultado [224](#)
 - desenfoque [225](#)
 - formato de máscara [222](#)
 - formatos de máscara especiales [226](#)
 - intervalo [225](#)
- reglas de selección
 - asignaciones dinámicas [292](#), [434](#)
 - selectores de puerto [291](#), [433](#)
 - transformación de combinación [376](#)
- reglas y directrices
 - propiedades de administración de ventana [298](#)
- rendimiento
 - utilizar variables para mejorar [47](#)
- restablecer
 - transformación de generador de secuencia [631](#)
- Restablecer (propiedad)
 - transformación de generador de secuencia [631](#)
- restricciones
 - optimización de origen [591](#)
- Reutilizable
 - transformación de generador de secuencia [632](#)

S

- salida de error HTTP
 - cómo habilitar una transformación de consumidor de servicio web [707](#)
- salida de fallos genéricos
 - cómo habilitar una transformación de consumidor de servicio web [707](#)
- salida desnormalizada
 - análisis de mensajes SOAP [723](#)
- salida distinta
 - transformación de ordenación [640](#)
- salida pivotada
 - análisis de mensajes SOAP [724](#)
- salida única
 - Transformación de enmascaramiento de datos [217](#)
- Script
 - crear en la transformación del Procesador de datos [258](#)
- seguridad de la capa de transporte
 - Transformación de consumidor de servicio web [697](#)
 - transformación de consumidor de servicio web REST [697](#)
- selector de puerto
 - reglas de selección [291](#), [292](#), [433](#), [434](#)
 - crear [293](#), [377](#), [436](#)
 - en expresiones dinámicas [299](#)
 - reglas de selección [291](#), [292](#), [433](#), [434](#)
 - transformación de combinación [375](#), [376](#)
- selectores de puerto
 - en transformaciones de combinación [380](#)

- servicio
 - Elemento de archivo WSDL [696](#)
- Servicio de Data Transformation
 - importar al repositorio de modelos [262](#), [263](#)
 - importar varios servicios [262](#)
- Servicio de integración de datos)
 - modo de reinicio [355](#)
- servicio web
 - asignar puertos a anyTypes [739](#)
 - grupos de sustitución [740](#)
 - tipos derivados [739](#)
- Sin estado
 - transformaciones de Java [325](#)
- sincronización
 - objetos de datos físicos [587](#)
 - objetos de datos personalizados [587](#)
- sincronizar con el editor
 - Transformación del Procesador de datos [261](#)
- snippets de código
 - cómo crear para transformaciones de Java [329](#)
- snippets de código Java
 - cómo crear para transformaciones de Java [329](#)
- solución de problemas
 - transformaciones de Java [339](#)
- subregistros
 - transformación de normalizador [540](#)
- sustitución de cadenas
 - transformación de SQL [659](#)

T

- tabla de almacenamiento
 - enmascaramiento de expresiones [212](#)
 - máscara de datos de sustitución [216](#)
- tabla de almacenamiento compartido
 - Transformación de enmascaramiento de datos [233](#)
- Tabla de problemas
 - cómo generar [164](#)
- tabla de registros únicos
 - crear [286](#)
- tamaño de archivo de memoria caché
 - Transformación de asociación [155](#)
 - transformación de consolidación [192](#)
 - Transformación de excepción de registros duplicados [279](#)
 - Transformación de generador de claves [399](#)
- tamaño de la memoria caché
 - automático [75](#)
 - Transformación de enmascaramiento de datos [233](#)
- tamaño de la memoria caché automático
 - descripción [75](#)
- Tareas humanas
 - excepciones de registros incorrectos [160](#)
- técnicas de enmascaramiento
 - cifrado [220](#)
 - enmascaramiento de datos [209](#)
- tipo de condición
 - propiedades avanzadas de la transformación de combinación [380](#)
 - transformación de combinación simple [379](#)
- tipo de condición avanzada
 - Transformación de combinación [380](#)
- tipo de condición simple
 - Transformación de combinación [379](#)
- tipos de datos
 - transformaciones de Java [320](#)
- tipos de datos primitivos Java
 - transformaciones de Java [320](#)

