



Informatica® Integración de Data Quality
para PowerCenter

10.1.1 HotFix 2

Guía del usuario

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.

Las partes de este software o la documentación están sujetas a derechos de autor de terceros. Se incluyen con el producto los avisos obligatorios de terceros.

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

EXENCIÓN DE RESPONSABILIDAD: Informatica LLC proporciona esta documentación "tal cual" sin garantía de ningún tipo, ya sea expresa o implícita, incluidas, entre otras, las garantías implícitas de no incumplimiento, de adecuación para un propósito determinado y de validez para el comercio. Informatica LLC no garantiza que este software o esta documentación estén libres de errores. La información proporcionada en este software o en esta documentación puede contener imprecisiones técnicas o errores tipográficos. La información de este software y esta documentación está sujeta a cambios en cualquier momento sin previo aviso.

AVISOS

Este producto de Informatica (el "Software") incluye ciertos controladores (los "Controladores DataDirect") de DataDirect Technologies, una empresa operativa de Progress Software Corporation ("DataDirect") que están sujetos a los términos y condiciones siguientes:

1. LOS CONTROLADORES DATADIRECT SE PROPORCIONAN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUIDAS, ENTRE OTRAS, LAS GARANTÍAS IMPLÍCITAS DE NO INCUMPLIMIENTO, DE ADECUACIÓN PARA UN PROPÓSITO DETERMINADO Y DE VALIDEZ PARA EL COMERCIO.
2. EN NINGÚN CASO DATADIRECT NI SUS PROVEEDORES DE TERCEROS SERÁN RESPONSABLES ANTE EL USUARIO FINAL POR NINGÚN DAÑO DIRECTO, INDIRECTO, FORTUITO, ESPECIAL, CONSECUENTE, NI DE NINGÚN OTRO TIPO, RESULTANTE DEL USO DE LOS CONTROLADORES ODBC, INDEPENDIENTEMENTE DE SI SE HA AVISADO O NO DE LOS POSIBLES DAÑOS POR ADELANTADO. ESTAS LIMITACIONES SE APLICAN A TODAS LAS DEMANDAS JUDICIALES, INCLUIDAS, ENTRE OTRAS, AQUELLAS POR INCUMPLIMIENTO DE CONTRATO, INCUMPLIMIENTO DE LA GARANTÍA, NEGLIGENCIA, RESPONSABILIDAD ESTRUCTA, TERGIVERSACIÓN Y OTROS AGRAVIOS.

La información contenida en esta documentación está sujeta a cambios sin previo aviso. Si encuentra algún problema en esta documentación, escribanos 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: 2018-07-03

Tabla de contenido

Prefacio	4
Documentación de Informatica	4
Informatica Network.	4
Base de conocimiento de Informatica.	4
Documentación de Informatica	4
Matrices de disponibilidad de productos de Informatica.	5
Informatica Velocity.	5
Catálogo de soluciones de Informatica.	5
Servicio internacional de atención al cliente de Informatica.	5
 Capítulo 1: Integración de Informatica Data Quality.....	6
Resumen de integración de Data Quality.	6
Características y funciones de Data Quality.	7
Programas de instalación de integración.	8
Exportación de objetos en PowerCenter.	8
Instalación de datos de referencia.	8
Matriz de integración.	9
Reglas y directrices relativas a la integración de productos.	10
Comportamiento de las transformaciones de Data Quality en PowerCenter.	10
 Capítulo 2: Instalación.....	14
Visión general del proceso de instalación.	14
Requisitos del sistema.	14
Antes de la instalación.	15
Instalación del complemento Servidor.	15
Instalación en Windows.	15
Instalación en UNIX.	16
Instalación del complemento Cliente.	16
Registro del complemento con PowerCenter.	17
 Capítulo 3: Requisitos de los datos de referencia.....	18
Visión general de los datos de referencia.	18
Requisitos de archivos de datos de referencia.	19
Requisitos de los datos de referencia de direcciones.	20
Requisitos de los datos de llenado de identidad.	21
Requisitos del modelo probabilístico y del modelo clasificador.	23
Ubicaciones del modelo probabilístico y del modelo clasificador.	23
Asignación de memoria del servicio de integración de PowerCenter.	24
 Índice.....	25

Prefacio

Los usuarios de PowerCenter® pueden trabajar con las asignaciones y los objetos de datos de referencia que importen desde un entorno de Data Quality. La *Guía del usuario de la integración de Informatica Data Quality para PowerCenter* está destinada a usuarios de PowerCenter que ejecutan asignaciones de calidad de datos y gestionan los objetos de datos de referencia asociados.

Documentación de Informatica

Informatica Network

Informatica Network incluye el servicio internacional de atención al cliente de Informatica, la base de conocimiento de Informatica y otros recursos de producto. Para acceder a Informatica Network, visite <https://network.informatica.com>.

Un miembro puede:

- Acceder a todos sus recursos de Informatica en un solo lugar.
- Busque recursos de producto, como documentación, preguntas frecuentes y mejores prácticas en la base de conocimiento.
- Vea la información de disponibilidad del producto.
- Revisar los casos de asistencia.
- Buscar su red de grupos de usuarios de Informatica locales y colaborar con sus iguales.

Base de conocimiento de Informatica

Utilice la base de conocimiento de Informatica para buscar recursos de producto como documentación, artículos de procedimientos, mejores prácticas y PAM en la red de Informatica.

Para acceder a la base de conocimiento, visite <https://kb.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

Para obtener la documentación más reciente del producto, consulte la base de conocimiento de Informatica en https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

Si tiene preguntas, comentarios o ideas relacionadas con esta documentación, póngase en contacto con el equipo de documentación de Informatica enviando un correo electrónico a infa_documentation@informatica.com.

