



Informatica®  
10.5.6

# Guía de datos de referencia

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.

Informatica y el logotipo de Informatica son marcas comerciales o marcas comerciales registradas de Informatica LLC en Estados Unidos y en las diversas 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.

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.

Algunas 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>.

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](mailto: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: 2024-09-27

# Tabla de contenido

<b>Prefacio .....</b>	<b>7</b>
Recursos de Informatica .....	7
Informatica Network. ....	7
Base de conocimiento de Informatica. ....	7
Documentación de Informatica. ....	7
Matrices de disponibilidad de producto de Informatica. ....	8
Informatica Velocity. ....	8
Catálogo de soluciones de Informatica. ....	8
Servicio internacional de atención al cliente de Informatica. ....	8
 <b>Capítulo 1: Introducción a los datos de referencia.....</b>	 <b>9</b>
Descripción de los datos de referencia. ....	9
Datos de referencia de Informatica. ....	10
Datos de referencia definidos por el usuario. ....	10
Tablas de referencia. ....	11
Estructura de la tabla de referencia. ....	12
Privilegios del almacén de datos de referencia. ....	12
Parámetros y tablas de referencia. ....	13
Objetos de datos de referencia y control de versiones. ....	13
 <b>Capítulo 2: Tablas de referencia en la herramienta Analyst.....</b>	 <b>14</b>
Resumen de las tablas de referencia de la herramienta Analyst. ....	14
Propiedades de la tabla de referencia. ....	15
Propiedades generales de la tabla de referencia. ....	15
Propiedades de columna de la tabla de referencia. ....	16
Crear una tabla de referencia en el editor de tablas de referencia. ....	16
Cómo crear una tabla de referencia a partir de datos de perfil. ....	17
Crear una tabla de referencia a partir de datos de columnas de perfil. ....	17
Crear una tabla de referencia a partir de patrones de valor. ....	18
Cómo crear una tabla de referencia a partir de un archivo sin formato. ....	20
Propiedades de archivo sin formato de Analyst Tool. ....	20
Crear una tabla de referencia a partir de un archivo sin formato. ....	20
Crear una tabla de referencia a partir de una tabla de base de datos. ....	21
Crear una tabla de referencia a partir de una tabla de base de datos. ....	21
Trabajar con tablas de referencia en un repositorio de modelos con versión. ....	22
Actualizaciones de la tabla de referencia. ....	23
Administrar columnas. ....	23
Administrar filas. ....	24
Buscar y reemplazar valores. ....	25
Exportar datos de tablas de referencia. ....	25

Habilitar y deshabilitar ediciones en una tabla de referencia sin administrar. . . . .	26
Actualizar los valores de tabla de referencia. . . . .	27
Eventos de traza de auditoría. . . . .	27
Visualizar eventos de traza de auditoría. . . . .	27
Reglas y directrices para tablas de referencia. . . . .	28

### **Capítulo 3: Datos de referencia en la herramienta del desarrollador. . . . . 29**

Resumen de los datos de referencia de la herramienta del desarrollador. . . . .	29
Datos de referencia y transformaciones. . . . .	30
Trabajar con objetos de datos de referencia en un repositorio de modelos con versión. . . . .	30
Desproteger objetos de datos de referencia. . . . .	30
Proteger objetos de datos de referencia. . . . .	31
Tablas de referencia. . . . .	31
Propiedades de datos de tabla de referencia. . . . .	32
Crear un objeto de tabla de referencia. . . . .	32
Cómo crear una tabla de referencia a partir de un archivo sin formato. . . . .	33
Crear una tabla de referencia a partir de un origen relacional. . . . .	34
Conjuntos de contenido. . . . .	36
Conjuntos de caracteres. . . . .	36
Modelos de clasificador. . . . .	37
Conjuntos de patrones. . . . .	37
Modelos probabilísticos. . . . .	37
Expresiones regulares. . . . .	38
Conjuntos de tokens. . . . .	39
Reglas y directrices de los modelos probabilísticos y de clasificador. . . . .	40
Crear un conjunto de contenido. . . . .	42
Cómo crear un objeto de datos de referencia en un conjunto de contenido. . . . .	42

### **Capítulo 4: Modelos de clasificador. . . . . 43**

Introducción a los modelos de clasificador. . . . .	43
Estructura del modelo de clasificador. . . . .	44
Puntuaciones de clasificador. . . . .	44
Ejemplo de transformación de clasificador. . . . .	44
Opciones del modelo de clasificador. . . . .	45
Datos de referencia del modelo de clasificador. . . . .	47
Datos de etiqueta del modelo de clasificador. . . . .	47
Administración de etiquetas del modelo de clasificador. . . . .	48
Configuración del modelo de clasificador. . . . .	49
Crear un modelo de clasificador. . . . .	49
Añadir datos de un origen de datos a un modelo de clasificador. . . . .	50
Agregar una fila de datos de referencia a un modelo de clasificador. . . . .	51
Cómo añadir una etiqueta a un modelo de clasificador. . . . .	51
Asignar una etiqueta a filas de datos de referencia. . . . .	51

Identificar valores de etiqueta no utilizados. . . . .	52
Eliminar las filas de un modelo de clasificador. . . . .	52
Cómo eliminar una etiqueta de un modelo de clasificador. . . . .	52
Cómo compilar un modelo de clasificador. . . . .	52
Operaciones de filtro y operaciones de búsqueda. . . . .	53
Usar un valor de datos para filtrar las filas de datos de referencia. . . . .	53
Usar un valor de etiqueta para filtrar las filas de datos de referencia. . . . .	53
Buscar un valor en una fila de datos de referencia. . . . .	53
Operaciones de copiar y pegar. . . . .	54
Cómo copiar un modelo de clasificador en otro conjunto de contenido. . . . .	54
Cómo importar un modelo de clasificador desde otro conjunto de contenido. . . . .	54

## **Capítulo 5: Modelos probabilísticos. . . . . 55**

Introducción a los modelos probabilísticos. . . . .	55
Estructura del modelo probabilístico. . . . .	56
Ejemplo de transformación de etiquetador. . . . .	57
Ejemplo de transformación de analizador. . . . .	58
Opciones del modelo probabilístico. . . . .	58
Vista Datos del modelo probabilístico. . . . .	58
Vista Etiqueta del modelo probabilístico. . . . .	60
Datos de referencia del modelo probabilístico. . . . .	61
Datos de etiqueta del modelo probabilístico. . . . .	61
Etiqueta de desbordamiento. . . . .	62
Propiedades del modelo probabilístico. . . . .	63
Configuración del modelo probabilístico. . . . .	63
Crear un modelo probabilístico vacío. . . . .	64
Cómo crear un modelo probabilístico a partir de un objeto de datos. . . . .	64
Añadir datos de un origen de datos a un modelo probabilístico. . . . .	65
Agregar una fila de datos de referencia a un modelo probabilístico. . . . .	66
Cómo añadir una etiqueta a un modelo probabilístico. . . . .	66
Asignar una etiqueta a un valor de datos de referencia. . . . .	66
Asignar una etiqueta a varios valores de datos. . . . .	67
Eliminar las filas de un modelo probabilístico. . . . .	68
Eliminar una etiqueta de un modelo probabilístico. . . . .	68
Cómo compilar el modelo probabilístico. . . . .	68
Buscar filas de datos en un modelo probabilístico. . . . .	68
Filtrar valores de datos de referencia por asignación de etiqueta. . . . .	69
Buscar valores de etiqueta no utilizados. . . . .	69
Operaciones de copiar y pegar. . . . .	69
Cómo copiar un modelo probabilístico en otro conjunto de contenido. . . . .	70
Cómo importar un modelo probabilístico desde otro conjunto de contenido. . . . .	70
Copiar las filas de datos de referencia en el portapapeles. . . . .	70

**Apéndice A: Datos de referencia e Informatica Data Engineering Integration. 71**

Datos de referencia e Informatica Data Engineering Integration. . . . . 71

    Datos de referencia para validación de direcciones. . . . . 71

    Instalación de los archivos de datos de referencia de direcciones. . . . . 72

**Índice. . . . . 73**

# Prefacio

Lea la *Guía de datos de referencia de Informatica* para obtener información sobre los archivos y objetos de datos de referencia que se pueden utilizar en Informatica Developer y en Informatica Analyst. Los datos de referencia sirven para identificar, analizar, mejorar o quitar valores de los datos de origen según las necesidades empresariales particulares.

## 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](mailto: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](mailto: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](mailto: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 los datos de referencia

Este capítulo incluye los siguientes temas:

- [Descripción de los datos de referencia, 9](#)
- [Datos de referencia de Informatica, 10](#)
- [Datos de referencia definidos por el usuario, 10](#)
- [Tablas de referencia, 11](#)
- [Objetos de datos de referencia y control de versiones, 13](#)

## Descripción de los datos de referencia

Las transformaciones de Informatica pueden utilizar datos de referencia para analizar y actualizar datos. Puede crear objetos de datos de referencia en la Developer Tool y la herramienta Analyst. También puede importar archivos y objetos de datos de referencia al Repositorio de modelos y al sistema de archivos.

Puede crear y editar los siguientes tipos de datos de referencia:

### **Tablas de referencia**

Una tabla de referencia contiene las versiones estándar y alternativa de un conjunto de valores de datos. Añada una tabla de referencia a una transformación en la Herramienta del desarrollador para verificar que los valores de datos de origen son exactos y tienen el formato correcto.

La mayoría de las tablas de referencia contienen al menos dos columnas. Una columna contiene la versión estándar o preferida de un valor y las otras columnas contienen versiones alternativas. 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.

### **Conjuntos de contenido**

Un conjunto de contenido es un objeto del repositorio de modelos que especifica los valores de datos de referencia en el repositorio o en un archivo. Al añadir un conjunto de contenido a una transformación, la transformación busca los datos de entrada para los valores que coinciden con los patrones de datos en el conjunto de contenido.

Se pueden descargar e instalar los siguientes tipos de datos de referencia de Informatica:

#### **Archivos de datos de referencia de direcciones**

Archivos de datos de referencia que contienen datos para las direcciones válidas en un país. La transformación del validador de direcciones lee datos de referencia. No se pueden crear o editar archivos de datos de referencia de direcciones.

Los datos de referencia de direcciones son válidos durante un período definido, por lo que deberá actualizarlos regularmente; por ejemplo, cada trimestre.

#### **Archivos de poblaciones de identidad**

Archivos de datos de referencia que contienen información sobre identidades personales, de domicilios y de empresas. La transformación de coincidencia y la transformación de comparación utilizan archivos de llenado para encontrar identidades potenciales en datos de entrada. No se pueden crear ni editar archivos de llenado de identidad.

## Datos de referencia de Informatica

Puede adquirir y descargar los datos de referencia de direcciones y datos de llenado de identidad en Informatica.

Los datos de referencia se descargan e instalan en un proceso aparte de los de las aplicaciones de Informatica. Póngase en contacto con el administrador para obtener información sobre los datos de referencia que se deben instalar en el sistema.

Los datos de direcciones se adquieren mediante suscripción en los países cuyas direcciones deberá confirmar. Puede descargar los datos de direcciones más recientes de Informatica en cualquier momento durante el período de suscripción vigente.

## Datos de referencia definidos por el usuario

Se pueden utilizar los valores de un objeto de datos para crear un objeto de datos de referencia.

Por ejemplo, se puede seleccionar un objeto de datos o una columna de perfil que contenga los valores específicos de un proyecto o de una organización. Cree objetos de datos de referencia personalizados a partir de los valores de columna.

Puede crear un objeto de datos de referencia a partir de una columna de datos para comprobar lo siguiente:

- Las filas de datos de la columna contienen el mismo tipo de información.
- Un valor de origen es válido. El objeto de referencia podría contener una lista de los valores válidos o el objeto de referencia podría contener una lista de valores que no son válidos.

La siguiente tabla muestra ejemplos comunes de columnas de datos de proyecto que puede contener datos de referencia:

Información	Ejemplo de datos de referencia
Códigos de unidades en stock (SKU)	Utilice una columna SKU para crear una tabla de referencia de códigos SKU válidos para una organización. Utilice la tabla de referencia para encontrar códigos SKU correctos o incorrectos dentro de un conjunto de datos.
Códigos de empleado	Utilice una columna de código o ID de empleado para crear una tabla de referencia de códigos de empleado válidos. Utilice la tabla de referencia para encontrar errores en los datos de los empleados.
Números de cuenta de clientes	Ejecute un perfil en una columna de cuentas de clientes para identificar los patrones de número de cuenta. Utilice el perfil para crear un conjunto de tokens de patrones de datos incorrectos. Utilice el conjunto de tokens para encontrar números de cuenta que no tengan la estructura de número de cuenta correcta.
Nombres de clientes	Cuando la columna de nombre de cliente contiene nombres, segundos nombres y apellidos, puede crear un modelo probabilístico que defina la estructura prevista de las cadenas de la columna. Utilice el modelo probabilístico para buscar cadenas de datos que no deberían estar en la columna.

## Tablas de referencia

Puede crear y actualizar tablas de referencia en la herramienta del analista y en Developer tool.

Las tablas de referencia almacenan metadatos en el repositorio de modelos. Las tablas de referencia pueden almacenar datos de columna en el almacén de datos de referencia o en otra base de datos. Cuando el almacén de datos de referencia almacena los datos de columna, los servicios de Informática identifican la tabla como una tabla de referencia administrada. Cuando otra base de datos almacena los datos de columna, los servicios de Informática identifican la tabla como una tabla de referencia sin administrar.

Puede utilizar una base de datos IBM DB2, Microsoft SQL Server, Microsoft Azure SQL, Oracle o PostgreSQL como almacén de datos de referencia.

### Opciones del almacén de datos de referencia en el servicio de administración de contenido

El servicio de administración de contenido es un servicio de aplicaciones que administra la conexión con el almacén de datos de referencia. Puede administrar y actualizar las propiedades del servicio de administración de contenido en la Herramienta del administrador.

### Reglas y directrices para la configuración del almacén de datos de referencia

Tenga en cuenta las siguientes reglas y directrices cuando configure el almacén de datos de referencia:

- Compruebe la base de datos y el esquema que el servicio de administración de contenido usará para los datos de referencia antes de crear una tabla de referencia administrada.
- De forma opcional, puede especificar un esquema para un almacén de datos de referencia de IBM DB2, Microsoft SQL Server, Microsoft Azure SQL u Oracle.

Si no se especifica un esquema de tabla de referencia en el servicio de administración de contenido, el servicio utilizará el esquema que especifica la conexión de base de datos. Si no especifica un esquema en el servicio de administración de contenido o en la conexión de base de datos, el servicio utilizará el esquema de base de datos predeterminado.

- Debe especificar el esquema en un almacén de datos de referencia de PostgreSQL.
- Utilice un controlador JDBC para conectarse a un almacén de datos de referencia de PostgreSQL. Para obtener información sobre cómo seleccionar el controlador JDBC para un almacén de datos de referencia de PostgreSQL, consulte la *Guía de instalación de PowerCenter y Data Quality*.