- tipos de evento
 - Transformación del Procesador de datos [253](#)
- tipos derivados
 - cómo analizar mensajes SOAP [725](#)
 - servicios web [739](#)
- todos los grupos
 - cómo se visualizan en una transformación de consumidor de servicio web [705](#)
- transformación de agregación
 - agrupar por puertos [142](#)
 - asignaciones dinámicas [139](#)
 - cómo ordenar datos [146](#)
 - crear una transformación reutilizable [148](#)
 - entorno no nativo [150](#)
 - entrada ordenada [145](#)
 - expresiones de agregado [140](#)
 - expresiones de no agregado [145](#)
 - funciones de agregado [141](#)
 - funciones de agregado anidadas [142](#)
 - memorias caché [145](#)
 - motor de Blaze [150](#)
 - motor de Databricks Spark [151](#)
 - motor de Spark [151](#)
 - puertos [139](#)
 - resumen [138](#)
 - uso de variables [47](#)
- Transformación de agregación
 - Combinación de estrategia de actualización [689](#)
 - cómo crear una transformación no reutilizable [148](#)
 - consejos [149](#)
 - desarrollo [139](#)
 - propiedades avanzadas [147](#)
 - solución de problemas [149](#)
 - tamaño de la memoria caché [72](#)
- transformación de analizador
 - entorno no nativo [573](#)
- Transformación de analizador
 - resumen [562](#)
- Transformación de asociación
 - niveles de seguimiento [155](#)
 - propiedades avanzadas [155](#)
 - resumen [153](#)
- transformación de búsqueda
 - asignaciones dinámicas [426](#)
 - búsqueda no almacenada en caché [453](#)
 - cómo crear búsquedas no conectadas [443](#)
 - condición de búsqueda [422](#)
 - conectada [416](#), [417](#)
 - conflictos de nombre de puerto [429](#)
 - consulta predeterminada [419](#)
 - crear una transformación no reutilizable [442](#)
 - crear una transformación reutilizable [441](#)
 - desarrollo [419](#)
 - entorno no nativo [446](#)
 - memoria caché dinámica [458](#)
 - memoria caché persistente [454](#)
 - memorias caché [451](#)
 - motor de Blaze [446](#)
 - motor de Databricks Spark [449](#)
 - motor de Spark [447](#)
 - no conectadas [416](#), [418](#)
 - objeto de datos parametrizados relacionales [430](#)
 - parámetros y variables de asignación [419](#)
 - propiedades avanzadas [439](#)
 - propiedades de consulta [426](#)
 - propiedades de tiempo de ejecución [438](#)
 - puertos dinámicos [427](#)
 - reemplazar la búsqueda [421](#)
- transformación de búsqueda (*continuado*)
 - reemplazo de consulta de búsqueda [419](#)
 - reglas y directrices para la condición de búsqueda [424](#)
 - resumen [415](#)
 - tamaño de la memoria caché, reducir [419](#)
- Transformación de búsqueda
 - almacenar en memoria caché [425](#)
 - tamaño de la memoria caché [72](#)
- transformación de clasificador
 - entorno no nativo [183](#)
- Transformación de clasificador
 - algoritmo clasificador [176](#)
 - modelo de clasificador [176](#)
 - resumen [175](#)
- transformación de coincidencia
 - entorno no nativo [502](#)
 - motor de Blaze [502](#)
 - motor de Spark [502](#)
- Transformación de coincidencia
 - análisis de campos definido [479](#)
 - análisis de identidades definido [479](#)
 - análisis de origen dual [479](#)
 - análisis de un solo origen [479](#)
 - casos de uso [477](#)
 - códigos de estado de persistencia [496](#), [497](#)
 - conceptos del análisis de coincidencia [478](#)
 - configuración de una operación de análisis de coincidencia [501](#)
 - creación de un mapplet para el análisis de coincidencia [500](#)
 - datos agrupados [480](#)
 - descripciones de los estados de persistencia [496](#), [497](#)
 - ejemplo de estrategia de coincidencia [281](#)
 - factores de rendimiento [488](#)
 - flujo de proceso del análisis de campos [504](#)
 - flujo del proceso de análisis de identidad [517](#)
 - formatos de salida [524](#)
 - opciones de puntuación del clúster [483](#)
 - Opciones de vista de salida de coincidencia [508](#)
 - Propiedades de coincidencia en la vista de salida de coincidencia [526](#)
 - Propiedades de la vista Salida de coincidencia [509](#)
 - Propiedades de salida en la vista de salida de coincidencia [526](#)
 - puertos de entrada predefinidos [495](#)
 - puertos de salida predefinidos [495](#)
 - puntuaciones de coincidencia nulas [482](#)
 - puntuaciones de enlace [484](#)
 - puntuaciones del controlador [484](#)
 - tablas de índices de identidad [487](#)
 - visualización de datos de rendimiento [489](#)
 - visualización de datos del análisis del clúster [489](#)
- transformación de combinación
 - asignaciones dinámicas [375](#)
 - entorno no nativo [393](#)
 - motor de Blaze [393](#)
 - motor de Databricks Spark [394](#)
 - motor de Spark [394](#)
 - parámetro de expresión [381](#)
 - puertos [374](#)
 - puertos dinámicos [381](#)
 - reglas de selección [376](#)
 - rendimiento [392](#)
 - selectores de puerto [375](#)
 - usar selectores de puerto [380](#)
- Transformación de combinación
 - condiciones [378](#)
 - memorias caché [373](#)
 - propiedades avanzadas [372](#)
 - reglas y directrices [393](#)
 - resumen [371](#)