Matrices de disponibilidad de productos 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 de datos y destinos admitidos por una versión de un producto. Si es miembro de la red de Informatica, puede acceder a las PAM en <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity es un conjunto de sugerencias y mejores prácticas desarrollado por los servicios profesionales de Informatica. Desarrollado a partir de la experiencia real de cientos de proyectos de administración de datos, Informatica Velocity representa el conocimiento conjunto de nuestros asesores, los cuales han trabajado con organizaciones de todo el mundo para planificar, desarrollar, implementar y mantener con éxito soluciones de administración de datos.

Si es miembro de la red de Informatica, puede acceder a los 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. Al aprovechar cualquiera de los cientos de soluciones de los desarrolladores y los socios de Informatica, puede mejorar la productividad y acelerar el tiempo de implementación en los proyectos. Puede acceder al 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 del soporte en línea en la red de Informatica.

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:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

Si es miembro de la red de Informatica, puede utilizar el soporte en línea en <http://network.informatica.com>.

CAPÍTULO 1

Integración de Informatica Data Quality

Este capítulo incluye los siguientes temas:

- [Resumen de integración de Data Quality, 6](#)
- [Características y funciones de Data Quality, 7](#)
- [Programas de instalación de integración, 8](#)
- [Exportación de objetos en PowerCenter, 8](#)
- [Instalación de datos de referencia, 8](#)
- [Matriz de integración, 9](#)
- [Reglas y directrices relativas a la integración de productos, 10](#)
- [Comportamiento de las transformaciones de Data Quality en PowerCenter, 10](#)

Resumen de integración de Data Quality

Los usuarios de Informatica pueden crear y ejecutar asignaciones en Informatica Data Quality. Los usuarios guardan las asignaciones y otros objetos en el repositorio de modelos. Los usuarios pueden exportar los objetos que se van a usar en PowerCenter. Como usuario de PowerCenter, puede importar los objetos al repositorio de PowerCenter y puede añadir los objetos a sesiones.

Utilice los objetos para analizar y mejorar el contenido y la estructura de los datos.

Puede utilizar los objetos que los usuarios han creado en Data Quality 10.1.1 HotFix 2 con las siguientes versiones de PowerCenter:

- PowerCenter 10.1.1 con la revisión más reciente instalada.
- PowerCenter 10.1.
- PowerCenter 9.6.1 con la revisión más reciente instalada.

Si utiliza las versiones actuales de Data Quality y PowerCenter, puede importar objetos cuyo origen sea el repositorio de modelos sin necesidad de instalar nada más. Si utiliza una versión anterior de PowerCenter, instale y registre la versión actual del complemento de integración de Informatica Data Quality.

Características y funciones de Data Quality

Informatica Data Quality usa las aplicaciones de Informatica para crear objetos, guardar objetos en el repositorio de modelos y ejecutar asignaciones.

Puede que necesite instalar el complemento de Informatica Data Quality para PowerCenter si no usa las versiones actuales del producto.

Utilice Informatica Data Quality para diseñar y ejecutar asignaciones que consigan los siguientes objetivos:

- Crear perfiles de los datos. La creación de perfiles revela el contenido y la estructura de los datos. La creación de perfiles es un paso clave en cualquier proyecto de datos, ya que permite identificar las ventajas y desventajas de los datos y, además, le ayuda a definir el plan de proyecto.
- Crear cuadros de mando para revisar la calidad de los datos. Un cuadro de mando es una representación gráfica de las mediciones de calidad de un perfil.
- Estandarizar valores de datos. Estandarice datos para eliminar errores e inconsistencias detectadas cuando se ejecuta un perfil. Se pueden estandarizar variaciones en la puntuación, formato y ortografía. Por ejemplo, puede asegurarse de que los valores de ciudad, estado y código postal sean uniformes.
- Analizar registros. Analice registros de datos para mejorar la estructura de los registros y obtener información adicional de los datos. Un campo de datos de forma libre se puede dividir en campos que contengan distintos tipos de información. También se puede añadir información a los registros. Por ejemplo, puede marcar registros de clientes como de clientes empresariales o personales.
- Validar direcciones postales. La validación de direcciones evalúa y mejora la exactitud y factibilidad de entrega de los datos de direcciones postales. La validación de direcciones corrige los errores en las direcciones y completa las direcciones parciales comparando registros de direcciones respecto a datos de referencia de agencias postales nacionales. La validación de direcciones también puede añadir información postal para agilizar la entrega del correo y reducir los costes del mismo.
- Buscar registros duplicados. El análisis de registros duplicados compara un conjunto de registros con otro para encontrar valores similares o coincidentes en las columnas de datos seleccionadas. El usuario establece el nivel de similitud que indica una buena coincidencia entre los valores de los campos. También se puede establecer el peso relativo que se le da a cada columna en los cálculos de coincidencia. Por ejemplo, se puede priorizar la información de apellido sobre la de nombre de pila.
- Crear y ejecutar reglas de calidad de datos. Informatica proporciona reglas creadas previamente que se pueden ejecutar o editar para adaptarse a los objetivos de cada proyecto. Se pueden crear reglas en la herramienta Developer.
- Colaborar con usuarios de Informatica. Las reglas y tablas de datos de referencia que se añaden al repositorio de modelos están disponibles para los usuarios de las herramientas Developer y Analyst. Los usuarios pueden colaborar en proyectos; asimismo, varios usuarios pueden ser propietarios de objetos en distintas etapas de un proyecto.
- Exportar asignaciones a PowerCenter. Puede exportar asignaciones a PowerCenter para volver a utilizar los metadatos para la integración de datos físicos o para crear servicios web.

Programas de instalación de integración

Si utiliza distintas versiones de Data Quality y PowerCenter para crear y guardar objetos, instale el complemento de Informatica Data Quality para PowerCenter. Descargue los programas de instalación del complemento de Informatica.

Instale la versión del complemento para la versión de PowerCenter que desea utilizar. El complemento tiene programas de instalación distintos para las versiones Cliente y Servidor.