## Estructura de la tabla de referencia

La mayoría de las tablas de referencia contienen al menos dos columnas. Una columna contiene las versiones correctas o requeridas de los valores de datos. El resto de columnas contienen versiones diferentes de los valores, incluidas las versiones alternativas que pueden aparecer en los datos de origen.

La columna que contiene los valores correctos o necesarios se llama columna válida. Cuando una transformación lee una tabla de referencia en una asignación, la transformación busca valores en las columnas no válidas. Cuando la transformación encuentra un valor no válido, devuelve el valor correspondiente de la columna válida. También puede configurar una transformación para devolver un único valor común en lugar de los valores válidos.

La columna válida puede contener datos con formato correcto, como códigos ZIP. Puede contener datos que sean relevantes para un proyecto, tales como números de unidades en stock (SKU) que sean exclusivas para una organización. También puede crear una columna válida de datos incorrectos, tales como valores que contienen errores de datos conocidos que desea buscar.

Por ejemplo, puede crear una tabla de referencia que contenga una lista de números de SKU válidos en una organización minorista. Puede añadir la tabla de referencia a una transformación de etiquetador y crear una asignación con la transformación. Ejecute la asignación con una tabla de base de datos de productos. Cuando se ejecuta la asignación, el etiquetador crea una columna que identifica los registros del producto que no contienen números SKU válidos.

### Tablas de referencia y la transformación de analizador

Cree una tabla de referencia con una única columna para utilizar los datos de la tabla en una operación de análisis basado en patrones. Configure la transformación de analizador para realizar análisis basados en patrones e importe los datos de referencia a la configuración de la transformación.

## Privilegios del almacén de datos de referencia

El servicio de administración de contenido utiliza privilegios para restringir acciones de usuario en las tablas de referencia. Utilice las opciones de seguridad de la herramienta del administrador para revisar o actualizar los privilegios del servicio.

Para trabajar con tablas de referencia, debe tener los siguientes privilegios en el servicio de administración de contenido:

- Crear tablas de referencia
- Editar datos de tabla de referencia
- Editar metadatos de tabla de referencia

Para editar datos de una tabla de referencia sin administrar, compruebe también que se haya configurado el objeto de tabla de referencia para permitir ediciones.

**Nota:** Si edita los metadatos de una tabla de referencia sin administrar en una aplicación de base de datos, utilice la herramienta del analista para sincronizar el repositorio de modelos con la tabla. Deberá sincronizar el repositorio de modelos y la tabla de base de datos antes de usar la tabla de referencia sin administrar en Developer tool.

## Parámetros y tablas de referencia

Puede utilizar parámetros para identificar tablas de referencia en el repositorio de modelos. Puede crear un parámetro en Developer tool que identifique la tabla de referencia. También puede agregar la ubicación de la tabla de referencia a un archivo de parámetro.

Al crear un parámetro en Developer tool, puede agregarlo a una transformación en una asignación. Al agregar la ubicación de la tabla de referencia a un archivo de parámetro, especifique el archivo al ejecutar una asignación en la línea de comandos. En cada caso, el Servicio de integración de datos lee la tabla de referencia que el parámetro identifica al ejecutar la asignación.

Puede añadir un parámetro que identifica una tabla de referencia a las siguientes transformaciones:

- Transformación de conversión de mayúsculas y minúsculas
- Transformación de etiquetador
- Transformación de analizador en el modo de análisis de tokens
- Transformación de estandarizador

**Nota:** Utilice el comando `infacmd ms runMapping` para ejecutar una asignación en la línea de comandos.

## Objetos de datos de referencia y control de versiones

Si el repositorio de modelos que almacena los objetos de datos de referencia se integra con una aplicación de control de versiones, podrá aplicar el control de versión a los objetos. También podrá aplicar el control de versiones a las tablas de referencia y a los conjuntos de contenido.

Puede proteger y desproteger objetos de datos de referencia de un repositorio de modelos que admita el control de versiones. Puede deshacer una desprotección, obtener una versión anterior de un objeto y restaurar un objeto a una versión anterior. Cuando los objetos de datos de referencia no estén bajo el control de versiones, el repositorio de modelos bloqueará cualquier objeto de datos de referencia que edite. Otros usuarios no podrán editar un objeto bloqueado en el que trabaje. Cuando cierre el objeto, el repositorio de modelos liberará el bloqueo y otros usuarios podrán editar el objeto.

**Nota:** El control de versiones se aplica a los metadatos que el repositorio de modelos almacena para un objeto de tabla de referencias sin administrar. El control de versiones no se aplica a los datos de una tabla de referencias sin administrar. No podrá ver ni restaurar los datos de referencia de una versión anterior de una tabla de referencias sin administrar.

## CAPÍTULO 2

# Tablas de referencia en la herramienta Analyst

Este capítulo incluye los siguientes temas:

- [Resumen de las tablas de referencia de la herramienta Analyst, 14](#)
- [Propiedades de la tabla de referencia, 15](#)
- [Crear una tabla de referencia en el editor de tablas de referencia, 16](#)
- [Cómo crear una tabla de referencia a partir de datos de perfil, 17](#)
- [Cómo crear una tabla de referencia a partir de un archivo sin formato, 20](#)
- [Crear una tabla de referencia a partir de una tabla de base de datos, 21](#)
- [Trabajar con tablas de referencia en un repositorio de modelos con versión, 22](#)
- [Actualizaciones de la tabla de referencia, 23](#)
- [Eventos de traza de auditoría, 27](#)
- [Reglas y directrices para tablas de referencia, 28](#)

## Resumen de las tablas de referencia de la herramienta Analyst

Cree tablas de referencia en el espacio de trabajo Diseño de la Herramienta del analista.

Puede crear una tabla de referencia a partir de un archivo sin formato, de un origen de datos del repositorio de modelos y de una tabla de otra base de datos.

Puede crear una tabla de referencia desde una columna de perfil o un subconjunto de datos en una columna de perfil. También puede crear una tabla de referencia a partir de los patrones de columna que elija desde un perfil.

Al crear o actualizar una tabla de referencia, puede configurar las propiedades en la tabla y las columnas de datos que contiene.

# Propiedades de la tabla de referencia

Puede ver y actualizar las propiedades de la tabla de referencia en la Herramienta del analista. Una tabla de referencia muestra las propiedades generales y las propiedades de columna. Las propiedades generales incluyen el nombre de la tabla de referencia, la fecha de creación, el nombre de la conexión de la base de datos y el nombre de la columna válida. Las propiedades de columna incluyen los nombres de columna y los valores de precisión y de escala.

Puede ver las propiedades en modo de solo lectura. Para actualizar las propiedades, edite o desproteja la tabla de referencia.

## Propiedades generales de la tabla de referencia

Las propiedades generales contienen información acerca del objeto de tabla de referencia.

En la tabla siguiente, se describen las propiedades generales:

Propiedad	Descripción
Nombre	El nombre de la tabla de referencia.
Descripción	La descripción que un usuario ha introducido para la tabla de referencia.
Ubicación	La ubicación del objeto de tabla de referencia en el repositorio de modelos.
Columna válida	El nombre de la columna válida de la tabla de referencia.
Creada el	Fecha y hora de creación del nombre de la tabla de referencia.
Creada por	Nombre de inicio de sesión del usuario que creó la tabla de referencia.
Última modificación	Fecha y hora de la actualización más reciente de la tabla de referencia.
Última modificación por	Nombre de inicio de sesión del usuario que ha realizado la actualización más reciente.
Nombre de la conexión	Nombre de la conexión de la base de datos que almacena los valores de datos de referencia.
Tipo	Tipo de la tabla de referencia. La tabla de referencia puede ser administrada o no administrada.

## Propiedades de columna de la tabla de referencia

Las propiedades de columna incluyen información sobre los metadatos de la columna.

La tabla siguiente describe las propiedades de la columna:

Propiedad	Descripción
Nombre	El nombre de la columna.
Tipo de datos	<p>El tipo de datos para los datos de cada columna. Puede seleccionar uno de los siguientes tipos de datos:</p> <ul style="list-style-type: none"><li>- bigint</li><li>- fecha y hora</li><li>- decimal</li><li>- entero</li><li>- cadena</li></ul> <p><b>Nota:</b> No puede crear una tabla de referencia administrada desde una tabla de Oracle que utilice un número de tipo de datos, si el tipo de datos no especifica valores de precisión y escala.</p>
Precisión	<p>La precisión de cada columna. La precisión es el número máximo de dígitos o de caracteres que la columna puede admitir.</p> <p>Los valores de precisión que configura dependen del tipo de datos.</p>
Escala	<p>La escala de cada columna. La escala es el número máximo de dígitos que una columna puede admitir después del separador decimal. Se aplica a las columnas de decimales.</p> <p>Los valores de escala que configura dependen del tipo de datos.</p>
Descripción	Descripción opcional de cada columna.
Acepta valores NULL	Indica si la columna puede contener valores nulos.
Clave	<p>Identifica una columna de clave principal de una tabla que se importa de un origen de base de datos.</p> <p>Si actualiza una tabla de referencia no administrada, no podrá añadir valores duplicados a una columna cuyo origen sea una columna de clave principal en la base de datos de origen. Puede añadir valores duplicados a cualquier columna de una tabla de referencia administrada.</p>

## Crear una tabla de referencia en el editor de tablas de referencia

Defina la estructura de la tabla y añada datos a una tabla de referencia en el editor de tablas de referencia.

1. Haga clic en **Nueva > Tabla de referencia**.  
Se abre el asistente **Nueva tabla de referencia**.
2. Seleccione la opción **Usar el editor de tablas de referencia** y haga clic en **Siguiente**.
3. Utilice la opción **Añadir nueva columna** para agregar columnas a la tabla.
4. Configure las propiedades de cada columna.  
Las propiedades incluyen el nombre de columna, el tipo de datos, la precisión y la escala.



Si la columna contiene datos que una transformación puede devolver en una búsqueda de datos de referencia, seleccione la opción **Válido**.

5. Si lo desea, agregue una columna para incluir descripciones de bajo nivel como metadatos en la tabla de referencia.
6. También puede introducir una nota de auditoría para la tabla.  
La nota de auditoría aparece en el registro de traza de auditoría.
7. Haga clic en **Siguiente**.
8. Introduzca un nombre para la tabla de referencia y seleccione una ubicación para el objeto de tabla de referencia en el repositorio de modelos.
9. Haga clic en **Finalizar**.

## Cómo crear una tabla de referencia a partir de datos de perfil

Puede utilizar datos de perfil para crear tablas de referencia que se relacionen con los datos de origen del perfil. Utilice las tablas de referencia para buscar diferentes tipos de información en los datos de origen.

Puede utilizar un perfil para crear o actualizar una tabla de referencia de las siguientes formas:

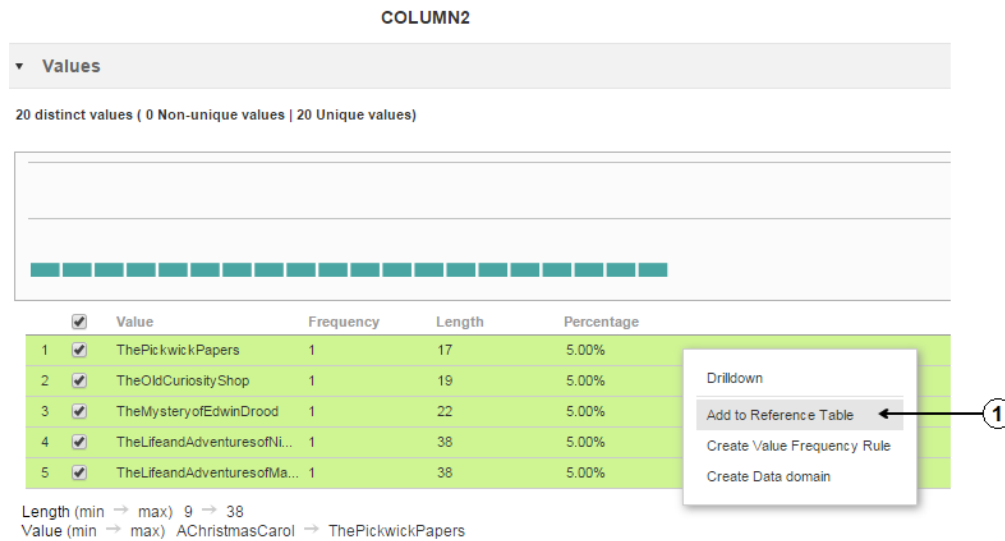
- Seleccione una columna del perfil y añádala a la tabla de referencia.
- Busque una columna del perfil y añada un subconjunto de los datos de la columna a la tabla de referencia.
- Seleccione una columna en el perfil y añada los valores de patrón de esa columna a una tabla de referencia.

## Crear una tabla de referencia a partir de datos de columnas de perfil

Puede crear una tabla de referencia a partir de uno o varios valores de la columna de datos de perfil. Seleccione una columna en un perfil y los valores de columna que desee añadir a la tabla de referencia.

1. Abra el espacio de trabajo **Biblioteca** en la Herramienta del analista.
2. Seleccione la categoría de activo **Perfiles**.  
La biblioteca muestra una lista de los perfiles en el repositorio de modelos.
3. Abra el perfil que contiene la columna para agregar a una tabla de referencia.  
El resumen de perfil enumera los nombres de las columnas del perfil.
4. Revise los datos de columna.  
Para ver los datos de columna, haga clic en el nombre de columna.
5. En la vista detallada de perfil, seleccione los valores de datos para agregar a la tabla de referencia. Puede seleccionar los valores de uno en uno o todos a la vez.
6. Haga clic en el nombre de columna y seleccione **Agregar a la tabla de referencia**.

La siguiente imagen muestra una columna de datos en la vista detallada de perfil:



El número 1 identifica la opción **Añadir a tabla de referencia** en la imagen.

7. Se abre el asistente **Añadir a tabla de referencia**.

Seleccione la opción **Crear una tabla de referencia**.

**Nota:** También puede seleccionar una opción para agregar los datos a una tabla de referencia actual.

8. Haga clic en **Siguiente**.

El nombre de columna aparece de forma predeterminada como el nombre de la tabla de referencia. Si lo desea, actualice el nombre.

9. También puede introducir una descripción y el valor predeterminado.

La Herramienta del analista utiliza el valor predeterminado para cualquier registro de tabla que no contenga ningún valor.

10. Haga clic en **Siguiente**.

11. Compruebe las propiedades de la columna.

Si lo desea, cree una columna para los metadatos descriptivos de bajo nivel.

12. Haga clic en **Siguiente**.

13. Revise el nombre y la descripción de la tabla de referencia.

Si lo desea, puede introducir una nota de auditoría.