Transformación de combinación (*continuado*)
 tamaño de la memoria caché [72](#)
 tipo de condición [379](#), [380](#)
 tipo de condición avanzada [380](#)
 tipo de condición simple [379](#)

transformación de comparación
 entorno no nativo [189](#)

Transformación de comparación
 resumen [184](#)

transformación de consolidación
 entorno no nativo [206](#)
 niveles de seguimiento [192](#)
 ordenar registros [192](#)
 propiedades avanzadas [192](#)

Transformación de consolidación
 resumen [190](#)

Transformación de consumidor de servicio web
 asignación de entrada [701](#)
 asignación de salida [704](#)
 asignar nodos de salida [704](#)
 asignar puertos de entrada [701](#)
 autenticación por cookies [700](#)
 cómo añadir encabezados HTTP [700](#)
 cómo habilitar la salida de error HTTP [707](#)
 cómo habilitar la salida de fallos genéricos [707](#)
 cómo visualizar claves [705](#)
 compresión SOAP [710](#)
 Control de errores [710](#)
 crear [714](#)
 fallos SOAP genéricos [710](#)
 habilitar la optimización de inserción [713](#)
 mensajes de solicitud de servicio web simultáneos [707](#)
 mensajes SOAP [696](#)
 nombre de seguridad WS dinámica [700](#)
 operaciones [697](#)
 optimización de inserción [712](#)
 optimización de primera selección [712](#)
 optimización del filtro [712](#)
 propiedades avanzadas [707](#)
 resumen [695](#)
 seguridad [697](#)
 seguridad de la capa de transporte [697](#)
 URL de punto final [700](#)
 URL dinámica de servicio web [700](#)
 visualizar claves [702](#)

transformación de consumidor de servicio web REST
 crear [615](#)
 seguridad [697](#)
 seguridad de la capa de transporte [697](#)

Transformación de consumidor de servicio web REST
 asignación de elementos a puertos [607](#)
 asignación de entrada [610](#)
 asignación de puertos de salida [613](#)
 asignación de salida [612](#)
 cómo asignar puertos de entrada [611](#)
 compatibilidad del servidor proxy [600](#)
 configuración [602](#)
 configuración de mensaje [602](#)
 configuración de URL de base [614](#)
 entrada de asignación [600](#)
 entrada ordenada [614](#)
 identificación de recurso [603](#)
 Método Delete [606](#)
 Método Get [604](#)
 Método Post [605](#)
 Método Put [605](#)
 Métodos HTTP [604](#)
 nivel de seguimiento [614](#)