Ejecute el programa de instalación cliente en el equipo cliente de PowerCenter para ver los mapplets y las asignaciones del repositorio y añadirlos a las sesiones. Ejecute el programa de instalación Servidor en el equipo del servicio de integración de PowerCenter para ejecutar las sesiones en flujos de trabajo.

Después de instalar el complemento en el equipo del servicio de integración, registre el complemento con el repositorio de PowerCenter.

Exportación de objetos en PowerCenter

Use Developer Tool para exportar objetos como mapplets y asignaciones del repositorio de modelos de Informatica. Puede exportar objetos desde el repositorio de modelos al repositorio de PowerCenter y también desde el repositorio de modelos al sistema de archivos. Use PowerCenter Repository Manager para importar objetos XML en el repositorio de PowerCenter.

- Para obtener información sobre la exportación de objetos del repositorio de modelos, consulte la *Guía de asignación de Informatica Developer* y la *Guía de Informatica Developer Tool*.
- Para obtener información sobre la importación de objetos en el repositorio de PowerCenter, consulte la *Guía del repositorio de PowerCenter*.

Instalación de datos de referencia

Si un mapplet o una asignación lee archivos de datos de referencia, instale los archivos de datos de referencia en el equipo del servicio de integración de PowerCenter.

Puede utilizar Developer Tool para exportar los datos de referencia desde el repositorio de modelos cuando se exporta un mapplet o asignación. Developer Tool exporta los datos de referencia como un archivo comprimido. Extraiga el archivo y copie los datos de referencia en el equipo del servicio de integración.

También puede usar el programa de instalación de contenido de Data Quality para añadir datos de referencia al equipo del servicio de integración. Ejecute el programa de instalación de contenido para instalar archivos de datos de referencia de direcciones y archivos de datos de poblaciones de identidad.

Matriz de integración

El nivel de integración entre Data Quality y PowerCenter depende de las versiones de producto respectivas. Puede exportar objetos y datos de referencia desde varias versiones de Data Quality a varias versiones de PowerCenter. No puede exportar objetos ni datos de referencia de PowerCenter a Data Quality.

La siguiente tabla muestra las distintas versiones del producto que puede integrar:

Nombre de producto	PowerCenter 10.1.1 HotFix 2	PowerCenter 10.1.1 HotFix 1	PowerCenter 10.1.1	PowerCenter 10.1	PowerCenter 10.0	PowerCenter 9.6.1
Data Quality 10.1.1 HotFix 2	Sí	No	No	Instalar complementos	No	Instalar complementos
Data Quality 10.1.1 HotFix 1	No	Sí	No	Instalar complementos	No	Instalar complementos
Data Quality 10.1.1	No	No	Sí	Instalar complementos	No	Instalar complementos
Data Quality 10.1	No	No	No	Sí	No	Instalar complementos
Data Quality 10.0	No	No	No	No	Sí	Instalar complementos
Data Quality 9.6.1	No	No	No	No	No	Sí

Los valores de tabla representan los siguientes tipos de integración:

Sí

Puede exportar objetos y archivos desde un entorno de Data Quality a otro de PowerCenter. No necesita instalar o registrar ningún complemento. Puede ejecutar objetos de calidad de datos en un repositorio de PowerCenter que actualice a la versión especificada.

Instalar complementos

Puede exportar objetos y archivos desde un entorno de Data Quality a otro de PowerCenter. Instale y registre los complementos de cliente y servidor para la versión de PowerCenter que utilice. Puede ejecutar objetos de calidad de datos en un repositorio de PowerCenter que actualice a la versión especificada.

No

No podrá utilizar los objetos y archivos que defina en un entorno de Data Quality en otro de PowerCenter.

Reglas y directrices relativas a la integración de productos

Puede exportar objetos de calidad de datos de distintas versiones de Informatica Data Quality a distintas versiones de PowerCenter. Instale el complemento de integración de la versión de Informatica Data Quality desde la que desee exportar los objetos.

Considere las siguientes reglas y directrices cuando integre Informatica Data Quality con PowerCenter:

- No instale el complemento de integración si la versión de producto del repositorio de PowerCenter es la misma que la versión de producto del repositorio de modelos.
- No se pueden ejecutar dos complementos de integración de Data Quality simultáneamente en un solo equipo de PowerCenter. PowerCenter utiliza el complemento instalado más recientemente. Por ejemplo, si instala el complemento de integración de Data Quality 9.6.1 en un equipo con PowerCenter 9.5.1, no podrá seguir importando y ejecutando objetos de calidad de datos procedentes de un repositorio de modelos de la versión 9.5.1. Podrá seguir ejecutando cualquier objeto de calidad de datos que se haya importado en el repositorio de PowerCenter antes de instalar el complemento 9.6.1.
- No se pueden editar mapplets o asignaciones de calidad de datos en un repositorio de PowerCenter.

Comportamiento de las transformaciones de Data Quality en PowerCenter

Quando se importa un mapplet de calidad de datos o una asignación del repositorio de modelos al repositorio de PowerCenter, el comportamiento de las transformaciones de calidad de datos puede cambiar.

Consideraciones sobre la transformación de consolidaciones

Informatica Data Quality actualizó la estrategia predeterminada de la transformación de consolidación en la versión 9.5.1. En la versión 9.5.1 y posteriores, la transformación de consolidación selecciona el registro con el ID de fila más alto como registro superviviente. En la versión 9.5 y anteriores, la transformación de consolidación utiliza el valor más frecuente que no está en blanco de una columna especificada para identificar el registro superviviente.

Si actualiza un repositorio de PowerCenter que contiene una transformación de consolidación a la versión 9.5.1 o posterior, la actualización mantiene el valor más frecuente que no está en blanco como valor predeterminado. Si importa una asignación que contiene una transformación de consolidación con valores predeterminados desde un repositorio de modelos de la versión 9.5.1 o un repositorio posterior, la transformación utiliza el ID de fila más alto para identificar el registro superviviente.