14. Seleccione una ubicación del repositorio de modelos para el objeto de tabla de referencia.

15. Haga clic en **Finalizar**.

## Crear una tabla de referencia a partir de patrones de valor

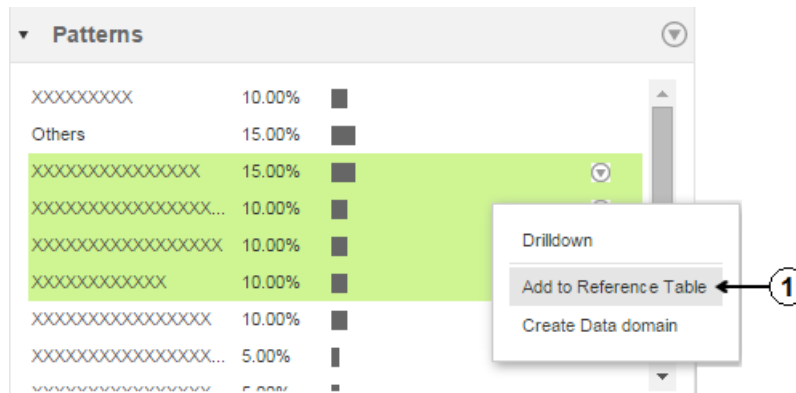
Puede crear una tabla de referencia a partir de patrones de columna en una columna de perfil. Los patrones representan la composición de los valores de datos en uno varios campos de columna. Seleccione una columna del perfil y los patrones que desee agregar a la tabla de referencia que va a crear.

1. Abra el espacio de trabajo **Biblioteca** en la Herramienta del analista.
2. Seleccione la categoría de activo **Perfiles**.

La biblioteca muestra una lista de los perfiles en el repositorio de modelos.

3. Abra el perfil que contenga los patrones de valor para agregar a la tabla de referencia.  
El resumen de perfil enumera los nombres de las columnas del perfil.
4. Seleccione la columna que defina los datos de patrón que desee agregar a la tabla de referencia.
5. Revise los patrones de datos de la columna.  
Para ver los datos de columna, haga clic en el nombre de columna.
6. En la vista detallada del perfil, seleccione los patrones de columna que desee agregar.
7. Haga clic con el botón derecho en los patrones que ha seleccionado y pulse **Añadir a tabla de referencia**.

La siguiente imagen muestra los patrones de datos de una columna en la vista detallada del perfil:



El número 1 identifica la opción **Añadir a tabla de referencia** en la imagen.

8. Se abre el asistente **Añadir a tabla de referencia**.  
Seleccione la opción **Crear una tabla de referencia**.  
**Nota:** También puede seleccionar una opción para agregar los datos a una tabla de referencia actual.
9. Haga clic en **Siguiente**.  
El nombre de columna aparece de forma predeterminada como el nombre de la tabla de referencia. Si lo desea, actualice el nombre.
10. También puede introducir una descripción y el valor predeterminado.  
La Herramienta del analista utiliza el valor predeterminado para cualquier registro de tabla que no contenga ningún valor.
11. Haga clic en **Siguiente**.
12. Compruebe las propiedades de la columna.  
Si lo desea, cree una columna para los metadatos descriptivos de bajo nivel.
13. Haga clic en **Siguiente**.
14. Revise el nombre y la descripción de la tabla de referencia.  
Si lo desea, puede introducir una nota de auditoría.
15. Seleccione una ubicación del repositorio de modelos para el objeto de tabla de referencia.
16. Haga clic en **Finalizar**.

# Cómo crear una tabla de referencia a partir de un archivo sin formato

Puede importar datos de referencia a partir de un archivo CSV. Utilice el asistente **Nueva tabla de referencia** para importar los datos del archivo.

Tiene que configurar las propiedades de cada archivo sin formato que utilice para crear una tabla de referencia.

## Propiedades de archivo sin formato de Analyst Tool

Cuando se importa un archivo sin formato como tabla de referencia, se deben configurar las propiedades de cada columna del archivo. Las opciones que se configuren determinarán la forma en que Analyst Tool leerá los datos del archivo.

En la siguiente tabla se describen las propiedades que se pueden configurar a la hora de importar datos de un archivo para una tabla de referencia:

Propiedades	Descripción
Delimitadores	Carácter utilizado para separar columnas de datos. Utilice el campo Otros para introducir un delimitador distinto. Los delimitadores deben ser caracteres imprimibles y no pueden ser el carácter de escape ni el carácter de comillas en caso de estar seleccionado. No se pueden seleccionar caracteres multibyte no imprimibles como delimitadores.
Calificador de texto	El carácter de comillas que define los límites de las cadenas de texto. Seleccione Sin comillas, Comillas simples o Comillas dobles. Si selecciona un carácter de comillas, el asistente ignora los delimitadores que aparezcan entre pares de comillas.
Nombres de columna	Importa nombres de columna de la primera línea. Seleccione esta opción si los nombres de columna aparecen en la primera fila. El asistente utiliza los datos de la primera fila en la vista previa de los nombres de columna. De forma predeterminada, no está habilitada.
Valores	Opción para iniciar la importación de valores desde una línea. Indica el número de fila en la vista previa a partir de la cual el asistente empieza a leer cuando importa el archivo.

## Crear una tabla de referencia a partir de un archivo sin formato

Cuando crea datos de tablas de referencia a partir de un archivo sin formato, la tabla utiliza la estructura de la columna del archivo e importa los datos del archivo.

1. Haga clic en **Nueva > Tabla de referencia**.  
Aparece el asistente **Nueva tabla de referencia**.
2. Seleccione la opción para **Importar un archivo sin formato**.
3. Haga clic en **Siguiente**.
4. Haga clic en **Seleccionar archivo** para seleccionar el archivo sin formato.
5. Seleccione una página de códigos que coincida con los datos del archivo sin formato.

6. Haga clic en **Cargar** para cargar los datos del archivo.
7. Haga clic en **Siguiente**.
8. Configure las propiedades del archivo sin formato.  
Las propiedades identifican el delimitador que el archivo utiliza y si la primera línea del archivo contiene nombres de columna.
9. Para obtener una vista previa de las propiedades que ha configurado, actualice el panel **Vista previa**.
10. Haga clic en **Siguiente**.
11. Configure las propiedades de cada columna.  
Las propiedades incluyen el nombre de columna, el tipo de datos, la precisión y la escala.  
Si la columna contiene datos que una transformación puede devolver en una búsqueda de datos de referencia, seleccione la opción **Válido**.
12. Si lo desea, agregue una columna para incluir descripciones de bajo nivel como metadatos en la tabla de referencia.
13. También puede introducir una nota de auditoría para la tabla.  
La nota de auditoría aparece en el registro de traza de auditoría.
14. Haga clic en **Siguiente**.
15. Introduzca un nombre para la tabla de referencia y seleccione una ubicación para el objeto de tabla de referencia en el repositorio de modelos.
16. También puede escribir si lo desea una descripción de la tabla.
17. Haga clic en **Finalizar**.

## Crear una tabla de referencia a partir de una tabla de base de datos

Al crear una tabla de referencia a partir de una tabla de base de datos, se crea un objeto de metadatos en el repositorio de modelos. Si lo desea, puede importar la tabla de datos al almacén de datos de referencia.

Al crear una tabla de referencia administrada, se importan los datos de columna al almacén de datos de referencia. Al crear una tabla de referencia no administrada, se identifica la tabla de base de datos que almacena los datos de columna.

Antes de crear la tabla de referencia, compruebe que el dominio de Informatica contiene una conexión a la base de datos que contiene los datos de referencia. Si el dominio no contiene una conexión de base de datos, puede definir una en la herramienta Analyst.

Para definir una conexión de base de datos, haga clic en **Administrar > Conexiones**.

## Crear una tabla de referencia a partir de una tabla de base de datos

Para crear la tabla de referencia, conéctese a una base de datos y seleccione la tabla que contiene los datos de referencia.

1. Seleccione **Nueva > Tabla de referencia**.  
Aparece el asistente **Nueva tabla de referencia**.

2. Seleccione la opción **Establecer conexión con una tabla relacional**.  
Para crear una tabla de referencia que no almacene datos en el almacén de datos de referencia, seleccione **Tabla no administrada**.  
Para permitir que los usuarios editen una tabla de referencia no administrada, seleccione la opción **Editable**.  
Haga clic en **Siguiente**.
3. Seleccione la conexión de base de datos de la lista de conexiones.  
Haga clic en **Siguiente**.
4. En el panel **Tablas**, seleccione una tabla.
5. Revise las propiedades de la tabla en el panel **Propiedades**.  
Opcionalmente, haga clic en **Previsualización de datos** para ver los datos de la tabla.  
Haga clic en **Siguiente**.
6. En el panel **Atributos de columna**, seleccione la columna válida.  
Si crea una tabla de referencia administrada, puede realizar las siguientes acciones en el panel **Atributos de columna**:
  - Editar los nombres de las columnas de la tabla de referencia.
  - Añadir una columna de metadatos para las descripciones en el nivel de fila.
7. Si lo desea, agregue una columna para incluir descripciones de bajo nivel como metadatos en la tabla de referencia.
8. También puede introducir una nota de auditoría para la tabla.  
La nota de auditoría aparece en el registro de traza de auditoría.
9. Haga clic en **Siguiente**.
10. Introduzca un nombre para la tabla de referencia y seleccione una ubicación para el objeto de tabla de referencia en el repositorio de modelos.
11. Si lo desea, puede escribir una descripción para la tabla de referencia.
12. Haga clic en **Finalizar**.

## Trabajar con tablas de referencia en un repositorio de modelos con versión

Abra una tabla de referencia en modo de solo lectura. Para poder trabajar en la tabla de referencia, debe entrar en el modo de edición o desproteger la tabla de referencia en el repositorio de modelos.

1. En la barra de herramientas de Informatica, haga clic en **Abrir**.  
Se abre la biblioteca de activos.
2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.  
La tabla de referencia se abre en modo de solo lectura.
3. Para editar la versión actual de la tabla de referencia, haga clic en **Editar**.  
Para editar la tabla de referencia en un repositorio de modelos con versión, desproteja la tabla de referencia.

4. Cuando termine de trabajar en la tabla de referencia, haga clic en **Finalizar**. La Herramienta del analista guardará los cambios realizados en la tabla de referencia.

Si ha desprotegido la tabla de referencia en un repositorio de modelos con versión, proteja el objeto. El repositorio de modelos con versión no actualizará la tabla de referencia hasta que no se proteja el objeto.

## Actualizaciones de la tabla de referencia

Los datos empresariales que contiene una tabla de referencia pueden cambiar a lo largo del tiempo. Revise y actualice los datos y metadatos de una tabla de referencia para comprobar que la tabla contenga información precisa. Las tablas de referencia se actualizan en la herramienta Analyst. Puede actualizar los datos y metadatos de las tablas de referencia administradas y no administradas.

Puede realizar las siguientes operaciones en los datos y metadatos de una tabla de referencia:

### Administrar columnas

Puede añadir columnas, eliminar columnas y editar propiedades de columnas.

### Administrar filas

Puede añadir filas de datos a una tabla de referencia.

### Editar valores de datos de referencia

Puede editar un valor de datos de referencia.

### Reemplazar valores de datos

Utilice la opción **Buscar y reemplazar** para reemplazar los valores de datos que ya no son exactos o relevantes para la organización. Puede buscar un valor en una columna y reemplazarlo por otro. Puede reemplazar todos los valores de una columna por un solo valor.

### Exportar una tabla de referencia

Puede exportar una tabla de referencia a un archivo de valores separados por comas (CSV), un archivo de diccionario o un archivo de Excel.

### Habilitar o deshabilitar ediciones en una tabla no administrada

Actualice una tabla de referencia no administrada para habilitar o deshabilitar las ediciones a los datos y metadatos de la tabla.

### Actualizar los datos de la tabla de referencia

Vuelva a cargar los datos de la tabla de referencia en la herramienta Analyst para ver los cambios más recientes en los datos.

## Administrar columnas

Puede agregar columnas a una tabla de referencia y actualizar las propiedades de columna. También puede actualizar el estado editable de una tabla de referencia sin administrar.

1. Haga clic en **Abrir**.  
Se abre la biblioteca de activos.
2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.  
La tabla de referencia se abre en modo de solo lectura.

3. Para editar la versión actual de la tabla de referencia, haga clic en **Editar**.  
Para editar la tabla de referencia en un repositorio de modelos con versión, desproteja la tabla de referencia.
4. Abra el menú **Acciones** y seleccione **Modificar propiedades de la columna**.  
Se abre el cuadro de diálogo **Modificar propiedades de la columna**. Use las opciones del cuadro de diálogo para realizar las siguientes operaciones:
  - Agregar una columna.
  - Cambiar la columna válida de la tabla.
  - Cambiar un nombre de columna.
  - Actualizar el texto descriptivo de una columna.
  - Actualizar el estado editable de una tabla de referencia sin administrar.
  - Actualizar la nota de auditoría de la tabla.
5. Cuando finalice las operaciones, haga clic en **Aceptar**.

## Administrar filas

Puede añadir, editar o eliminar filas en una tabla de referencia.

1. Haga clic en **Abrir**.  
Se abre la biblioteca de activos.
2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.  
La tabla de referencia se abre en modo de solo lectura.
3. Para editar la versión actual de la tabla de referencia, haga clic en **Editar**.  
Para editar la tabla de referencia en un repositorio de modelos con versión, desproteja la tabla de referencia.
4. Edite las filas de datos. Puede editar las filas de datos de las siguientes maneras:
  - Para añadir una fila, seleccione **Acciones > Añadir fila**.  
En el cuadro de diálogo **Añadir fila**, introduzca un valor en la columna válida y al menos en otra columna. Si lo desea, puede introducir una nota de auditoría.  
Haga clic en **Aceptar** para agregar la fila.
  - Para actualizar un único valor de datos, haga clic en el valor y actualice los datos.  
Después de actualizar los datos, utilice las opciones de nivel de fila para aceptar o rechazar los datos. No se puede introducir una nota de auditoría cuando se introducen datos directamente en la fila de datos.
  - Para actualizar los valores de datos en una fila, seleccione **Acciones > Editar fila**.  
En el cuadro de diálogo **Editar fila**, introduzca un valor en uno o varias columnas. Si lo desea, puede introducir una nota de auditoría.  
Haga clic en **Aplicar** para actualizar los datos en las columnas que ha seleccionado.
  - Para actualizar los valores en varias filas, seleccione las filas para editar y haga clic en **Acciones > Editar fila**.  
En el cuadro de diálogo **Editar varias filas**, introduzca un valor en una o varias columnas. Si lo desea, puede introducir una nota de auditoría.  
Haga clic en **Aceptar** para actualizar los datos en las columnas que ha seleccionado.



- Para eliminar filas, seleccione las filas para eliminar y haga clic en **Acciones > Eliminar**.

En el cuadro de diálogo **Eliminar filas**, introduzca una nota de auditoría si lo desea.

Haga clic en **Aceptar** para eliminar las filas.

**Nota:** Utilice Developer tool para editar los datos de las filas en una tabla de referencia grande. Por ejemplo, si una tabla de referencia contiene más de 500 filas, edite la tabla en Developer tool.

## Buscar y reemplazar valores

Puede buscar y reemplazar valores de datos en una tabla de referencia. Utilice las opciones de buscar y reemplazar cuando una tabla contenga una o más instancias de un valor de datos que deba actualizar.

1. Haga clic en **Abrir**.

Se abre la biblioteca de activos.

2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.

La tabla de referencia se abre en modo de solo lectura.