Transformación de consumidor de servicio web REST (*continuado*)
 no reutilizables [615](#)
 personalización de vista de asignación de salida [613](#)
 proceso [602](#)
 propiedad de conexión [614](#)
 propiedades avanzadas [614](#)
 Puerto de RequestInput [607](#)
 puertos [607](#)
 puertos de argumento [608](#)
 puertos de código de respuesta [609](#)
 puertos de cookies [609](#)
 Puertos de encabezado HTTP [608](#)
 puertos de entrada [607](#)
 puertos de salida [607](#)
 puertos de salida XML [609](#)
 puertos de transferencia [607](#)
 Puertos URL [608](#)
 reglas de la asignación de entrada [610](#)
 reglas de la asignación de salida [613](#)
 resumen [600](#)
 reutilizables [615](#)
 salida de asignación [600](#)
 tipo de medio de Internet [614](#)
 validación de esquema XML [614](#)

transformación de conversión de mayúsculas y minúsculas
 entorno no nativo [174](#)

Transformación de conversión de mayúsculas y minúsculas
 resumen [172](#)

transformación de decisión
 entorno no nativo [272](#)

Transformación de decisión
 estrategia de ejemplo de problemas de calidad [160](#)
 propiedades avanzadas [272](#)
 resumen [266](#)

transformación de enmascaramiento de datos
 enmascaramiento de direcciones IP [228](#)
 entorno no nativo [237](#)

Transformación de enmascaramiento de datos
 archivo de valores predeterminados [230](#)
 caracteres de cadena de origen [223](#)
 descripción [208](#)
 desenfoque [225](#)
 diccionario para la máscara de sustitución [215](#)
 directorio de la memoria caché [233](#)
 directrices de enmascaramiento de expresiones [213](#)
 enmascaramiento aleatorio [209](#)
 enmascaramiento de datos dependiente [218](#)
 enmascaramiento de direcciones de correo electrónico [227](#)
 enmascaramiento de direcciones URL [230](#)
 enmascaramiento de expresiones [211](#)
 enmascaramiento de expresiones del nombre del diccionario [212](#)
 enmascaramiento de expresiones repetible [211](#)
 enmascaramiento de números de la seguridad social [229](#), [230](#), [232](#)
 enmascaramiento de números de teléfono [229](#)
 enmascaramiento de tarjetas de crédito [226](#)
 enmascaramiento de valores de fecha [225](#)
 formato de máscara [222](#)
 formatos de máscara especiales [226](#)
 intervalo [225](#)
 intervalo de confirmación de almacenamiento [233](#)
 máscara de sustitución [215](#)
 números de SIN repetibles [230](#)
 propiedades de la máscara de sustitución [216](#), [217](#)
 propiedades de tiempo de ejecución [233](#)
 salida única [233](#)
 SSN repetibles [229](#)
 tabla de almacenamiento [212](#)
 tabla de almacenamiento compartido [233](#)

Transformación de enmascaramiento de datos (*continuado*)
 tablas de almacenamiento [216](#)
 tamaño de la memoria caché [233](#)
 técnicas de enmascaramiento [209](#)

transformación de enrutador
 asignaciones dinámicas [619](#)
 entorno no nativo [625](#)
 parámetros [623](#)
 puertos dinámicos [623](#)

Transformación de enrutador
 cómo conectar en asignaciones [624](#)
 condición del filtro de grupo [620](#)
 ejemplo [620](#)
 grupos [620](#)
 propiedades avanzadas [625](#)
 puertos [624](#)
 resumen [618](#)

transformación de escritura
 resumen [747](#)

Transformación de estandarizador
 entorno no nativo [681](#)
 resumen [678](#)

transformación de estrategia de actualización
 asignaciones dinámicas [687](#)
 entorno no nativo [690](#)
 motor de Blaze [690](#)
 motor de Spark [691](#)

Transformación de estrategia de actualización
 Combinación de agregación [689](#)
 cómo crear [687](#)
 cómo reenviar filas rechazadas [688](#)
 expresiones [688](#)
 pasos para la configuración [687](#)
 propiedades avanzadas [688](#)
 resumen [686](#)

transformación de etiquetador
 entorno no nativo [414](#)

Transformación de etiquetador
 resumen [401](#)

Transformación de excepción
 Vista Asignación problemática [164](#)