Conversión a mapplets

Las siguientes transformaciones se pueden convertir en mapplets al importarlas al repositorio de PowerCenter:

- Consolidación
- Excepción
- Generador de claves
- Coincidencia

Quando se importa una asignación que contiene una o varias de estas transformaciones, PowerCenter convierte cada transformación en un mapplet que contiene la configuración expandida.

Cuando se importa un mapplet que contiene una o varias de estas transformaciones, PowerCenter expande la configuración de cada transformación en varias transformaciones.

PowerCenter puede añadir transformaciones a la transformación de calidad de datos o a la expansión del mapplet para completar la lógica de transformación. Por ejemplo, PowerCenter expande la transformación Coincidencia con transformaciones adicionales.

Consideraciones sobre las transformaciones de decisión

No puede exportar una asignación o un mapplet que contenga una transformación de decisión a PowerCenter si el script de transformación incluye un parámetro del sistema. La operación de exportación no puede convertir el parámetro del sistema en un valor que PowerCenter pueda utilizar. Antes de exportar una asignación o un mapplet con una transformación de decisión que utilice un parámetro del sistema, reemplace el parámetro con un valor adecuado.

Consideraciones sobre la transformación de excepciones

PowerCenter no puede utilizar una transformación de excepción originada en un repositorio de PowerCenter anterior a la versión 9.5. Si actualiza un repositorio de PowerCenter a la versión 9.5 o a una posterior y el repositorio contiene una asignación con una transformación de excepciones, exporte la asignación de nuevo. Cree de nuevo la asignación en Developer tool si fuera necesario.

Cuando ejecute una asignación que incluye una transformación de excepción, puede actualizar los registros incorrectos o las tablas de salida de registros duplicados en la Herramienta del analista. Si va a actualizar una tabla de registros incorrectos en la Herramienta del analista, deberá habilitar la alta precisión en la sesión que ejecuta la asignación.

Consideraciones sobre la transformación de coincidencia

No puede ejecutar una asignación de coincidencia de identidad en PowerCenter si la asignación lee o escribe datos de índice de identidad en tablas de base de datos. Para realizar el análisis de coincidencias de identidad en PowerCenter, configure la transformación de coincidencia en Developer tool para escribir datos de índice de identidad en archivos temporales.

Transformaciones multiestrategia

Puede definir varias estrategias de transformación de datos dentro de cada una de las siguientes transformaciones en Developer tool:

- Mayúsculas y minúsculas
- Clasificador
- Decisión
- Generador de claves
- Etiquetador
- Coincidencia
- Fusionar
- Analizador
- Estandarizador

Cuando se importa un mapplet o asignación que contiene una transformación de varias estrategias, cada estrategia se convierte en una transformación independiente.

Configuración de las particiones

Antes de ejecutar una tarea de sesión que contiene una asignación de calidad de datos en una malla, identifique la versión de PowerCenter en la que ha creado los nodos de malla. Si ha creado los nodos en PowerCenter 9.5.1 HotFix 2 o en una versión anterior de PowerCenter, la asignación podría fallar.

Para comprobar que las asignaciones de calidad de datos pueden ejecutarse en todos los nodos de la malla, realice una de las tareas siguientes:

- Revise las propiedades avanzadas en cualquier servicio de integración de PowerCenter que distribuye tareas entre nodos de la malla. Configure cada servicio de integración de PowerCenter para que omita los requisitos de recursos cuando distribuya tareas. De forma predeterminada, las propiedades avanzadas especifican que el servicio de integración de PowerCenter omite los requisitos de recursos.
- Vuelva a crear todos los nodos que haya creado en PowerCenter 9.5.1 HotFix 2 o en una versión de PowerCenter anterior. Cuando vuelva a crear el nodo, PowerCenter aplicará los archivos de configuración del nodo actual al nodo.

Cuando ejecuta una tarea de sesión en una malla, el proceso del Administrador de Data Transformation (DTM) distribuye los subprocesos de la sesión entre los nodos de la malla. Algunas transformaciones no admiten particiones en una malla. Cuando una transformación no admite particiones en una malla, el DTM crea un solo grupo de particiones para los subprocesos de la transformación y ejecuta los subprocesos en un solo nodo.

En la siguiente tabla se describe la configuración de las particiones para transformaciones de calidad de datos:

Transformación	Configuración de las particiones
Validación de direcciones	En la malla
Asociación	No admite particiones
Informe AV	Admite particiones localmente
Conversión de mayúsculas y minúsculas	En la malla
Clasificador	Admite particiones localmente
Clúster	No admite particiones
Comparación	En la malla
Consolidación	No admite particiones
Decisión	En la malla
Excepción	No admite particiones
Punto de ejecución	En la malla
Generador de claves	No admite particiones
Puntuación de clave	No admite particiones
Etiquetador	En la malla
Fusionar	En la malla
Generador de pares	No admite particiones
Analizador (análisis probabilístico)	Admite particiones localmente

Transformación	Configuración de las particiones
Analizador	En la malla
Estandarizador de tokens	En la malla
Analizador basado en peso	En la malla

CAPÍTULO 2

Instalación

Este capítulo incluye los siguientes temas:

- [Visión general del proceso de instalación, 14](#)
- [Requisitos del sistema, 14](#)
- [Antes de la instalación, 15](#)
- [Instalación del complemento Servidor, 15](#)
- [Instalación del complemento Cliente, 16](#)
- [Registro del complemento con PowerCenter, 17](#)

Visión general del proceso de instalación

El complemento de integración de Data Quality tiene un componente de servidor y un componente de cliente.