3. Para editar la versión actual de la tabla de referencia, haga clic en **Editar**.

Para editar la tabla de referencia en un repositorio de modelos con versión, desproteja la tabla de referencia.

4. Haga clic en **Acciones > Buscar y reemplazar**.

A continuación, aparece la barra de herramientas **Buscar y reemplazar**.

5. Especifique los criterios de búsqueda en la barra de herramientas:

- Introduzca un valor de datos en el campo **Buscar**.
- Seleccione las columnas para buscar. De forma predeterminada, la operación busca en todas las columnas.
- Introduzca un valor de datos en el campo **Reemplazar por**.

6. Utilice las siguientes opciones para reemplazar valores uno por uno o para reemplazar todos los valores:

- Utilice las opciones **Siguiente** y **Anterior** para encontrar valores uno por uno.
- Para reemplazar un valor, seleccione **Reemplazar**.
- Para mostrar todas las instancias del valor, seleccione **Resaltar todo**.
- Para reemplazar todas las instancias del valor, seleccione **Reemplazar todo**.

## Exportar datos de tablas de referencia

Exporte los datos de una tabla de referencia a un archivo separado por comas, un archivo de diccionario o un archivo de Microsoft Excel. Puede exportar los datos en modo de solo lectura.

1. Haga clic en **Abrir**.

Se abre la biblioteca de activos.

2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.

La tabla de referencia se abre en modo de solo lectura.

3. Haga clic en **Acciones > Exportar datos**.

Se abrirá el cuadro de diálogo **Exportar los datos a un archivo**.

La tabla siguiente describe las opciones del cuadro de diálogo:

Opción	Descripción
Nombre de archivo	Nombre del archivo que contendrá los datos. La operación de exportación crea el archivo.
Formato de archivo	Formato del archivo que contendrá los datos. Seleccione uno de los formatos siguientes: <ul style="list-style-type: none"><li>• csv. Archivo separado por comas. Formato predeterminado.</li><li>• xls. Archivo de Microsoft Excel.</li><li>• dic. Archivo de diccionario de Informatica.</li></ul>
Exportar nombres de campos como primera fila	Opción de nombre de columna. Seleccione la opción para indicar que la primera fila del archivo contiene los nombres de columna.
Página de códigos	Página de códigos de los datos de referencia. La página de códigos predeterminada es UTF-8.

4. Haga clic en **Aceptar** para exportar el archivo.

## Habilitar y deshabilitar ediciones en una tabla de referencia sin administrar

Puede habilitar o deshabilitar las actualizaciones de los valores de datos y las columnas en una tabla de referencia sin administrar.

Antes de cambiar el estado editable de la tabla de referencia, guarde la tabla.

1. Haga clic en **Abrir**.  
Se abre la biblioteca de activos.
2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.  
La tabla de referencia se abre en modo de solo lectura.
3. Para editar la versión actual de la tabla de referencia, haga clic en **Editar**.  
Para editar la tabla de referencia en un repositorio de modelos con versión, desproteja la tabla de referencia.
4. Abra el menú **Acciones** y seleccione **Modificar propiedades de la columna**.  
Se abre el cuadro de diálogo **Modificar propiedades de la columna**.
5. Seleccione o deseleccione la opción **Editable**.

**Nota:** Las siguientes condiciones se aplican a una tabla de referencia no administrada que permite actualizaciones del usuario:

- La tabla de referencia debe utilizar tipos de datos simples, como de cadena y de número.
- No defina ninguna restricción en los metadatos de la tabla de referencia ni especifique un valor predeterminado para ninguna columna.

## Actualizar los valores de tabla de referencia

Puede que necesite actualizar los valores que la Herramienta del analista muestra para la tabla de referencia.

Para volver a cargar los valores de la tabla de referencia, haga clic en **Acciones > Actualizar**. La Herramienta del analista obtiene de la base de datos las versiones actuales de los valores de datos.

## Eventos de traza de auditoría

Puede ver una traza de auditoría de los cambios que los usuarios han realizado en una tabla de referencia. Utilice la vista Traza de auditoría de la tabla de referencia para ver los eventos de traza de auditoría. Puede filtrar los eventos de traza de auditoría que muestra la Herramienta del analista.

La tabla siguiente describe las opciones de filtro que puede especificar:

Opción	Descripción
Fecha	Fechas de inicio y fin de las acciones para mostrar. Utilice las opciones de calendario para establecer las fechas.
Tipo	<p>El tipo del evento de traza de auditoría. Se pueden ver los siguientes tipos de eventos:</p> <ul style="list-style-type: none"><li>- Datos. Los eventos relacionados con los valores de datos en la tabla de referencia. Los eventos incluyen operaciones para añadir, eliminar y actualizar una fila.</li><li>- Metadatos. Los eventos relacionados con los metadatos de la tabla de referencia. Los eventos incluyen las operaciones para crear la tabla de referencia, agregar o quitar columnas y proteger la tabla de referencia.</li></ul> <p><b>Nota:</b> No se pueden ver al mismo tiempo eventos de datos y metadatos.</p>
Usuario	El usuario que ha editado la tabla de referencia. El filtro muestra el nombre completo y el nombre de inicio de sesión del usuario.
Estado	El estado de los eventos de registro de traza de auditoría. El estado se corresponde con la acción que se realiza en el editor de tablas de referencia. Por ejemplo, el estado puede indicar que un usuario ha creado la tabla de referencia o ha agregado una fila.

Los eventos de registro de traza de auditoría también incluyen los comentarios de traza de auditoría y los valores de columna insertados, actualizados o eliminados.

## Visualizar eventos de traza de auditoría

Consulte los eventos de traza de auditoría para conocer las actualizaciones que han realizado los usuarios en una tabla de referencias. Los eventos de traza de auditoría solo se pueden ver en modo de solo lectura.

1. Haga clic en **Abrir**.  
Se abre la biblioteca de activos.
2. Seleccione la categoría de activo **Tablas de referencia** y seleccione un nombre de tabla de referencia.  
La tabla de referencia se abre en modo de solo lectura.
3. Haga clic en **Traza de auditoría**.
4. Configure las opciones de filtro.

Puede filtrar por la fecha de la actualización, el tipo de actualización, el estado de la actualización y el nombre del usuario que realizó la actualización.

5. Haga clic en **Mostrar**.

Los eventos de registro se muestran para las opciones de filtro que haya especificado.

## Reglas y directrices para tablas de referencia

Utilice las siguientes reglas y directrices cuando trabaje con tablas de referencia en la Herramienta del analista:

- Al crear una tabla de referencia a partir de patrones de columna deducidos en un formato, la Herramienta del analista rellena la tabla de referencia con patrones de columna en un formato diferente.  
Por ejemplo, al crear una tabla de referencia para un patrón de columna X(5), la Herramienta del analista muestra el siguiente formato para el patrón de columna en la tabla de referencia: XXXXX.
- Cuando importe una tabla de base de datos de Oracle, compruebe la longitud de cualquier columna VARCHAR2 de la tabla. La Herramienta del analista no puede importar una tabla de base de datos de Oracle que contenga una columna VARCHAR2 con una longitud mayor que 1000.
- Para leer una tabla de referencia, necesita permisos de ejecución en la conexión a la base de datos que almacena los valores de datos de tabla. Por ejemplo, si el almacén de datos de referencia almacena los valores de datos, necesita permisos de ejecución en la conexión al almacén de datos de referencia. Necesita permisos de ejecución para acceder a la tabla de referencia en modo de lectura o escritura. Los permisos de la conexión de base de datos se aplican a todos los datos de referencia de la base de datos.
- Cuando ejecute una asignación con una transformación que especifica una tabla de referencia, la asignación utiliza la versión actual de la tabla de referencia en el repositorio de modelos. No puede seleccionar una versión histórica de la tabla de referencia al configurar la transformación.  
Si otro usuario restaura la tabla de referencia a una versión anterior en una sesión simultánea de Developer tool, las versiones de la tabla de referencia ya no serán idénticas entre las sesiones. Si configura y ejecuta una asignación que utiliza la tabla de referencia, la asignación podría fallar, porque la sesión actual no identifica la versión actual de la tabla de referencia. Para asegurarse de que la asignación utiliza la tabla de referencia actual, actualice el repositorio de modelos antes de ejecutar la asignación.
- Cuando configure una tabla de referencia no administrada para permitir ediciones, compruebe que la tabla de referencia utiliza tipos de datos simples, como de cadena y de número. Además, compruebe que la tabla de referencia no define ninguna restricción en los metadatos de la tabla de referencia ni emplea valores predeterminados para columnas.
- Todos los valores de microsegundos deben ser cero en los valores de datos de tabla de referencia que utiliza el tipo de dato de marca de hora.

## CAPÍTULO 3

# Datos de referencia en la herramienta del desarrollador

Este capítulo incluye los siguientes temas:

- [Resumen de los datos de referencia de la herramienta del desarrollador, 29](#)
- [Datos de referencia y transformaciones, 30](#)
- [Trabajar con objetos de datos de referencia en un repositorio de modelos con versión, 30](#)
- [Tablas de referencia, 31](#)
- [Conjuntos de contenido, 36](#)

## Resumen de los datos de referencia de la herramienta del desarrollador

Puede crear, actualizar y ver las propiedades de configuración de los objetos de datos de referencia en la Herramienta del desarrollador.

Utilice la Herramienta del desarrollador para crear y actualizar los siguientes tipos de objeto:

### **Tablas de referencia**

Una tabla de referencia contiene las versiones estándar y alternativa de un conjunto de valores de datos. Añada una tabla de referencia a una transformación en la Herramienta del desarrollador para verificar que los valores de datos de origen son exactos y tienen el formato correcto.

### **Conjuntos de contenido**

Un conjunto de contenido es un objeto del repositorio de modelos que especifica los valores de datos de referencia en el repositorio o en un archivo. Un conjunto de contenido contiene distintos tipos de datos de referencia que puede utilizar para realizar operaciones de búsqueda en las transformaciones de calidad de datos.

También puede trabajar con archivos de datos de referencia de direcciones y archivos de poblaciones de identidad en la Herramienta del desarrollador. Seleccione archivos de datos de referencia de direcciones al configurar una transformación del validador de direcciones. Seleccione archivos de poblaciones de identidad al configurar una transformación de coincidencia para el análisis de coincidencia de identidades.

# Datos de referencia y transformaciones

Varias transformaciones leen datos de referencia para realizar tareas de calidad de datos.

Las siguientes transformaciones pueden leer datos de referencia:

- Validador de direcciones. Lee datos de referencia de direcciones para verificar la exactitud de las direcciones.
- Conversión de mayúsculas y minúsculas. Lee tablas de datos de referencia para identificar las cadenas que deben cambiar mayúsculas y minúsculas.
- Clasificador. Lee los datos del conjunto de contenido para identificar el tipo de información de una cadena.
- Comparación. Lee datos de llenado de identidad durante los análisis duplicados.
- Etiquetador. Lee los datos del conjunto de contenido para identificar y etiquetar cadenas.
- Coincidencia. Lee datos de llenado de identidad durante los análisis duplicados.
- Analizador. Lee los datos del conjunto de contenido para analizar cadenas basándose en la información del contenido.
- Estandarizador. Lee tablas de datos de referencia para estandarizar cadenas con un formato común.

El conjunto de archivos del programa de instalación de contenido de Data Quality incluye objetos de datos de referencia de Informática que puede importar.

## Trabajar con objetos de datos de referencia en un repositorio de modelos con versión

Si trabaja con tablas de referencia o conjuntos de contenido en un repositorio de modelos con versión, el repositorio puede aplicar el control de versiones a los objetos. Para aplicar el control de versiones a un objeto, los usuarios deben proteger el objeto en el repositorio de modelos.

Si una tabla de referencia o un conjunto de contenido no están bajo el control de versiones, puede abrir y actualizar el objeto fuera del sistema de control de versiones. Cuando se abre el objeto, el repositorio de modelos bloquea el objeto para que otro usuario no pueda trabajar en él.

Si una tabla de referencia o un conjunto de contenido están bajo el control de versiones, abra el objeto en modo de solo lectura. Para trabajar en el objeto, desproteja el objeto en el repositorio de modelos. También puede desproteger el objeto y luego abrirlo. Proteja un objeto para crear una versión de él que incluya los últimos cambios.

### Desproteger objetos de datos de referencia

Para poder trabajar en una tabla de referencia o un conjunto de contenido que el usuario ha protegido en el repositorio de modelos, desproteja el objeto del repositorio.

1. En el Explorador de objetos, desplácese a una tabla de referencia o un conjunto de contenido.
2. Haga clic con el botón derecho en el nombre del objeto y seleccione **Abrir**.  
El objeto se abre en modo de solo lectura.
3. Haga clic con el botón derecho en el nombre del objeto y seleccione **Desproteger**.  
Ahora se puede editar el objeto.

## Proteger objetos de datos de referencia

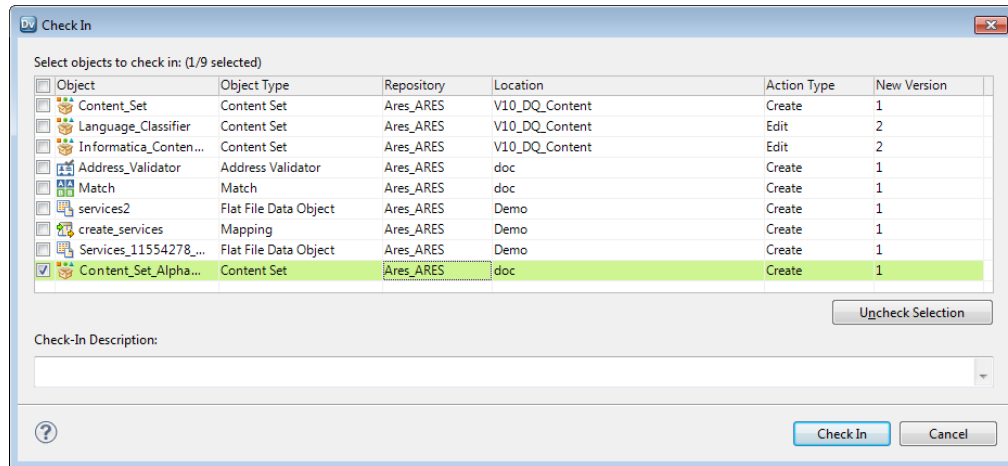
Cuando termine de trabajar en una tabla de referencia o un conjunto de contenido que ha desprotegido del repositorio de modelos, proteja el objeto.

Para ver la lista de los objetos desprotegidos en ese momento, abra la ficha **Objetos desprotegidos** que se encuentra debajo del editor de tablas de referencia.

1. Guarde todos los cambios que haya realizado en la tabla de referencia o el conjunto de contenido.
2. En el Explorador de objetos, desplácese a la tabla de referencia o al conjunto de contenido.
3. Haga clic con el botón derecho en el nombre del objeto y seleccione **Proteger**.

Se abre el cuadro de diálogo **Proteger**.