Transformación de excepción de registros duplicados
 clústeres [274](#)
 cómo generar una tabla de registros duplicados [276](#)
 configurar [286](#)
 consolidación automática [275](#)
 ejemplo [280](#)
 ejemplo de asignación [280](#)
 ejemplo de salida [283](#)
 ejemplo de salida del clúster [284](#)
 ejemplo de valores de configuración [282](#)
 Grupo de salida estándar [278](#)
 grupos de salida [278](#)
 niveles de seguimiento [279](#)
 ordenar registros [279](#)
 propiedades avanzadas [279](#)
 puertos [277](#)
 resumen [273](#)
 Vista de configuración [275](#)

Transformación de excepción de registros incorrectos
 asignación [159](#)
 configurar [166](#)
 ejemplo [167](#)
 ejemplo de salida de problemas [170](#)
 ejemplo de salida de registros incorrectos [169](#)
 flujo de proceso [158](#)
 grupos de puertos [161](#)
 grupos de salida [157](#)

Transformación de excepción de registros incorrectos (*continuado*)
 mapplet de ejemplo [167](#)
 problemas de calidad [160](#)
 propiedades avanzadas [165](#)
 puertos de entrada [161](#)
 puertos de problema de calidad en blanco [157](#)
 puertos de salida [162](#)
 resumen [156](#)
 salida de ejemplo [169](#)
 salida de ejemplo de registros correctos [171](#)
 Vista de configuración [162](#)

transformación de expresión
 configuración del puerto de salida [300](#)
 entorno no nativo [305](#)
 expresión dinámica [299](#)
 motor de Blaze [305](#)
 motor de Databricks Spark [305](#)
 motor de Spark [305](#)
 propiedades avanzadas [304](#)
 prueba [289](#)
 resumen [287](#)
 tipos de puerto [288](#)
 uso de variables [47](#)

transformación de filtro
 asignaciones dinámicas [307](#)
 condición de filtro [308](#)
 entorno no nativo [310](#)
 motor de Blaze [311](#)
 parámetro de condición de filtro [308](#)
 propiedades avanzadas [310](#)
 resumen [306](#)

Transformación de filtro
 consejos sobre el rendimiento [310](#)
 filas con valores nulos [310](#)

transformación de fusión
 entorno no nativo [561](#)

Transformación de fusión
 resumen [560](#)

Transformación de generador de claves
 niveles de seguimiento [399](#)
 propiedades avanzadas [399](#)
 resumen [395](#)

transformación de generador de secuencia
 ciclo [629](#)
 Ciclo [630](#)
 crear una clave compuesta [627](#)
 entorno no nativo [637](#)
 intervalo de valores [630](#)
 motor de Blaze [637](#)
 motor de Spark [637](#)
 propiedades [629](#), [631](#), [632](#)
 propiedades avanzadas [631](#)
 puerto NEXTVAL [627](#)
 puertos [626](#)
 restablecer [631](#)
 resumen [626](#)
 usar la función IIF para reemplazar claves que faltan [628](#)
 valor inicial [629](#)

Transformación de generador de secuencia
 crear [635](#)
 mantener orden de filas [631](#)
 propiedad Incrementar por [630](#)

transformación de incorporación
 cómo configurar el orden de clasificación [385](#)
 entrada ordenada [385](#)
 unión normal [382](#)

Transformación de incorporación
 cómo bloquear canales [391](#)