Lleve a cabo los siguientes pasos para instalar y registrar el complemento:

1. Ejecute el programa de instalación del servidor en el equipo del servicio de integración de PowerCenter.
2. Ejecute el programa de instalación del cliente en el equipo cliente de PowerCenter.
3. Registre el archivo `IDQZIntegration.xml` con el repositorio de PowerCenter.

Requisitos del sistema

El complemento de integración de Informatica Data Quality tiene los mismos requisitos del sistema que las aplicaciones y los servicios de PowerCenter en los que se instala.

Para obtener más información sobre los requisitos previos de PowerCenter y los requisitos del sistema, consulte la documentación de instalación de PowerCenter.

Antes de la instalación

Tenga en cuenta el siguiente comportamiento del programa de instalación si usa datos de referencia en PowerCenter.

El programa de instalación del complemento de integración escribe los archivos de propiedades en el equipo del servicio de integración de PowerCenter. Los archivos de propiedades especifican las ubicaciones de los archivos de datos de referencia que pueden leer las asignaciones de calidad de datos.

El programa de instalación escribe los archivos siguientes:

- AD50.cfg. Almacena las propiedades de la configuración para los datos de referencia de direcciones.
- CLASSIFIER.properties. Almacena las propiedades de la configuración para los datos del modelo clasificador.
- IDQTx.cfg. Almacena las propiedades de la configuración para los datos de poblaciones de identidad.
- NER.properties. Almacena las propiedades de la configuración para los datos del modelo probabilístico.

El programa de instalación añade los archivos al directorio siguiente:

```
[PowerCenter_installation]/server/bin
```

Si el programa de instalación encuentra un archivo de propiedades en el directorio `bin`, crea una copia de seguridad copia del archivo e instala un archivo nuevo. Para mantener la configuración actual de los datos de referencia después de la instalación, fusione el contenido de los archivos antiguos y nuevos.

Si instala PowerCenter 9.5.1 o una versión posterior, el proceso de instalación del servidor añade los archivos de propiedades en el directorio `bin` de manera predeterminada.

Requisitos previos de los datos de referencia de direcciones

Antes de instalar los datos de referencia de direcciones, detenga el servicio de integración de PowerCenter. Reinicie el servicio después de instalar los datos.

Instalación del complemento Servidor

El programa de instalación Servidor de integración de Informatica Data Quality añade los archivos de la aplicación de transformación de calidad de datos a la estructura de directorios del servicio de integración de PowerCenter.

Instalación en Windows

Ejecute el programa de instalación Servidor Windows para añadir el complemento al equipo del servicio de integración de PowerCenter en una plataforma Windows.

Cierre todas las aplicaciones antes de proceder con la instalación.

1. Inicie sesión en el equipo con una cuenta de usuario del sistema.
2. Extraiga el archivo Servidor de integración para Windows.
3. Acceda al directorio en el que ha extraído los archivos y ejecute `install.exe`.
4. Seleccione un idioma y haga clic en **Aceptar**.
5. En la pantalla de **Bienvenida**, haga clic en **Siguiente**.
6. Compruebe que no hay ningún servicio de PowerCenter ejecutándose y haga clic en **Aceptar**.

7. Especifique la ruta de acceso al directorio de instalación de PowerCenter. El programa de instalación asigna una ruta predeterminada a este directorio. Para seleccionar una ruta distinta, haga clic en **Elegir**.
8. Haga clic en **Siguiente**.
9. Revise el resumen previo a la instalación. Haga clic en **Anterior** para solucionar cualquier posible error.
10. Haga clic en **Instalar**.
11. Revise el resumen posterior a la instalación.
Para obtener más información sobre las tareas del proceso de instalación y ver las propiedades de la configuración de los componentes instalados, consulte los archivos de registro de la instalación.
12. Haga clic en **Listo**.

Instalación en UNIX

Ejecute el programa de instalación Servidor UNIX para añadir el complemento al equipo del servicio de integración de PowerCenter en una plataforma UNIX.

1. Inicie sesión en el equipo con una cuenta de usuario del sistema.
2. Cierre el resto de aplicaciones.
3. Extraiga el archivo Servidor de integración para UNIX.
4. Abra un shell de UNIX y acceda al directorio en el que ha extraído los archivos de instalación.
5. Introduzca `sh install.bin`.
6. Para aceptar English (Inglés) como idioma de instalación, pulse **Intro**.
7. Lea el texto de bienvenida y pulse **Intro**.
8. Escriba la ruta de acceso absoluta al directorio de instalación de PowerCenter, o pulse **Intro** para aceptar la ruta predeterminada.
9. Revise el resumen previo a la instalación. Para solucionar un error, escriba `back` (atrás) y pulse **Intro**.
10. Pulse **Intro** para iniciar la instalación.
11. Revise el resumen posterior a la instalación.
Para obtener más información sobre las tareas del proceso de instalación y ver las propiedades de la configuración de los componentes instalados, consulte los archivos de registro de la instalación.
12. Pulse **Intro** para finalizar el proceso de instalación.

Instalación del complemento Cliente

Ejecute el programa de instalación Cliente de integración de Data Quality en los equipos cliente de PowerCenter.

1. Cierre el resto de aplicaciones.
2. Extraiga el archivo Cliente de integración.
3. Abra el directorio en el que ha extraído los archivos.
4. Ejecute `install.exe`.
5. En la pantalla de **Bienvenida**, haga clic en **Siguiente**.

El programa de instalación muestra un mensaje solicitándole que cierre las aplicaciones cliente de PowerCenter antes de continuar con la instalación.

6. Compruebe que no hay ninguna aplicación cliente de PowerCenter ejecutándose y haga clic en **Aceptar**.
7. Especifique la ruta de acceso al directorio Cliente de PowerCenter. El programa de instalación asigna una ruta predeterminada a este directorio. Para seleccionar una ruta distinta, haga clic en **Elegir**.
8. Haga clic en **Siguiente**.
9. Revise el resumen previo a la instalación. Haga clic en **Anterior** para solucionar cualquier posible error.
10. Haga clic en **Instalar**.
11. Revise el resumen posterior a la instalación.
Para obtener más información sobre las tareas del proceso de instalación y ver las propiedades de la configuración de los componentes instalados, consulte los archivos de registro de la instalación.
12. Haga clic en **Listo**.