En el cuadro de diálogo se muestra la siguiente imagen:



4. Seleccione uno o varios objetos para proteger en el repositorio.

**Nota:** Puede proteger los objetos que no se hayan abierto en la sesión actual. Puede proteger cualquier objeto cuyo estado sea desprotegido.

5. Si lo desea, puede escribir una descripción de la operación.
6. Haga clic en **Proteger**.

La operación de protección actualiza el número de versión del objeto. Si es la primera vez que protege el objeto, el repositorio de modelos creará la versión uno (1) del objeto.

## Tablas de referencia

Puede añadir una tabla de referencia a una transformación en Developer tool. Configure la transformación para buscar valores de tabla de referencia en datos de entrada y para escribir los valores válidos correspondientes de la tabla de referencia como salida.

Para crear una tabla de referencia en Developer tool, utilice uno de los siguientes métodos:

- Cree una tabla de referencia vacía y especifique los valores de datos.
- Cree una tabla de referencia a partir de datos de un archivo sin formato.
- Cree una tabla de referencia a partir de los datos de una tabla de base de datos, un sinónimo o una vista.

## Propiedades de datos de tabla de referencia

Puede ver las propiedades de los datos y metadatos de la tabla de referencia y en Developer Tool. Developer Tool muestra las propiedades al abrir la tabla de referencia desde el repositorio de modelos.

Una tabla de referencia muestra las propiedades generales y las propiedades de columna. Puede ver las propiedades de la tabla de referencia en Developer Tool. Puede ver y editar las propiedades de la tabla de referencia en Analyst Tool.

La siguiente tabla describe las propiedades generales de una tabla de referencia:

Propiedad	Descripción
Nombre	El nombre de la tabla de referencia.
Descripción	La descripción opcional de la tabla de referencia.

La siguiente tabla describe las propiedades de columna de una tabla de referencia:

Propiedad	Descripción
Válido	Identifica la columna que contiene los datos de referencia válidos.
Nombre	El nombre de cada columna.
Tipo de datos	El tipo de datos de los datos de cada columna.
Precisión	La precisión de cada columna.
Escala	La escala de cada columna.
Descripción	La descripción del contenido de la columna. También puede añadir una descripción al crear la tabla de referencia.
Incluir una columna para las descripciones de nivel inferior	Indica que la tabla de referencia contiene una columna para las descripciones de los datos de columna.
Valor predeterminado	El valor predeterminado para los campos de la columna. También puede añadir un valor predeterminado al crear la tabla de referencia.
Nombre de la conexión	El nombre de la conexión con la base de datos que contiene los valores de datos de la tabla de referencia.

## Crear un objeto de tabla de referencia

Seleccione esta opción si desea crear una tabla de referencia vacía y añadir valores a mano.

1. Seleccione **Archivo > Nuevo > Tabla de referencia** en el menú de la herramienta del desarrollador.
2. En el asistente para nueva tabla, seleccione **Tabla de referencia vacía**.
3. Especifique un nombre para la tabla.
4. Seleccione un proyecto donde almacenar los metadatos de la tabla.

En el campo Ubicación, haga clic en **Examinar**. Se abre el cuadro de diálogo **Seleccionar ubicación** y se muestran los proyectos del repositorio. Seleccione el proyecto que necesite.

Haga clic en **Siguiente**.



5. Añada una o varias columnas a la tabla. Haga clic en la opción **Nuevo** para crear una columna.

En la siguiente tabla se enumeran las propiedades de columna que puede establecer:

Propiedad	Valor predeterminado
Válido	Borrado. Seleccione la columna que contiene los valores válidos.
Nombre	columna
Tipo de datos	cadena
Precisión	10
Escala	0
Descripción	Vacío. Si lo desea, puede escribir una descripción de la columna.

En la siguiente tabla se enumeran las propiedades que puede establecer para la tabla de referencia:

Propiedad	Valor predeterminado
Incluya una columna para las descripciones a nivel de fila	Borrado
Descripción	Vacío
Valor predeterminado	Vacío
Nota de auditoría	Vacío

6. Haga clic en **Finalizar**.

## Cómo crear una tabla de referencia a partir de un archivo sin formato

Puede crear una tabla de referencia a partir de los datos almacenados en un archivo sin formato.

1. Seleccione **Archivo > Nuevo > Tabla de referencia** en el menú de la herramienta Developer.
2. En el asistente para nueva tabla, seleccione **Tabla de referencia a partir de un archivo sin formato**.
3. Busque el archivo que desee utilizar como el origen de datos para la tabla.
4. Especifique un nombre para la tabla.
5. Seleccione un proyecto donde almacenar los metadatos de la tabla.

En el campo Ubicación, haga clic en **Examinar**. Se abre el cuadro de diálogo **Seleccionar ubicación** y se muestran los proyectos del repositorio. Seleccione el proyecto que necesite.

Haga clic en **Siguiente**.

6. Establezca UTF-8 como página de códigos.
7. Especifique el delimitador que utiliza el archivo sin formato.
8. Si el archivo sin formato contiene nombres de columna, seleccione la opción para importar nombres de columna desde la primera línea del archivo.

9. La tabla siguiente describe propiedades de tabla opcionales:

Propiedad	Valor predeterminado
Calificador de texto	Sin comillas
Iniciar importación en la línea	Línea 1
Delimitador de fila	\012 LF (\n)
Tratar delimitadores consecutivos como uno	Borrado
Carácter de escape	Vacío
Retener carácter de escape en datos	Borrado
Máximo de filas para vista previa	500

Haga clic en **Siguiente**.

10. Seleccione la columna que contiene los valores válidos.
11. La tabla siguiente describe propiedades opcionales:

Propiedad	Valor predeterminado
Incluya una columna para las descripciones a nivel de fila	Borrado
Nota de auditoría	Vacío
Valor predeterminado	Vacío
Máximo de filas para vista previa	500

Haga clic en **Finalizar**.

La tabla de referencia se abre en el espacio de trabajo de la herramienta Developer.

## Crear una tabla de referencia a partir de un origen relacional

Puede crear una tabla de referencia a partir de una tabla relacional, un sinónimo o una vista.

Al crear una tabla de referencia administrada, se importan los datos de columna al almacén de datos de referencia. Al crear una tabla de referencia no administrada, se identifica la tabla de base de datos que almacena los datos de columna.

Antes de crear la tabla de referencia, compruebe que el dominio de Informatica contiene una conexión a la base de datos que contiene los datos de referencia.

Puede configurar una conexión de base de datos en el explorador de conexiones. Si Developer tool no muestra el explorador de conexiones, seleccione **Ventana > Mostrar vista > Explorador de conexiones** en el menú de Developer tool.

## Crear una tabla de referencia a partir de un origen relacional

Para crear la tabla de referencia, conéctese a una base de datos y seleccione la tabla que contiene los datos de referencia.

1. Seleccione **Archivo > Nuevo > Tabla de referencia** en el menú de la Developer Tool.
2. En el asistente de creación de tablas, seleccione **Tabla de referencia desde un origen relacional**.  
Haga clic en **Siguiente**.
3. Seleccione una conexión de base de datos.  
En el campo Conexión, haga clic en **Examinar**. Se abre el cuadro de diálogo **Elegir conexión** y aparecen las conexiones de base de datos disponibles.  
Haga clic en **Aceptar** cuando seleccione una conexión.
4. Seleccione un recurso de base de datos.  
En el campo Recursos, haga clic en **Examinar**. El cuadro de diálogo **Seleccionar un recurso** se abre y muestra los recursos en la conexión de base de datos. Explore la base de datos y seleccione una tabla de base de datos, un sinónimo o una vista.  
Si lo desea, puede previsualizar la información de entidad en el recurso.
5. Especifique un nombre para la tabla.
6. Seleccione una ubicación para el objeto de tabla de referencia.  
En el campo Ubicación, haga clic en **Examinar**. Se abre el cuadro de diálogo **Seleccionar ubicación** y se muestran los proyectos del repositorio.  
Seleccione una ubicación y haga clic en **Siguiente**.
7. Para crear una tabla de referencia que no almacene datos en el almacén de datos de referencia, seleccione **Tabla no administrada**.  
Para permitir que los usuarios editen una tabla de referencia no administrada, seleccione la opción **Editable**.  
Haga clic en **Siguiente**.
8. Seleccione la columna que contiene los valores válidos.
9. En la siguiente tabla se describen propiedades opcionales que puede especificar:

Propiedad	Valor predeterminado
Incluya una columna para las descripciones a nivel de fila	Borrado
Descripción	Borrado
Valor predeterminado	Vacío
Nota de auditoría	Vacío
Máximo de filas para vista previa	500

10. Haga clic en **Finalizar**.

# Conjuntos de contenido

Un conjunto de contenido es un objeto del repositorio de modelos que almacena los datos o metadatos para otros objetos de datos de referencia. Un conjunto de contenido puede incluir conjuntos de caracteres, conjuntos de patrones, conjuntos de tokens, expresiones regulares, modelos probabilísticos y modelos de clasificador. Utilice un conjunto de contenido para definir y organizar objetos de datos de referencia relacionados con un único proyecto, tipo de información o propósito empresarial.

La Herramienta del desarrollador incluye conjuntos de caracteres y conjuntos de tokens definidos por el sistema que no aparecen en el repositorio de modelos. Para ver y utilizar los objetos definidos por el sistema, configure una estrategia en la transformación de etiquetador, de analizador o de estandarizador.

## Conjuntos de caracteres

Un juego de caracteres contiene expresiones que identifican caracteres e intervalos de caracteres específicos. Puede utilizar conjuntos de caracteres en transformaciones de etiquetador que utilizan el modo de etiquetado de caracteres.

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 conjuntos de caracteres para identificar un carácter o un rango de caracteres como parte de las operaciones 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.

## Propiedades de los juegos de caracteres

Configure las propiedades que determinan las operaciones de etiquetado de caracteres para un juego de caracteres.

En la tabla siguiente se describen las propiedades para un juego de caracteres definido por el usuario:

Propiedad	Descripción
Etiqueta	Define la etiqueta que una transformación de etiquetador aplica a los datos que coinciden con el juego de caracteres.
Modo estándar	Habilita una vista de edición simple que incluye campos para el rango inicial y el rango final.
Rango inicial	Especifica el primer carácter de un rango de caracteres.
Rango final	Especifica el último carácter de un rango de caracteres. Para un rango con un solo carácter, deje este campo vacío.
Modo avanzado	Habilita una vista de edición avanzada donde se pueden especificar manualmente rangos de caracteres mediante caracteres de rango y caracteres delimitadores.
Carácter de rango	Cambia provisionalmente el símbolo que representa un rango de caracteres. El carácter de rango vuelve a convertirse en el carácter predeterminado cuando se cierra el juego de caracteres.
Carácter delimitador	Cambia provisionalmente el símbolo que separa rangos de caracteres. El carácter delimitador vuelve a convertirse en el carácter predeterminado cuando se cierra el juego de caracteres.

## Modelos de clasificador

Un modelo de clasificador analiza cadenas de entrada y determina los tipos de información con más probabilidades de aparecer en las cadenas. Puede utilizar un modelo de clasificador en una transformación de clasificador.

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.

## 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.

### Propiedades del conjunto de patrones

Configure las propiedades necesarias para determinar los patrones de un conjunto de patrones.

En la tabla siguiente se describe la propiedad para el conjunto de patrones definido por el usuario:

Propiedad	Descripción
Patrón	Define los patrones que busca el analizador de patrones. Se pueden especificar varios patrones para un conjunto de patrones. Se pueden especificar patrones contruidos a partir de una combinación de caracteres comodín, caracteres y cadenas.

## Modelos probabilísticos

Un modelo probabilístico analiza valores de entrada y determina los tipos de información con más probabilidades de aparecer en los valores. Utilice un modelo probabilístico en una transformación de etiquetador y una transformación del analizador.

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.

## Expresiones regulares

En el contexto de los conjuntos de contenido, una expresión regular es una expresión que se puede usar en operaciones de análisis y de etiquetado. Utilice expresiones regulares para identificar una o más cadenas en los datos de entrada. Puede utilizar expresiones regulares en transformaciones de analizador que utilizan el modo de análisis de tokens. También puede utilizar expresiones regulares en transformaciones de etiquetador que utilizan el modo de etiquetado 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.

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.

## Propiedades de expresiones regulares

Configure las propiedades que determinan cómo identifica y escribe cadenas de salida una expresión regular.

En la tabla siguiente se describen las propiedades para una expresión regular definida por el usuario:

Propiedad	Descripción
Número de salidas	Define el número de puertos de salida que escribe la expresión regular.
Expresión regular	Define un patrón que utiliza la transformación de analizador para hacer coincidir cadenas.
Expresión de prueba	Contiene datos que se introducen para probar la expresión regular. Mientras se escriben datos en este campo, el campo resalta las cadenas que coinciden con la expresión regular.
Siguiente expresión	Se desplaza hasta la siguiente cadena que coincide con la expresión regular y cambia el tipo de letra de la cadena a negrita.
Expresión previa	Se desplaza hasta la siguiente cadena que coincide con la expresión regular y cambia el tipo de letra de la cadena a negrita.

## 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. También 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 específicos como parte de las operaciones de etiquetado y análisis. 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.

### Propiedades de conjuntos de tokens

Configure las propiedades necesarias para determinar las operaciones de etiquetado de un conjunto de tokens.

En la tabla siguiente se describen la propiedades para un juego de caracteres definido por el usuario:

Propiedad	Modo de conjuntos de tokens	Descripción
Nombre	N/A	Define el nombre del conjunto de tokens
Descripción	N/A	Describe el conjunto de tokens.
Opciones de conjunto de tokens	N/A	Define si el conjunto de tokens utiliza el modo de expresiones regulares
Etiqueta	Expresión regular	Define la etiqueta que una transformación de etiquetador aplica a los datos que coinciden con el conjunto de tokens.
Expresión regular	Expresión regular	Define un patrón que la transformación de etiquetador utiliza para buscar coincidencias en cadenas.
Probar expresión	Expresión regular	Contiene datos que se especifican para probar la expresión regular. A medida que se escriben datos en este campo, el campo resalta las cadenas que coinciden con la expresión regular.
Siguiente expresión	Expresión regular	Se desplaza hasta la siguiente cadena que coincide con la expresión regular y cambia el tipo de letra de esa cadena a negrita.

Propiedad	Modo de conjuntos de tokens	Descripción
Expresión previa	Expresión regular	Se desplaza hasta la cadena anterior que coincide con la expresión regular y cambia el tipo de letra de esa cadena a negrita.
Etiqueta	Carácter	Define la etiqueta que una transformación de etiquetador aplica a los datos que coinciden con el juego de caracteres.
Modo estándar	Carácter	Habilita una vista de edición simple que incluye campos para el rango inicial y el rango final.
Rango inicial	Carácter	Especifica el primer carácter de un rango de caracteres.
Rango final	Carácter	Especifica el último carácter de un rango de caracteres. En el caso de rangos de un carácter, deje este campo en blanco.
Modo avanzado	Carácter	Habilita una vista de edición avanzada donde se pueden especificar manualmente rangos de caracteres mediante caracteres de rango y caracteres delimitadores.
Carácter de rango	Carácter	Cambia temporalmente el símbolo que representa un rango de caracteres. El carácter de rango revierte al carácter predeterminado cuando se cierra el juego de caracteres.
Carácter delimitador	Carácter	Cambia temporalmente el símbolo que separa rangos de caracteres. El carácter delimitador revierte al carácter predeterminado cuando se cierra el juego de caracteres.