- Transformación de incorporación (*continuado*)
 - cómo guardar filas maestras en la memoria caché [392](#)
 - cómo unir datos de un mismo origen [389](#)
 - no ordenada [391](#)
 - ordenada [391](#)
 - outer join completo [384](#)
 - outer join de detalles [384](#)
 - resumen [682](#)
 - tipos de unión [382](#)
- transformación de Java
 - métodos API [347](#)
 - resumen [319](#)
- transformación de jerárquica a relacional
 - configurar puertos [316](#)
 - crear [317](#)
 - crear puertos [317](#)
 - desarrollo [317](#)
 - descripción [312](#)
 - ejemplo [313](#)
 - probar en el visor de datos [318](#)
 - puertos [315](#)
- transformación de lectura
 - resumen [583](#)
- Transformación de macro
 - asignaciones dinámicas [474](#)
 - descripción [473](#)
 - entorno no nativo [476](#)
 - instrucción macro [473](#)
 - parámetros [474](#)
 - puertos dinámicos [474](#)
- transformación de media ponderada
 - entorno no nativo [745](#)
- transformación de normalizador
 - arrastrar puertos desde otros objetos [555](#)
 - campos de ocurrencia múltiple [539](#)
 - crear [554](#)
 - editar un grupo de salida [552](#)
 - ejemplo de asignación [556](#)
 - ejemplo de grupos de entrada y salida [557](#)
 - ejemplo de uso de mayúsculas y minúsculas [556](#)
 - entorno no nativo [559](#)
 - GCID [539](#)
 - generación de claves [553](#)
 - grupos de salida [549](#), [551](#)
 - puertos de entrada [542](#)
 - subregistros [540](#)
- Transformación de normalizador
 - ejemplo de definición [557](#)
 - ejemplo de salida de asignación [558](#)
 - fusionar campos [543](#), [544](#)
 - jerarquía de entrada [541](#)
 - propiedades avanzadas [554](#)
 - registros de ocurrencia múltiple [540](#)
 - resumen [538](#)
- transformación de ordenación
 - asignaciones dinámicas [639](#)
 - entorno no nativo [647](#)
 - ficha Ordenar [640](#)
 - motor de Blaze [647](#)
 - motor de Databricks Spark [648](#)
 - motor de Spark [648](#)
- Transformación de ordenación
 - memorias caché [644](#)
 - resumen [638](#)
 - tamaño de la memoria caché [72](#)
- transformación de rango
 - asignaciones dinámicas [576](#)
 - entorno no nativo [581](#)
- transformación de rango (*continuado*)
 - motor de Blaze [581](#)
 - motor de Databricks Spark [582](#)
 - motor de Spark [582](#)
 - uso de variables [47](#)
- Transformación de rango
 - cómo definir grupos para [578](#)
 - memorias caché [580](#)
 - opciones [576](#)
 - propiedades avanzadas [580](#)
 - puerto de rango [578](#)
 - Puerto RANKINDEX [577](#)
 - puertos [577](#)
 - resumen [575](#)
 - tamaño de la memoria caché [72](#)
- Transformación de relacional a jerárquica
 - crear [598](#)
 - crear puertos [599](#)
 - desarrollo [598](#)
 - descripción [594](#)
 - ejemplo [595](#)
 - puertos [598](#)
- transformación de SQL
 - Parámetros INOUT [670](#)
 - cardinalidad entre fila de entrada y fila de salida [660](#)
 - cómo definir la conexión de la base de datos [668](#)
 - cómo definir la consulta [658](#)
 - cómo definir puertos de salida [652](#)
 - cómo restringir filas de salida [661](#)
 - conexión de tiempo de ejecución [674](#)
 - conjuntos de resultados [671](#)
 - continuar ante error de SQL [663](#)
 - creación desde un procedimiento almacenado [676](#)
 - crear como vacía [675](#)
 - descripción de puertos de entrada [652](#)
 - ejemplo [665](#)
 - enlace de parámetro [658](#)
 - instrucción de consulta [668](#)
 - llamar a un procedimiento almacenado [670](#)
 - número de filas afectadas [655](#)
 - número de salida de filas [661](#)
 - optimización de inserción [664](#)
 - optimización de primera selección [664](#)
 - parámetros de procedimiento almacenado [670](#)
 - procedimientos almacenados [669](#)
 - propiedades de la optimización de inserción [665](#)
 - puerto de valor de devolución [669](#)
 - puerto SQLException [654](#)
 - puertos [651](#)
 - puertos de transferencia [653](#)
 - sustitución de cadenas en consultas [659](#)
 - Vista Propiedades avanzadas [655](#)
- Transformación de SQL
 - resumen [650](#)
- transformación de unión
 - entorno no nativo [685](#)
 - motor de Databricks Spark [685](#)
 - outer join principal [383](#)
- transformación del generador de claves
 - entorno no nativo [400](#)
- transformación del procesador de datos
 - entorno no nativo [265](#)
- transformación del Procesador de datos
 - Configuración de XMap [251](#)
 - puertos de parámetros de servicio [243](#)
- Transformación del Procesador de datos
 - configuración de codificación [246](#)
 - configuración de control de salida [248](#)

Transformación del Procesador de datos (*continuado*)