Registro del complemento con PowerCenter

Registre el complemento con cualquier repositorio de PowerCenter que desee que contenga los objetos de calidad de datos.

El programa de instalación del servicio de integración escribe el archivo de complemento `IDQZIntegration.xml` en el directorio `$INFA_HOME\server\bin\native` del equipo del servicio de integración de PowerCenter. Compruebe que tenga acceso a este archivo. Si fuera necesario, copie o descargue por FTP el archivo en el equipo del servicio de integración de PowerCenter.

- Para registrar el complemento, ejecute el comando `pmrep RegisterPlugin`. Incluya la opción `-n` cuando ejecute el comando.

Para obtener más información acerca del comando `pmrep RegisterPlugin`, consulte la *Referencia de comando de Informatica*.

Nota: Si no tiene los privilegios necesarios para registrar el complemento, póngase en contacto con el administrador del servicio de repositorio de PowerCenter.

CAPÍTULO 3

Requisitos de los datos de referencia

Este capítulo incluye los siguientes temas:

- [Visión general de los datos de referencia, 18](#)
- [Requisitos de archivos de datos de referencia, 19](#)
- [Requisitos de los datos de referencia de direcciones, 20](#)
- [Requisitos de los datos de llenado de identidad, 21](#)
- [Requisitos del modelo probabilístico y del modelo clasificador, 23](#)

Visión general de los datos de referencia

Las transformaciones de calidad de datos pueden leer datos de referencia. Si alguna transformación del repositorio de PowerCenter lee datos de referencia, compruebe que los archivos de datos de referencia están presentes en el entorno de PowerCenter.

Las transformaciones de calidad de datos en PowerCenter pueden usar los siguientes tipos de datos de referencia:

Archivos de datos de referencia

Archivos de texto que contienen las versiones estándar de los términos empresariales y las versiones alternativas de los términos. Puede editar los archivos de datos de referencia.

Archivos de datos de referencia de direcciones

Archivos de datos que contienen datos completos para todas las direcciones postales de un país. Puede instalar archivos de datos de referencia de direcciones en el equipo del servicio de integración de PowerCenter. Los archivos de datos de referencia de direcciones no se pueden editar.

Archivos de llenado de identidad

Archivos de datos que definen distintos tipos de identidad y contienen algoritmos utilizados por la transformación de coincidencia. Puede instalar los archivos de llenado en el equipo de los servicios de PowerCenter. Los archivos de llenado no se pueden editar.

Conjuntos de contenido

Archivos de datos que contienen datos de referencia especificados por una o varias transformaciones. Cuando se exporta una asignación que lee los conjuntos de contenido, el proceso de exportación exporta los datos del conjunto de contenido con los metadatos de transformación. En el caso de los

archivos del modelo clasificador y el modelo probabilístico, el proceso exporta los datos de los archivos de la estructura de directorios de Data Quality.

Requisitos de archivos de datos de referencia

Una asignación de calidad de datos pueden leer objetos de datos de referencia definidos en el repositorio de modelos.

Cuando un usuario de Informatica Data Quality exporta una asignación que lee un objeto de datos de referencia, el proceso de exportación escribe los datos de referencia en el sistema de archivos como uno o varios archivos de datos de referencia. Cuando se usa la asignación en PowerCenter, debe comprobar que los archivos están instalados en una ubicación que el servicio de integración pueda leer.

Debe especificar la ubicación de archivo de los archivos de datos de referencia al exportar la asignación. Exporte los archivos en la estructura de directorios de PowerCenter o exporte los archivos en el sistema de archivos y copie los archivos en el equipo del servicio de integración de PowerCenter.

Las ubicaciones de los archivos de datos de referencia en la estructura de directorios de PowerCenter deben corresponder a las ubicaciones de la tabla de referencia en la estructura del repositorio de modelos. La estructura de directorios debe seguir este formato:

```
<Directorio_Raíz_PowerCenter>\services\<Nombre_Proyecto_Repositorio_Modelos>
\<Nombre_Carpeta_Proyecto_Repositorio_Modelos>
```

Nota: PowerCenter lee archivos de datos de referencia en formato UTF-8. La herramienta Developer exporta los datos de referencia a PowerCenter en formato UTF-8.

Cómo instalar datos de referencia desde el programa de instalación de contenido de Data Quality

El conjunto de archivos del programa de instalación de contenido de Data Quality incluye los archivos de datos de referencia que escribe en la estructura de directorios de PowerCenter.

Al ejecutar el programa de instalación de contenido, debe identificar el directorio de instalación de PowerCenter. De forma predeterminada, el programa de instalación de contenido añade la ruta de acceso siguiente a la raíz:

```
\services\DQContent\INFA_Content\
```

Edite la ruta de acceso para que coincida con la estructura del repositorio de modelos o copie los archivos de datos de referencia en la ruta de acceso requerida después de ejecutar el programa de instalación de contenido.

Cómo usar una variable de entorno para definir la ubicación de archivos de datos de referencia

Puede definir una variable de entorno *INFA_CONTENT* en el equipo del servicio de integración de PowerCenter para establecer la ruta de acceso que el servicio de integración usa para leer los archivos de datos de referencia.

Use esta variable de entorno si no puede instalar los archivos de datos de referencia en la ubicación requerida en el equipo del servicio de integración de PowerCenter. Instale los archivos en una ubicación a la que el servicio de integración de PowerCenter pueda acceder y cree la variable de entorno *INFA_CONTENT* con la ruta de acceso ya instalada.