## Reglas y directrices de los modelos probabilísticos y de clasificador

Cada modelo probabilístico y de clasificador del repositorio de modelos identifica un archivo en la estructura del directorio de Informatica. Los archivos contienen los valores de datos y las etiquetas que se añaden al



modelo en la herramienta del desarrollador. Los archivos también contienen la lógica de compilación que define las asociaciones entre los valores de datos y las etiquetas.

Tenga en cuenta las siguientes reglas y directrices cuando trabaje con modelos probabilísticos o de clasificador:

- Cuando se ejecuta una asignación que incluye un modelo, el servicio de integración de datos aplica la lógica del modelo compilado a los datos de entrada de la transformación. El servicio de integración de datos no lee los valores de datos ni las etiquetas del modelo cuando se ejecuta la asignación.
- También puede eliminar los valores de datos y las etiquetas de un modelo probabilístico o un modelo de clasificador. Por ejemplo, puede optar por eliminar datos confidenciales o datos propietarios de un modelo. Puede eliminar valores de datos y etiquetas individuales en la herramienta del desarrollador. Puede eliminar todos los valores de datos y las etiquetas cuando exporta un modelo desde el repositorio de modelos.

**Nota:** Si elimina todos los valores de datos y las etiquetas de un modelo, no podrá compilar el modelo.

- Cuando se eliminan uno o varios valores de datos o etiquetas de un modelo, la lógica del modelo compilado ya no representa los datos actuales en el archivo de modelo. Para sincronizar la lógica del modelo y los valores de datos y las etiquetas, compile el modelo de nuevo. No compile el modelo si desea mantener la lógica del modelo actual.
- Para proteger los datos en un modelo de clasificador o un modelo probabilístico, realice una copia de seguridad del archivo del modelo en la estructura del directorio de Informática. Realice una copia de seguridad del archivo antes de eliminar todos los valores de datos y las etiquetas de un modelo.
- Busque los archivos del modelo en el equipo host del servicio de administración de contenido.

Los archivos del modelo probabilístico tienen la ubicación y la extensión de nombre de archivo predeterminadas que se especifican a continuación:

```
<Directorio_de_instalación_de_Informatica>/tomcat/bin/ner/<nombre de archivo>.ner
```

Los archivos del modelo de clasificador tienen la ubicación y la extensión del nombre de archivo predeterminadas siguientes:

```
<Directorio_de_instalación_de_Informatica>/tomcat/bin/classifier/<nombre de archivo>.classifier
```

- Si actualiza la instalación de Informática, deberá compilar los modelos probabilísticos y de clasificador antes de utilizarlos en una asignación. Si un modelo no contiene datos, reemplace el archivo actual en la estructura del directorio de Informática por el archivo de copia de seguridad que contiene los datos.

## Administrar etiquetas en modelos de clasificador y modelos probabilísticos

Para revisar y actualizar las etiquetas en un modelo probabilístico o un modelo de clasificador, utilice el cuadro de diálogo **Administrar etiquetas**.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Abra el cuadro de diálogo **Administrar etiquetas**.

El cuadro de diálogo enumera las etiquetas del modelo.

## TEMAS RELACIONADOS

- [“Administración de etiquetas del modelo de clasificador” en la página 48](#)
- [“Configuración del modelo probabilístico” en la página 63](#)

## Crear un conjunto de contenido

Cree un conjunto de contenido para administrar objetos de datos de referencia que hacen referencia a un único proyecto, tipo de información o propósito empresarial.

1. En la vista **Explorador de objetos**, seleccione un proyecto o una carpeta para almacenar el conjunto de contenido.
2. Haga clic en **Archivo > Nuevo > Conjunto de contenido**.
3. Escriba un nombre para el conjunto de contenido.
4. Si lo desea, seleccione **Explorar** para cambiar la ubicación del repositorio de modelos para el conjunto de contenido.
5. Haga clic en **Finalizar**.

## Cómo crear un objeto de datos de referencia en un conjunto de contenido

En un conjunto de contenido, puede crear un conjunto de caracteres, un conjunto de patrones, un conjunto de tokens, una expresión regular, un modelo probabilístico y un modelo de clasificador.

1. Abra un conjunto de contenido en el editor y seleccione la vista **Contenido**.
2. Seleccione un tipo de objeto de datos de referencia.
3. Haga clic en **Añadir**.
4. Especifique un nombre para el objeto de datos de referencia.  
También puede escribir una descripción del objeto.
5. Configure las propiedades del objeto de datos de referencia.
6. Haga clic en **Finalizar**.

**Sugerencia:** Puede copiar objetos de datos de referencia desde un conjunto de contenido a otro. Utilice las opciones **Copiar en** y **Pegar desde** para crear una copia de un objeto en un conjunto de contenido. Utilice la tecla **Ctrl** para seleccionar varios objetos de conjunto de contenido.

## CAPÍTULO 4

# Modelos de clasificador

Este capítulo incluye los siguientes temas:

- [Introducción a los modelos de clasificador, 43](#)
- [Estructura del modelo de clasificador, 44](#)
- [Puntuaciones de clasificador, 44](#)
- [Ejemplo de transformación de clasificador, 44](#)
- [Opciones del modelo de clasificador, 45](#)
- [Datos de referencia del modelo de clasificador, 47](#)
- [Datos de etiqueta del modelo de clasificador, 47](#)
- [Configuración del modelo de clasificador, 49](#)
- [Operaciones de filtro y operaciones de búsqueda, 53](#)
- [Operaciones de copiar y pegar, 54](#)

## Introducción a los modelos de clasificador

Un modelo de clasificador es un objeto de datos de referencia en un conjunto de contenido. Utilice un modelo de clasificador para analizar cadenas de texto largas que contienen varios valores. Un modelo de clasificador identifica el tipo de información más común en cada cadena.

Añada un modelo de clasificador a una transformación de clasificador. La transformación busca valores comunes entre los datos del modelo de clasificador y los datos en cada fila de entrada. La transformación utiliza los valores comunes para categorizar el tipo de información que representa cada fila.

Puede utilizar un modelo de clasificador cuando los datos de entrada tienen las siguientes características:

- Los datos de entrada contienen texto. Los modelos de clasificador aplican procesos de lenguajes naturales a los datos de texto para identificar los tipos de información en el texto. Los procesos de lenguajes naturales detectan palabras relevantes en la cadena de entrada. Los procesos de lenguajes naturales omiten las palabras que no son relevantes.
- Las cadenas de datos de entrada contienen varios valores. Por ejemplo, puede crear una columna de datos que contenga el contenido de un mensaje de correo electrónico en cada campo.

La transformación de clasificador lee los tipos de datos de las cadenas. La transformación no impone ningún límite a la longitud de las cadenas de entrada.

Puede compilar modelos de clasificador en la herramienta Developer. Al compilar un modelo, se crean asociaciones entre valores de datos similares en el modelo. La transformación de clasificador utiliza los datos compilados para buscar información en los datos de entrada.

## Estructura del modelo de clasificador

Un modelo de clasificador contiene valores de datos de referencia y valores de etiqueta. Los valores de datos de referencia representan los datos que desea clasificar. Los valores de etiqueta especifican los tipos de información que una transformación de clasificador puede identificar en los datos.

Los modelos de clasificador también contienen datos de compilación. La transformación de clasificador utiliza los datos de compilación para medir las similitudes entre los datos de referencia del modelo y los datos de entrada de la transformación. Al compilar un modelo de clasificador, se crean o actualizan los datos de compilación. Cuando una transformación de clasificador compara los datos de entrada con los datos del modelo, la transformación devuelve los valores de etiqueta que describen cada fila de datos de entrada.

Developer tool escribe los valores de datos de referencia, los valores de etiqueta y los datos de compilación en un archivo de la estructura del directorio de Informatica. El objeto del modelo de clasificador del repositorio de modelos almacena el nombre de archivo. Al guardar un modelo de clasificador, se escriben los valores de datos de referencia y los valores de etiqueta actuales en el archivo. Cuando se compila el modelo, se actualizan los datos de la compilación en el archivo. Puede leer el nombre de archivo desde las propiedades del modelo en Developer tool.

## Puntuaciones de clasificador

Una transformación de clasificador compara cada fila de datos de entrada con cada fila de datos de referencia en un modelo de clasificador. La transformación calcula una puntuación para cada comparación. Las puntuaciones representan los niveles de similitud entre la fila de entrada y las filas de datos de referencia.

Cuando ejecute una asignación que contenga una transformación de clasificador, la asignación devuelve la etiqueta que identifica la fila de datos de referencia con la puntuación más alta. El intervalo de la puntuación es 0 a 1. Una puntuación alta indica una coincidencia alta entre los datos de entrada y los datos del modelo.

Revise las puntuaciones del clasificador para comprobar que la salida de etiqueta describe exactamente cada fila de datos de entrada. También puede revisar las puntuaciones para comprobar que el modelo de clasificador es adecuado para los datos de entrada. Si la salida de la transformación contiene un gran porcentaje de puntuaciones bajas, el modelo de clasificador podría ser inapropiado. Para mejorar las comparaciones, compile el modelo de nuevo. Si el modelo compilado no mejora las puntuaciones, reemplácelo en la transformación.

## Ejemplo de transformación de clasificador

Puede utilizar un modelo de clasificador y una transformación de clasificador para categorizar los mensajes de correo electrónico basándose en el texto que contienen.

Por ejemplo, imagine que es gestor de datos en el centro de atención al cliente de un fabricante de software. Debe revisar los mensajes de correo electrónico que el centro de atención recibe de los clientes. La organización tiene clientes en varios países, por lo que en el centro de atención se reciben correos electrónicos en muchos idiomas. Decide ordenar los correos electrónicos por idioma, de modo que pueda enviar cada correo electrónico al departamento más indicado para responder al cliente.

Para ordenar los mensajes de correo electrónico, realice los pasos siguientes:

1. Escriba los mensajes de correo electrónico en un único archivo o una tabla de base de datos.
2. Cree un objeto de datos en el repositorio de modelos que lee el archivo o la tabla de base de datos.
3. Cree objetos de datos en el repositorio de modelos para cada idioma que los mensajes utilicen.
4. Cree un modelo de clasificador que contenga el texto de muestra para cada idioma.

**Nota:** Puede utilizar datos de muestra de los mensajes de correo electrónico como datos de origen para el modelo.

5. Agregue el modelo de clasificador a una transformación de clasificador reutilizable.
6. Configure una asignación para aplicar la transformación de clasificador a los datos del mensaje.

Para configurar la asignación, realice los pasos siguientes:

- Agregue la transformación de clasificador y los objetos de datos a la asignación.
- Conecte un puerto de entrada de la transformación de clasificador al objeto de datos de origen.
- Conecte los puertos de salida de la transformación de clasificador a los objetos de datos de destino.

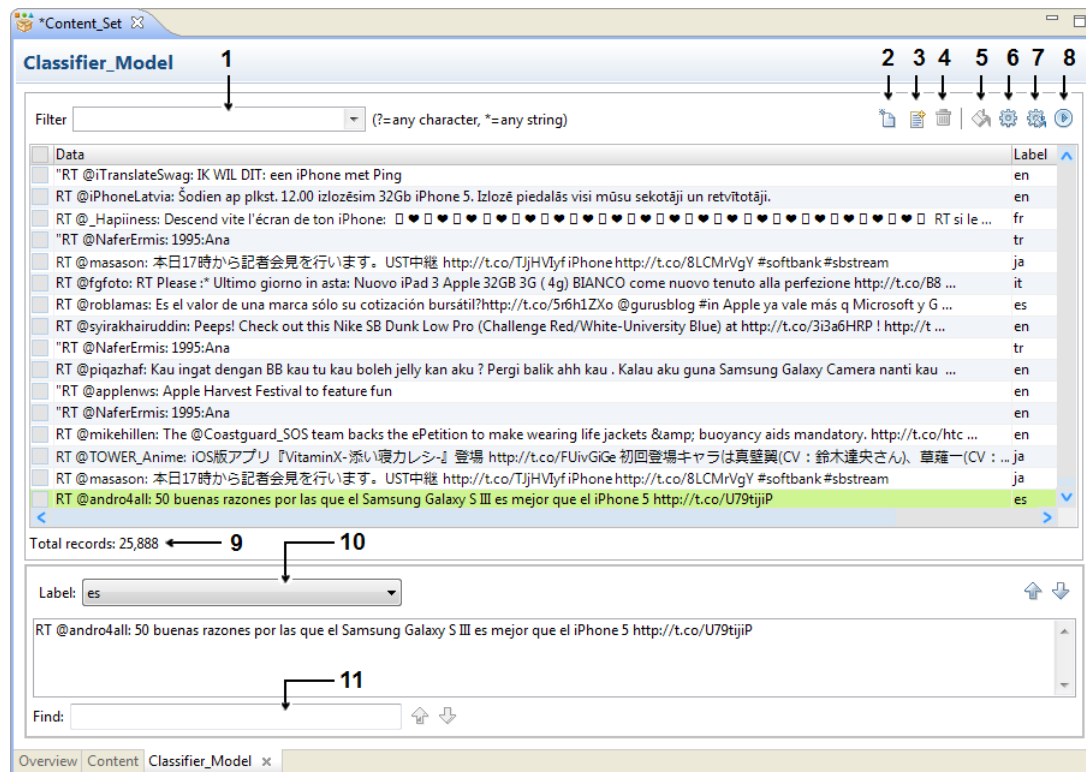
Al ejecutar la asignación, la transformación de clasificador analiza los mensajes de correo electrónico y escribe el texto de los correos electrónicos en el destino de datos correcto. Puede compartir los destinos de datos con los miembros del equipo de cada departamento.

## Opciones del modelo de clasificador

Developer tool muestra los datos del modelo de clasificador en un editor con un panel superior y otro inferior. El panel superior muestra cada fila de datos de referencia y todas las etiquetas que asigne a los datos. El panel inferior muestra el contenido de una fila que seleccione.

Utilice el panel superior para revisar las filas de datos de referencia y para identificar cualquier fila que no utilice una etiqueta. Utilice el panel inferior para revisar el contenido de una fila y para asignar una etiqueta a la fila. El panel superior muestra aproximadamente 100 caracteres de datos en cada fila. El panel inferior muestra todos los datos de la fila que seleccione.

La siguiente imagen muestra el editor del modelo de clasificador:



El editor incluye las opciones siguientes:

1. Campo Filtro  
Filtra la lista de filas de datos de referencia según el valor de los datos o la etiqueta que especifique.
2. Añadir fila  
Inserta una fila de datos de referencia en blanco.
3. Anexar datos  
Importa los datos de un objeto de datos del repositorio de modelos.
4. Eliminar  
Elimina las filas de datos de referencia que seleccione. Utilice las casillas de verificación para seleccionar las filas.
5. Asignar etiqueta  
Asigna una etiqueta a una o más filas de datos de referencia que seleccione. Utilice las casillas de verificación para seleccionar las filas.
6. Editar propiedades  
Muestra las propiedades del modelo de clasificador.
7. Administrar etiquetas  
Abre el cuadro de diálogo **Administrar etiquetas**. Utilice el cuadro de diálogo para agregar o eliminar valores de etiqueta del modelo de clasificador.
8. Compilar  
Compila el modelo de clasificador.