- configuración de procesamiento [250](#)
- Configuración de XML [251](#)
- configuraciones [246](#)
- crear [257](#)
- descripción [240](#), [241](#)
- exportar como un servicio [262](#)
- probar en el visor de datos [261](#)
- puertos [242](#)
- puertos de entrada [242](#)
- puertos de parámetros de servicio [243](#)
- puertos de salida [243](#)
- registros de usuario [256](#)
- vistas [241](#)

Transformación del procesamiento de datos
componente de inicio [244](#)

transformación del validador de direcciones
entorno no nativo [137](#)

transformación Java

- entorno no nativo [344](#)
- motor de Blaze [344](#)
- motor de Spark [345](#)

transformaciones

- configurar puertos en Excel [69](#)
- copiar metadatos [69](#)
- activas [33](#)
- archivos de memoria caché [74](#)
- cómo crear [60](#)
- cómo desarrollar [41](#)
- cómo editar reutilizables [58](#)
- conectadas [33](#)
- controlar errores [54](#)
- crear particiones de la memoria caché [77](#)
- editar en Excel [70](#)
- entorno Hadoop [36](#)
- expresiones [43](#)
- Generador de secuencia [626](#)
- Java [319](#)
- memorias caché [72](#)
- no conectadas [33](#)
- no reutilizables [59](#)
- optimización del tamaño de memoria caché [78](#)
- resumen [32](#)
- reutilizables [58](#), [59](#)
- tamaño de la memoria caché [75](#), [77](#)
- validación de expresiones [45](#)
- varios grupos [42](#)

transformaciones activas

- descripción [33](#)
- Java [319](#), [320](#)
- Rango [575](#)

transformaciones conectadas

- Generador de secuencia [626](#)
- Java [319](#)
- Rango [575](#)

transformaciones de bloqueo

- descripción [42](#)

transformaciones de Java

- activas [320](#)
- almacenar metadatos [356](#)
- Ámbito de transformación [325](#)
- asignaciones con fallo en [349](#)
- classpath a nivel de asignación [327](#)
- código Java [328](#)
- cómo compilar [338](#)
- cómo comprobar valores nulos en [353](#)
- cómo crear [337](#), [338](#)
- cómo crear puertos [324](#)

transformaciones de Java (*continuado*)

- cómo crear snippets de código Java [329](#)
- cómo definir valores nulos en [355](#)
- cómo diseñar [323](#)
- cómo identificar el origen de errores de compilación [340](#)
- cómo obtener el tipo de fila de entrada [350](#)
- cómo restablecer variables en [355](#)
- conversión de tipo de datos [320](#)
- errores ajenos al código de usuario [340](#)
- errores de código de usuario [340](#)
- errores de compilación [339](#)
- ficha Al final [333](#)
- ficha Aplicaciones auxiliares [331](#), [332](#)
- ficha Código completo [334](#)
- ficha En entrada [333](#)
- ficha Funciones [334](#)
- ficha Importaciones [331](#), [332](#)
- método defineJExpression [348](#)
- método failSession [349](#)
- método generateRow [350](#)
- método getInRowType [350](#)
- Método getMetadata [351](#)
- método incrementErrorCount [351](#)
- método invokeJExpression [352](#)
- método isNull [353](#)
- método logError [354](#)
- método logInfo [354](#)
- método resetNotification [355](#)
- método setNull [355](#)
- Método storeMetadata [356](#)
- métodos API [347](#)
- pasivas [320](#)
- procesamiento al nanosegundo [325](#)
- procesamiento de alta precisión [325](#)
- propiedades avanzadas [325](#)
- puertos de entrada [324](#)
- puertos de salida [324](#)
- recuperando metadatos [351](#)
- registros [354](#)
- Sin estado [325](#)
- solución de problemas [339](#)
- tipos de datos primitivos Java [320](#)
- valores de puerto predeterminados [324](#)

transformaciones de lectura

- crear a partir de objetos de datos relacionales [592](#)

transformaciones de servicio web

- Columna Ubicación [731](#)

transformaciones no conectadas

- transformación de búsqueda [418](#)

transformaciones no reutilizables

- descripción [59](#)

transformaciones pasivas

- Generador de secuencia [626](#)
- Java [319](#), [320](#)
- resumen [33](#)

transformaciones reutilizables

- cómo editar [58](#)
- descripción [58](#)