Requisitos de los datos de referencia de direcciones

Una asignación que realiza la validación de direcciones lee datos de referencia de direcciones. Puede adquirir y descargar los datos de referencia de direcciones de Informatica. Copie los archivos de datos en el equipo del servicio de integración de PowerCenter.

La siguiente tabla describe los tipos de datos de referencia de direcciones que puede instalar:

Tipo de datos de referencia	Uso
Lote	Operaciones generales de validación de direcciones.
CAMEO	Se usa para añadir datos demográficos de clientes a los registros de direcciones.
Certificado	Se usa para comprobar que los registros de direcciones cumplen los estándares de certificación que define un proveedor de servicios de correo. Los siguientes países definen estándares de certificación: <ul style="list-style-type: none">- 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).
Lista de sugerencias	Se usa para encontrar versiones válidas alternativas de un registro de dirección incompleto. Use los datos de la lista de sugerencias cuando configure una asignación de validación de direcciones para procesar registros de direcciones únicos en tiempo real.
Geocodificación	Se usa para añadir coordenadas de latitud y de longitud a los registros de direcciones.
Suplementario	Se usa para identificar el área geográfica o el área de población que contiene una dirección. La transformación del validador de direcciones puede añadir datos suplementarios a los registros de direcciones de los siguientes países: <ul style="list-style-type: none">- Japón. Identifica el distrito urbano que contiene la dirección.- Serbia. Identifica el código postal a nivel de calle para la dirección.- Reino Unido. Añade un identificador de punto de entrega para la dirección.- Estados Unidos. Identifica el centro de la población que contiene la dirección.

También puede usar el programa de instalación de contenido de Data Quality para instalar los datos de referencia de direcciones. Ejecute el programa de instalación de contenido del servidor en el equipo del servicio de integración de PowerCenter para instalar los archivos.

El programa de instalación de contenido de Data Quality crea un directorio `/av/` en la estructura de directorios de PowerCenter y escribe los datos de referencia de direcciones en ese directorio. El servicio de integración de PowerCenter lee la ruta de acceso a los datos de referencia de direcciones en el archivo de configuración `AD50.cfg`.

El programa de instalación de PowerCenter escribe el archivo `AD50.cfg` en la ruta de acceso siguiente:

```
[instalación_de_PowerCenter]/server/bin
```

Cuando instale datos de referencia de direcciones, compruebe que las siguientes propiedades estén configuradas en AD50.cfg:

Ubicación de datos de referencia

Introduzca la ruta en los datos de referencia de direcciones en la propiedad `ReferenceDataLocation`.

Datos de claves de licencia

Introduzca los datos de la licencia en la propiedad `LicenseKey`. Si tiene más de una licencia, introduzca cada licencia en una cadena de caracteres separados por comas.

Valores de carga previa

Introduzca al menos la abreviatura de un país como valor de carga previa para cada tipo de datos de referencia que lee una asignación. Introduzca ALL para aplicar una configuración de carga previa para todos los países.

El servicio de integración puede utilizar un método diferente para cargar datos en cada país. Por ejemplo, puede especificar una carga previa completa para datos de lotes/interactivos de Estados Unidos y una carga previa parcial para los datos de lotes/interactivos de Reino Unido. El servicio de integración también puede utilizar un método de carga previa diferente para cada tipo de datos. Por ejemplo, puede especificar una carga previa completa para los datos de lotes/interactivos de Estados Unidos y una carga previa parcial para los datos de geocodificación de Estados Unidos.

La configuración de carga previa completa sustituye a la configuración de carga previa parcial y esta última sustituye a la configuración que indica que no hay carga previa de datos. Por ejemplo, si introduce ALL para que no haya carga previa de datos e introduce USA para la carga previa completa, el servicio de integración carga todos los datos de Estados Unidos que hay en la memoria y no carga datos de ningún otro país. Si no tiene un requisito de carga previa, introduzca ALL para que no haya carga previa de datos para ningún tipo de datos de referencia que vaya a utilizar.

No especifique un valor de carga previa para los datos suplementarios.

De forma opcional, puede editar otras propiedades del archivo AD50.cfg.

Nota: Informatica Data Quality 9.x no lee el archivo AD50.cfg. Los usuarios de Informatica Data Quality 9.x introducen la ubicación de los datos de referencia, la información de la clave de licencia y los valores de la carga previa mediante la herramienta Administrator.

Requisitos de los datos de llenado de identidad

Una asignación que realiza el análisis de coincidencia de identidad lee los archivos de datos de referencia llamados poblaciones de identidad. Puede adquirir y descargar los archivos de poblaciones de identidad de Informatica. Copie los archivos de población en el equipo del servicio de integración de PowerCenter o en un directorio compartido de un equipo al que pueda acceder el servicio de integración de PowerCenter.

También puede usar el programa de instalación de contenido de Data Quality para instalar los archivos de población. Ejecute el programa de instalación de contenido del servidor en el equipo del servicio de integración de PowerCenter para instalar los archivos.

Los archivos de población se deben copiar o instalar en un directorio con el nombre `predeterminado`.

El servicio de integración de PowerCenter puede leer la ubicación del directorio de archivos de población desde el archivo de configuración IDQTx.cfg o desde la variable de entorno SSAPR. El servicio de integración de PowerCenter busca la ubicación en el archivo IDQTx.cfg. Si el archivo IDQTx.cfg no está presente o no especifica una ubicación, el servicio de integración de PowerCenter buscará la ubicación en la variable de entorno SSAPR.

Nota: El archivo IDQTx.cfg y la variable de entorno SSAPR especifican la ruta de acceso al directorio principal del directorio *predeterminado*. La ruta de acceso no incluye el nombre del directorio.