9. Total de registros  
Indica el número de filas de datos de referencia en el modelo de clasificador.
10. Campo Etiqueta  
Muestra un valor de etiqueta que puede aplicar a la fila de datos de referencia actual.
11. Campo Buscar  
Busca un valor de datos que especifique en la fila de datos de referencia actual.

## Datos de referencia del modelo de clasificador

Un modelo de clasificador contiene una columna de datos de referencia que puede incluir oraciones, párrafos o páginas de texto. Los datos de referencia representan los distintos tipos de entrada de texto que una transformación de clasificador puede leer en una asignación. Cuando cree un modelo, compruebe que los datos de referencia incluyen los tipos de texto que espera encontrar al ejecutar la asignación.

Puede utilizar los datos de origen de la asignación para crear un modelo de clasificador. Seleccione una muestra de los datos de origen y copie la muestra de datos en el modelo.

Tenga en cuenta las siguientes reglas y directrices cuando trabaje con datos de referencia del modelo de clasificador:

- Un campo de datos de referencia puede ser de cualquier longitud. Puede especificar páginas de texto en cada campo de datos.
- Los datos de referencia se importan desde un objeto de datos.
- Los valores de los datos de referencia no se pueden editar. Sin embargo, se puede eliminar una fila de datos.
- Cuando compile un modelo clasificador, el proceso de compilación no tendrá en cuenta los valores de número en los datos de referencia.

## Datos de etiqueta del modelo de clasificador

Un modelo de clasificador contiene una o más etiquetas descriptivas que resumen los tipos de información de las filas de datos de referencia. Asigne una etiqueta a cada fila de datos de referencia.

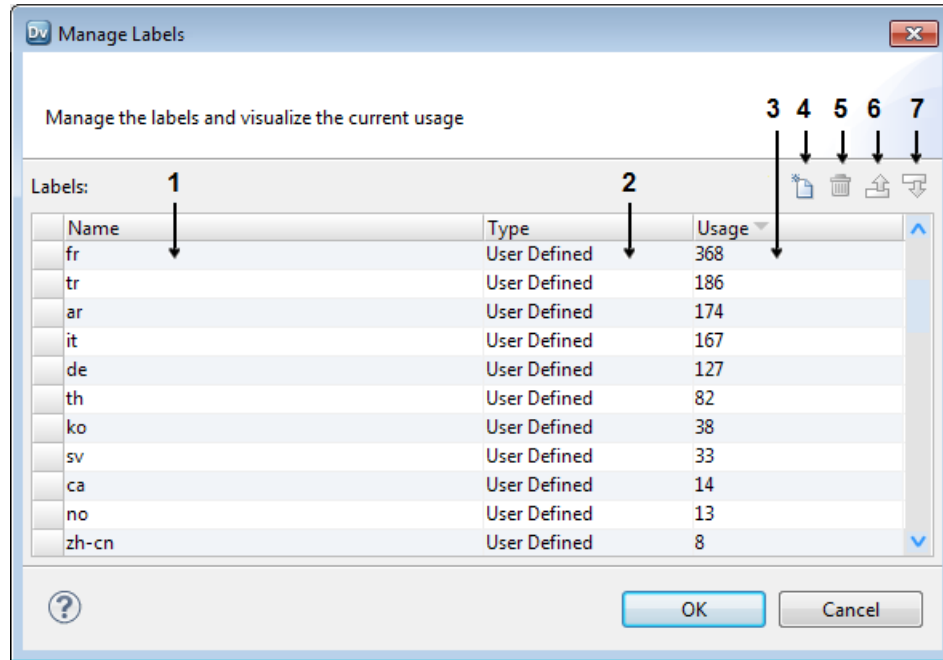
Al agregar datos de un origen de datos a un modelo de clasificador, puede especificar una columna como una columna de datos de etiqueta. También puede crear etiquetas en el modelo.

Las etiquetas son independientes de los valores de los datos de referencia que describen. Si elimina las filas de datos de referencia que utilizan una etiqueta, no se elimina la etiqueta del modelo. Si elimina una etiqueta, no eliminará los valores de datos de referencia asociados a la etiqueta.

## Administración de etiquetas del modelo de clasificador

Utilice el cuadro de diálogo **Administrar etiquetas** para revisar y actualizar los valores de etiqueta de un modelo de clasificador. También puede ordenar y actualizar los valores de etiqueta.

La siguiente imagen muestra el cuadro de diálogo **Administrar etiquetas**:



El cuadro de diálogo **Administrar etiquetas** contiene los elementos siguientes:

1. Columna Nombre.  
Contiene los valores de etiqueta que la transformación de clasificador puede aplicar a las filas de datos de entrada. Puede ordenar las etiquetas por nombre.
2. Columna Tipo.  
Identifica el origen de los valores de etiqueta. El modelo de clasificador identifica todas las etiquetas como valores definidos por el usuario.
3. Columna Uso.  
Indica el número de filas de datos de referencia que utilizan cada etiqueta. Puede ordenar las etiquetas por el número de filas.
4. Botón Añadir.  
Agrega una etiqueta al modelo de clasificador. Introduzca un valor de etiqueta en la columna Nombre de la fila.  
**Nota:** Para actualizar un valor de etiqueta, haga doble clic en el valor y escriba el valor que necesite.
5. Botón Eliminar.  
Elimina una etiqueta del modelo de clasificador.
6. Flecha arriba.  
Hace que la etiqueta suba una fila en el cuadro de diálogo.
7. Flecha abajo.  
Hace que la etiqueta baje una fila en el cuadro de diálogo.



# Configuración del modelo de clasificador

Los pasos para configurar un modelo de clasificador comienzan con los datos que desea clasificar. El contenido de los datos de referencia que añade al modelo debe reflejar los datos que conecta a la transformación de clasificador. La transformación compara los patrones y valores de datos de los datos de entrada con los patrones y valores de datos del modelo de clasificador.

Para crear un modelo de clasificador que pueda utilizar en una transformación de clasificador, realice las tareas siguientes:

1. Identifique los valores de datos de referencia y los valores de etiqueta que agregar al modelo.  
Puede utilizar un fragmento de los datos que desea clasificar. Cree un objeto de datos en el repositorio de modelos que lea el fragmento de datos.
2. Cree un conjunto de contenido y agregue un modelo de clasificador a ese conjunto.
3. Añada los valores de datos de referencia al modelo.
4. Añada los valores de etiqueta al modelo.  
Puede importar los datos del objeto de datos del repositorio de modelos. También puede introducir una única fila de datos de referencia o una única etiqueta.
5. Asigne una etiqueta a cada fila de datos de referencia.  
Puede asignar una etiqueta a varias filas en una sola operación.
6. Compile el modelo.

Después de compilar el modelo de clasificador, puede utilizar el modelo en una transformación de clasificador.

## Crear un modelo de clasificador

Utilice un objeto de datos como origen para los datos del modelo de clasificador.

Un modelo de clasificador funciona de forma óptima cuando utiliza los datos de entrada de la transformación de clasificador como origen para los datos de referencia del modelo.

1. En Object Explorer, abra o cree un conjunto de contenido.
2. Seleccione la vista **Contenido**.
3. Seleccione **Modelos de clasificador** y haga clic en **Añadir**.  
Se abrirá el asistente del modelo de clasificador.
4. Especifique un nombre para el modelo de clasificador.  
También puede escribir una descripción para el modelo.
5. Examine el repositorio de modelos y seleccione el objeto de datos que contiene los datos que importar.  
No seleccione un objeto de datos de medios sociales.  
Haga clic en **Siguiente**.
6. Revise las columnas del objeto de datos y seleccione una o más columnas que agregar al modelo.  
Puede añadir columnas de datos de referencia y una columna de etiquetas en la misma operación.
  - Para importar una columna de datos como datos de referencia, seleccione el nombre de columna y haga clic en **Datos**.  
Puede seleccionar varias columnas de datos. Developer tool fusiona el contenido de las columnas que seleccione en una única columna.

- Para importar una columna de datos como valores de etiqueta, seleccione el nombre de columna y haga clic en **Etiqueta**.

Al importar datos de referencia y valores de etiqueta, Developer tool asigna la etiqueta de cada fila a la cadena de datos de referencia de la misma fila. Puede obtener una vista previa de los datos antes de seleccionar las columnas. Puede cambiar las asignaciones de etiquetas después de crear el modelo.

Haga clic en **Siguiente**.

7. Seleccione el número de filas que quiera importar desde el origen de datos.

De forma predeterminada, Developer tool importará todas las filas del origen de datos. Si especifica un número, el modelo cuenta las filas desde el inicio del conjunto de datos.

8. Haga clic en **Finalizar** y guarde el modelo.

Después de crear el modelo, compruebe las asignaciones de etiqueta y compile el modelo.

## Añadir datos de un origen de datos a un modelo de clasificador

Puede importar varias filas de valores de datos de referencia o de etiqueta a un modelo de clasificador en una única operación.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.

2. Seleccione el nombre del modelo y haga clic en **Editar**.

3. Haga clic en **Anexar datos**.

Se abrirá el asistente del modelo de clasificador.

4. Examine el repositorio de modelos y seleccione el objeto de datos que contiene los datos que importar.

No seleccione un objeto de datos de medios sociales.

Haga clic en **Siguiente**.

5. Revise las columnas del objeto de datos y seleccione una o más columnas que agregar al modelo.

Puede añadir columnas de datos de referencia y una columna de etiquetas en la misma operación.

- Para importar una columna de datos como datos de referencia, seleccione el nombre de columna y haga clic en **Datos**.

Puede seleccionar varias columnas de datos. Developer tool fusiona el contenido de las columnas que seleccione en una única columna.

- Para importar una columna de datos como valores de etiqueta, seleccione el nombre de columna y haga clic en **Etiqueta**.

Al importar datos de referencia y valores de etiqueta, Developer tool asigna la etiqueta de cada fila a la cadena de datos de referencia de la misma fila. Puede obtener una vista previa de los datos antes de seleccionar las columnas. Puede cambiar las asignaciones de etiquetas después de crear el modelo.

Haga clic en **Siguiente**.

6. Seleccione el número de filas que quiera importar desde el origen de datos.

De forma predeterminada, Developer tool importará todas las filas del origen de datos. Si especifica un número, el modelo cuenta las filas desde el inicio del conjunto de datos.

7. Haga clic en **Finalizar** y guarde el modelo.

## Agregar una fila de datos de referencia a un modelo de clasificador

Puede agregar una única fila de datos de referencia a un modelo de clasificador.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Haga clic en **Añadir fila**.

Developer tool agrega una fila debajo de la fila actual en los datos de referencia.

4. Introduzca los valores de datos de referencia en la fila.

Puede usar métodos abreviados de Windows para pegar datos en la fila.

## Cómo añadir una etiqueta a un modelo de clasificador

Puede agregar una única etiqueta a un modelo de clasificador.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Abra el cuadro de diálogo **Administrar etiquetas**.  
El cuadro de diálogo enumera las etiquetas del modelo.
4. Haga clic en **Nueva**.  
Developer tool agrega una fila en la parte inferior de la lista de etiquetas.
5. Haga doble clic en el valor predeterminado en la columna Nombre e introduzca un nombre de etiqueta.
6. Haga clic en **Aceptar**.

Después de crear la etiqueta, puede asignarla a una o más filas de datos de referencia. La columna Uso del cuadro de diálogo **Administrar etiquetas** indica el número de filas que utilizan la etiqueta.

## Asignar una etiqueta a filas de datos de referencia

Puede asignar una etiqueta a una o varias filas de datos de referencia en una única operación.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Seleccione una o más filas de datos de referencia. Utilice las opciones de las casillas de verificación para seleccionar las filas.

**Nota:** Puede utilizar la opción de filtro para mostrar todas las filas que contienen un valor de datos que especifique. Utilice la opción de la casilla de verificación Seleccionar todo para seleccionar todas las filas que contienen el valor.

4. Haga clic en **Asignar etiqueta**.

Developer tool muestra la lista de etiquetas del modelo de clasificador.

5. Seleccione un valor de etiqueta y haga clic en **Asignar**.

Developer tool actualiza las filas de datos de referencia que ha seleccionado con el valor de etiqueta.

También puede compilar el modelo para añadir los nombres de etiqueta a la lógica del modelo de clasificador.

## Identificar valores de etiqueta no utilizados

Utilice el cuadro de diálogo **Administrar etiquetas** para buscar cualquier valor de etiqueta que siga sin utilizarse en el modelo de clasificador. El cuadro de diálogo **Administrar etiquetas** muestra los datos de uso de los valores de etiqueta en el modelo de clasificador. Utilice los datos de uso para comprobar el número de filas de datos de referencia que utilizan un valor de etiqueta y para buscar valores de etiqueta no utilizados.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Abra el cuadro de diálogo **Administrar etiquetas**. El cuadro de diálogo enumera las etiquetas del modelo de clasificador.
4. Revise los datos de la columna Uso de cada etiqueta.

La columna Uso indica el número de filas de datos de referencia que utilizan la etiqueta. Si un valor de etiqueta no se utiliza, la columna Uso tiene el valor de cero.

## Eliminar las filas de un modelo de clasificador

Puede eliminar una o más filas de datos de referencia de un modelo de clasificador en una única acción.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Seleccione una o más filas de datos de referencia. Utilice las opciones de las casillas de verificación para seleccionar las filas.
4. Haga clic en **Eliminar**.

Developer tool elimina las filas que ha seleccionado del modelo de clasificador.

Para deshacer la operación, presione las teclas Ctrl + Z del teclado.

## Cómo eliminar una etiqueta de un modelo de clasificador

Utilice el cuadro de diálogo **Administrar etiquetas** para eliminar una etiqueta de un modelo de clasificador.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Abra el cuadro de diálogo **Administrar etiquetas**.
4. Haga clic en **Eliminar**.
5. Haga clic en **Sí** para confirmar la operación.

Developer tool elimina la etiqueta del modelo. Developer tool no elimina ninguna fila de datos de referencia que utilice la etiqueta.

6. Haga clic en **Aceptar** para cerrar el cuadro de diálogo.

Para deshacer la operación, presione las teclas Ctrl + Z del teclado.

## Cómo compilar un modelo de clasificador

Cada vez que edita un valor de etiqueta o un valor de datos de referencia en un modelo de clasificador, debe compilar el modelo. Cuando compila el modelo, actualiza los datos de la compilación en el modelo.

- Para actualizar los datos de la compilación, abra el modelo en la herramienta Developer y haga clic en **Compilar**.

# Operaciones de filtro y operaciones de búsqueda

Utilice las opciones de filtro para mostrar u ocultar las filas de datos de referencia que cumplan los criterios que especifique. Al aplicar un filtro, puede realizar acciones adicionales en las filas de datos que muestre el modelo de clasificador. Por ejemplo, puede aplicar un valor de etiqueta a todas las filas de datos.