Transformador

- descripción [240](#)

Transmisión por secuencias

- descripción [240](#)

tratar fallo como error

- cómo habilitar una transformación de consumidor de servicio web [707](#)

U

- umbral inferior
 - configurar [162](#), [275](#)
- umbral superior
 - configurar [162](#), [275](#)
- URL de punto final
 - Transformación de consumidor de servicio web [700](#)
 - Transformación de consumidor de servicio web REST [608](#)
- URL dinámica
 - Transformación de consumidor de servicio web [700](#)
- Usar fusión de Hive
 - opción [688](#)

V

- validación
 - valores predeterminados [54](#), [57](#)
- Valor actual (propiedad)
 - transformación de generador de secuencia [629](#)
- Valor final (propiedad)
 - transformación de generador de secuencia [629](#)
- valor inicial
 - transformación de generador de secuencia [629](#)
- Valor inicial (propiedad)
 - transformación de generador de secuencia [629](#)
- valores de cadena
 - enmascaramiento de datos de claves [213](#)
 - enmascaramiento de datos personalizados [210](#)
- valores de fecha
 - enmascaramiento de datos aleatorio [211](#)
- valores de fecha y hora
 - enmascaramiento de datos [214](#)
- valores de puerto
 - transformaciones de Java [324](#)
- valores nulos
 - cómo comprobar en transformaciones de Java [353](#)
 - cómo definir para transformaciones Java [355](#)
 - omisión [54](#)
 - reemplazar con una constante [52](#)
- valores numéricos
 - enmascaramiento aleatorio [210](#)
 - enmascaramiento de claves [214](#)
- valores predeterminados
 - definidos por el usuario [51](#)
 - enmascaramiento de datos [230](#)
 - introducir [54](#), [57](#)
 - puertos de entrada [50](#), [51](#)
 - puertos de salida [50](#), [51](#)
 - puertos de transferencia [50](#), [51](#)
 - reglas de [56](#)
 - resumen [50](#)
 - validación [54](#), [57](#)
- valores que faltan
 - reemplazar con el generador de secuencia [628](#)

- variables
 - inicializaciones [50](#)
 - orden de evaluación de puertos [49](#)
 - resultados de procedimiento almacenado, captura [49](#)
 - resumen [47](#)
 - transformaciones de Java [331](#)
- variables de asignación
 - en el reemplazo de SQL de búsqueda [419](#)
- variables de instancia
 - transformaciones de Java [331](#)
- variables estáticas
 - transformaciones de Java [331](#)
- variables locales
 - resumen [47](#)
- varias coincidencias
 - transformación de búsqueda [446](#)
- varios grupos
 - transformaciones [42](#)
- vínculo
 - uno a muchos [63](#)
 - uno a uno [63](#)
- Vista Asignación problemática
 - Transformación de excepción de registros incorrectos [164](#)
- Vista de ayuda de script
 - Transformación del Procesador de datos [241](#)
- Vista de componentes
 - Transformación del Procesador de datos [241](#)
- Vista de configuración
 - Ejemplo de transformación de excepción de registros incorrectos [168](#)
 - Transformación de excepción de registros duplicados [275](#)
 - Transformación de excepción de registros incorrectos [162](#)
 - Transformación del Procesador de datos [246](#)
- Vista de eventos
 - Transformación del Procesador de datos [254](#)
- Vista de origen hexadecimal, procesador de datos
 - descripción [241](#)
- vista de referencias
 - transformación de jerárquica a relacional [316](#)
 - Transformación de relacional a jerárquica [598](#)
- Vista de referencias
 - Transformación del Procesador de datos [245](#)
- Vista de repositorio
 - Transformación del Procesador de datos [241](#)
- Vista del Procesador de datos de origen hexadecimal
 - descripción [241](#)
- Vista Eventos del procesador de datos
 - Transformación del Procesador de datos [254](#)
 - visualización del registro de eventos [256](#)
- vista Normalizador
 - descripción [541](#)
- vista Puertos
 - salida de transformación de SQL [652](#)
- Vista Tiempo de ejecución
 - transformación de SQL [668](#)