El programa de instalación de PowerCenter escribe el archivo IDQTx.cfg en la ruta de acceso siguiente:

```
[instalación_de_PowerCenter]/server/bin
```

El archivo IDQTx.cfg usa la propiedad *IdentityReferenceDataLocation* para especificar la ubicación del directorio de archivos de población. Actualice la propiedad con la ubicación actual del directorio.

Por ejemplo, puede establecer la siguiente ruta en la propiedad *IdentityReferenceDataLocation*:

```
[PowerCenter_Installation]/services/DQContent/INFA_Content/identity/
```

Puede establecer una ruta de acceso absoluta o relativa al directorio predeterminado en la propiedad *IdentityReferenceDataLocation*. Establezca una ruta de acceso absoluta a un directorio en el equipo del servicio de integración de PowerCenter o a un directorio compartido en un equipo al que pueda acceder el servicio. Establezca una ruta de acceso relativa a un directorio en el equipo del servicio de integración de PowerCenter.

El servicio de integración de PowerCenter lee una ruta de acceso relativa desde el siguiente directorio:

```
[instalación_de_PowerCenter]/server/bin
```

Por ejemplo, para especificar una ruta de acceso relativa a un directorio de archivos de población en el directorio *bin*, establezca la siguiente ruta de acceso relativa:

```
./
```

Propiedades de IDQTx.cfg

Cuando defina la propiedad *IdentityReferenceDataLocation* en el archivo IDQTx.cfg, también puede establecer las propiedades siguientes:

IdentityCacheDir

La ruta de acceso al directorio que almacena los archivos de datos temporales creados cuando se ejecuta la asignación. La ruta de acceso identifica un directorio principal. El servicio de integración de PowerCenter escribe los archivos temporales en los directorios de la ubicación que ha especificado.

La propiedad tiene el siguiente valor predeterminado:

```
./identityCache
```

IdentityIndexDir

La ruta de acceso al directorio que contiene los archivos de índice temporales creados cuando se ejecuta la asignación. El análisis de coincidencia de identidades utiliza el índice para ordenar los registros en grupos antes de analizar las coincidencias. La ruta de acceso identifica un directorio principal. El servicio de integración de PowerCenter escribe los archivos de índice en los directorios de la ubicación que ha especificado.

La propiedad tiene el siguiente valor predeterminado:

```
./identityIndex
```

Nota: Un usuario de la herramienta del desarrollador puede establecer las rutas de acceso de los directorios de memoria caché y de índice en la transformación de coincidencia. El servicio de integración de PowerCenter utiliza la ruta de acceso del archivo IDQTx.cfg si la transformación de coincidencia no especifica la ruta de acceso.

Requisitos del modelo probabilístico y del modelo clasificador

Los modelos probabilístico y clasificador son archivos de datos de referencia que identifican los tipos de información de una cadena de datos. Antes de ejecutar una asignación que lea un archivo del modelo probabilístico o clasificador, configure PowerCenter para que lea los archivos.

Realice las siguientes tareas de configuración:

- Compruebe las ubicaciones del archivo de modelo en la estructura de directorios de PowerCenter.
- Actualice la asignación de memoria de Java en el servicio de integración de PowerCenter.

Ubicaciones del modelo probabilístico y del modelo clasificador

Cuando se exporta una asignación que lee datos del modelo probabilístico o clasificador, debe especificar una ubicación de directorio para el archivo de datos del modelo exportado. Exporte el archivo a la estructura de directorios de PowerCenter o exporte el archivo al sistema de archivos y cópielo en el equipo de servicios de PowerCenter.

De forma predeterminada, el servicio de integración de PowerCenter lee los archivos de modelos del directorio siguiente:

```
[instalación_de_PowerCenter]/server/bin
```

Si los archivos no están presentes en la ubicación predeterminada, el servicio de integración de PowerCenter lee las ubicaciones de archivo de los archivos de propiedad. Actualice los archivos de propiedad con las rutas de acceso a los archivos de modelo.

La siguiente tabla describe los archivos de propiedad:

Nombre de archivo	Descripción
CLASSIFIER.properties	Contiene la propiedad <i>ClassifierFileLocation</i> . Actualice la propiedad para identificar la ruta de acceso a los archivos del modelo clasificador.
NER.properties	Contiene la propiedad <i>NERFileLocation</i> . Actualice la propiedad para identificar la ruta de acceso a los archivos del modelo probabilístico.

El servicio de integración de PowerCenter lee los archivos de propiedad del directorio siguiente:

```
[instalación_de_PowerCenter]/server/bin
```

Puede establecer una ruta de acceso absoluta o relativa para identificar las ubicaciones de los archivos del modelo en cada archivo de propiedad. Si establece una ruta de acceso relativa, el servicio de integración de PowerCenter añade la ruta de acceso al directorio siguiente:

```
[instalación_de_PowerCenter]/server/bin
```

Nota: Use barras diagonales para definir las rutas de acceso del directorio para los archivos de modelo en todas las plataformas.

Asignación de memoria del servicio de integración de PowerCenter

Un servicio de integración de PowerCenter utiliza grandes cantidades de memoria para leer un archivo del modelo clasificador o del modelo probabilístico. Si el servicio de integración de PowerCenter no puede acceder a la memoria necesaria, no podrá leer el archivo del modelo ni procesar los datos.

Revise la asignación de memoria en el servicio de integración de PowerCenter. Si ejecuta sesiones que hagan referencia a un archivo del modelo clasificador o a un archivo del modelo probabilístico, establezca la asignación de memoria mínima en 512 MB.

Puede usar la propiedad del proceso de la *memoria máxima del SDK de Java* para asignar la memoria. Para revisar o actualizar la propiedad, inicie sesión en Administrator Tool y seleccione el servicio de integración de PowerCenter que ejecuta las tareas de sesión. Puede encontrar la propiedad de la *memoria máxima del SDK de Java* en la vista **Procesos**.

INDICE

|
Informatica Data Quality [6](#)

Integración de Data Quality [6](#)