Utilice la opción de filtro para realizar las tareas siguientes:

- Buscar las filas de datos de referencia que contienen un valor que haya introducido.
- Buscar las filas de datos de referencia que utilicen una etiqueta que haya seleccionado.
- Buscar las filas de datos de referencia que no utilicen una etiqueta.

También puede buscar un valor de datos dentro de una fila de datos de referencia.

## Usar un valor de datos para filtrar las filas de datos de referencia

Utilice el filtro para comprobar que una o más filas de datos de referencia contienen los valores de datos esperados.

1. Abra el conjunto de contenido en el que se encuentra el modelo de clasificador.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Introduzca un valor en el campo Filtro.

Puede incluir caracteres comodín en el valor que escriba.

Developer tool muestra las filas de datos de referencia que contienen el texto del filtro.

## Usar un valor de etiqueta para filtrar las filas de datos de referencia

Utilice el filtro para mostrar u ocultar las filas de datos de referencia que utilizan una etiqueta que seleccione.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Seleccione un valor de etiqueta desde el menú Filtro.

Developer tool muestra las filas de datos de referencia que usan el valor de etiqueta.

**Nota:** Para buscar filas de datos de referencia que no utilizan una etiqueta, seleccione la opción **Sin etiqueta** del menú Filtro.

## Buscar un valor en una fila de datos de referencia

Utilice el campo Buscar para buscar un valor de datos en una fila que seleccione.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Seleccione una fila de datos de referencia.
4. Introduzca un valor en el campo Buscar.

El modelo resalta las instancias del valor en la fila de datos de referencia.

5. Utilice la flecha hacia arriba o la flecha hacia abajo para buscar más instancias del valor en la fila.

# Operaciones de copiar y pegar

Puede copiar un modelo de clasificador de un conjunto de contenido a otro en un repositorio de modelos. Copie un modelo de clasificador para compartir recursos con otros usuarios de la herramienta Developer.

Puede copiar un modelo a otro conjunto de contenido, o importar un modelo en el conjunto de contenido actual. Puede importar varios modelos de varios conjuntos de contenido en el repositorio en una única operación.

Cuando copia un modelo, el servicio de administración de contenido crea una copia del archivo de datos del modelo en el equipo del servicio. Cada modelo utiliza un archivo de datos diferente.

## Cómo copiar un modelo de clasificador en otro conjunto de contenido

Puede copiar un modelo de clasificador de un conjunto de contenido a otro en un repositorio de modelos. Cuando copie un modelo de clasificador, especifique el objeto de modelo y los conjuntos de contenido de origen y destino.

1. Abra el conjunto de contenido que contiene el modelo de clasificador.
2. Seleccione un modelo de clasificador y haga clic en **Copiar en**.
3. Examine el repositorio de modelos y seleccione un conjunto de contenido.

Puede copiar el modelo de clasificador en un conjunto de contenido del proyecto actual o de otro proyecto.

4. Haga clic en **Aceptar**.

La herramienta Developer copia el modelo de clasificador en el conjunto de contenido seleccionado.

## Cómo importar un modelo de clasificador desde otro conjunto de contenido

Puede importar un modelo de clasificador de un conjunto de contenido a otro en un repositorio de modelos. Cuando importe un modelo de clasificador, especifique uno o más objetos de modelo y los conjuntos de contenido de origen y destino.

1. Abra el conjunto de contenido que contendrá el modelo de clasificador.
2. Seleccione un modelo de clasificador y haga clic en **Pegar desde**.
3. Examine el repositorio de modelos y seleccione un modelo de clasificador.

Puede pegar el modelo de clasificador desde un conjunto de contenido en el proyecto actual o en otro proyecto.

4. Haga clic en **Aceptar**.

La herramienta Developer pega el modelo de clasificador en el conjunto de contenido actual.

## CAPÍTULO 5

# Modelos probabilísticos

Este capítulo incluye los siguientes temas:

- [Introducción a los modelos probabilísticos, 55](#)
- [Estructura del modelo probabilístico, 56](#)
- [Ejemplo de transformación de etiquetador, 57](#)
- [Ejemplo de transformación de analizador, 58](#)
- [Opciones del modelo probabilístico, 58](#)
- [Datos de referencia del modelo probabilístico, 61](#)
- [Datos de etiqueta del modelo probabilístico, 61](#)
- [Propiedades del modelo probabilístico, 63](#)
- [Configuración del modelo probabilístico, 63](#)
- [Operaciones de copiar y pegar, 69](#)

## Introducción a los modelos probabilísticos

Un modelo probabilístico es un objeto de datos de referencia que crea en un conjunto de contenido. Utilice un modelo probabilístico para analizar una cadena de datos que contenga varios valores de datos. Un modelo probabilístico identifica el tipo de información de cada valor en la cadena. Puede añadir un modelo probabilístico a una transformación de etiquetador y a una transformación de analizador.

Utilice un modelo probabilístico en una transformación de etiquetador para aplicar una etiqueta descriptiva a cada valor de una cadena de entrada. La transformación de etiquetador escribe las etiquetas en un solo puerto de salida. Utilice un modelo probabilístico en una transformación de analizador para escribir cada valor de una cadena de entrada en un puerto que representa la información del valor. La transformación de analizador crea un puerto de salida para cada tipo de información.

Puede diseñar y compilar un modelo probabilístico en la Herramienta del desarrollador. Al definir un modelo probabilístico, añade una serie de filas de datos al modelo y asigna una etiqueta a cada valor de cada fila. Cuando compila un modelo probabilístico, la Herramienta del desarrollador crea asociaciones entre los valores de datos y las etiquetas que haya añadido. La transformación de etiquetador y la transformación de analizador utilizan procesos de lenguajes naturales para comparar los datos del modelo probabilístico con los datos del puerto de entrada.

Los procesos de lenguajes naturales utilizan las siguientes técnicas para identificar los tipos de información de los valores de datos:

- Los procesos de lenguajes naturales pueden reconocer valores de datos similares y aplicar la misma etiqueta a los valores.
- Los procesos de lenguajes naturales pueden comparar un valor de datos con los valores adyacentes de la cadena. Los procesos de lenguajes naturales analizan la secuencia de valores para comprender el uso de cada cadena y comprobar los tipos de información que las cadenas representan.

## Estructura del modelo probabilístico

Un modelo probabilístico contiene filas de valores de datos de referencia y valores de etiqueta. Los valores de datos de referencia representan los valores diferentes que pueden aparecer en los datos de entrada de la transformación. Los valores de etiqueta identifican los tipos de información que se espera que contengan los datos de entrada.

Los modelos probabilísticos también contienen datos de compilación. La transformación de etiquetador y la de analizador utilizan los datos de compilación para medir las similitudes entre los datos de referencia del modelo y los datos de entrada de la transformación. Al compilar un modelo probabilístico, se crean o actualizan los datos de compilación.

Una fila de datos puede contener un solo valor o varios valores. Cada fila de datos puede tener una estructura diferente. Puede asignar la misma etiqueta a varios valores distintos de una fila de datos. También puede asignar una etiqueta distinta a valores idénticos que aparecen en diferentes posiciones de una fila. El Servicio de integración de datos considera las posiciones relativas de los valores en la cadena de entrada cuando la asignación se ejecuta. Asigne cada etiqueta a al menos un valor de datos antes de compilar el modelo probabilístico.

Developer tool escribe los valores de datos de referencia, los valores de etiqueta y los datos de compilación en un archivo de la estructura del directorio de Informatica. El objeto del modelo probabilístico del repositorio de modelos almacena el nombre de archivo. Al guardar un modelo probabilístico, se escriben los valores de datos de referencia y los valores de etiqueta actuales en el archivo. Cuando se compila el modelo, se actualizan los datos de la compilación en el archivo. Puede leer el nombre de archivo desde las propiedades del modelo en Developer tool.

**Nota:** Para optimizar las capacidades del modelo probabilístico, compruebe que cada fila de datos contiene varios valores de datos de referencia. El orden de los valores en cada fila debe corresponderse en lo posible con el orden en el que aparecen los valores en los datos de entrada de la transformación. Si las filas de datos contienen valores de datos de referencia únicos, la transformación de etiquetador o la transformación de analizador no pueden aplicar procesos de lenguaje natural durante el análisis probabilístico.



## Ejemplo de transformación de etiquetador

La base de datos de clientes de una empresa de seguros contiene múltiples errores de entrada de datos. Usted trabaja como gestor de datos en la empresa de seguros. Configura una asignación con una transformación de etiquetador para determinar los distintos tipos de datos que contiene cada columna.

La siguiente tabla describe datos de ejemplo de la base de datos de clientes:

ID de fila	Campo 1	Campo 2	Campo 3
1	19132954	AIM SECURITIES	PETRIE TAYBRO
2	10110169	JASE TRAPANI	BANK OF NEW YORK
3	10111786	WANGER ASSET MANAGEMENT, LLP	JAN SEEDORF
4	10112299	FELIX LEVINGER	HARVARD MAGAZINE
5	10112036	DESCHÊNES & FILS LTÉE (QUEBEC)	RICHARD TREMBLAY
6	BERGER ASSOCIATES	10111101	DAREEN HULSMAN
7	19131385	EAGLE FINANCIAL GROUP INC	PATRICK MCKINNIE
8	LAKENYA PASKETT	WHITEHALL FINANCIAL GROUP	15954710

Al ejecutar la asignación, la transformación de etiquetador compara los datos de entrada con los datos de referencia del modelo probabilístico. La transformación de etiquetador selecciona una etiqueta para los datos de cada puerto de entrada. La transformación escribe las etiquetas en un puerto de salida. Cada fila de salida contiene un conjunto de etiquetas que definen la estructura de datos en la fila de entrada correspondiente.

La siguiente tabla describe las etiquetas que la transformación de etiquetador añade al puerto de salida:

ID de fila	Etiquetas de salida
1	número organización contacto
2	número contacto organización
3	número organización contacto
4	número contacto organización
5	número organización contacto
6	organización número contacto
7	organización número contacto
8	contacto organización número

## Ejemplo de transformación de analizador

Un supermercado almacena las descripciones de los productos en una única columna en una tabla de base de datos. Las descripciones de los productos contienen varios valores de datos que representan los distintos tipos de información. Usted trabaja como gestor de datos en el supermercado. Desea crear columnas para los diferentes tipos de información en las descripciones de los productos.

Configura una asignación con una transformación de analizador para organizar los valores de datos en los campos correctos.

El fragmento de datos siguiente contiene la descripción del producto para zumo de naranja:

Sunnydream Zumo de naranja Sin azúcar 340 g

La siguiente tabla describe los datos de salida que la transformación de analizador crea a partir de los datos de entrada:

Nombre de producto	Tipo de producto	Detalles del producto	Tamaño del producto
Sunnydream	Zumo de naranja	Sin azúcar	340 g

## Opciones del modelo probabilístico

Al editar un modelo probabilístico, puede trabajar en las vistas Datos o Etiqueta.

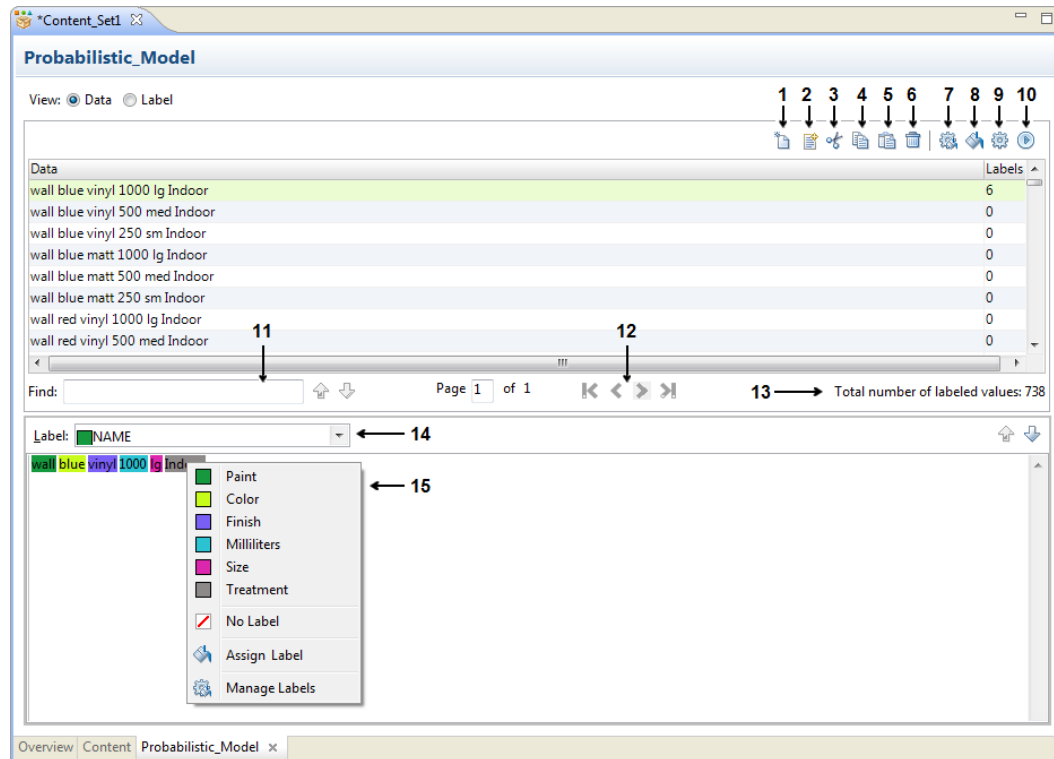
Utilice la vista Datos para añadir filas de datos de referencia al modelo y para asignar etiquetas a los valores de datos de cada fila. Utilice la vista Etiqueta para revisar la información detallada sobre el uso de los valores de etiqueta en el modelo. Puede agregar etiquetas al modelo probabilístico en la vista Datos y en la vista Etiqueta.

### Vista Datos del modelo probabilístico

La vista Datos muestra las filas de datos de referencia en el modelo probabilístico y el número de valores de etiqueta que asigna a cada fila. La vista Datos también muestra el número total de etiquetas que ha asignado a los valores del modelo actual.

Al seleccionar una fila de datos de referencia, los valores de la fila aparecen en un editor debajo del campo Buscar. Para asignar una etiqueta a un valor de datos de referencia de una fila, haga clic con el botón derecho en el valor en el editor y seleccione un valor de etiqueta.

La siguiente imagen muestra las opciones del modelo probabilístico que puede utilizar cuando selecciona la vista Datos:



La vista Datos incluye las opciones siguientes:

1. Añadir fila  
Inserta una fila de datos en blanco.
2. Anexar datos  
Importa los datos de un objeto de datos del repositorio de modelos.
3. Cortar  
Elimina una fila de datos del modelo probabilístico y añade la fila de datos al portapapeles.
4. Copiar  
Copia una fila de datos al portapapeles.
5. Pegar  
Pega una fila de datos del portapapeles en el modelo probabilístico.
6. Eliminar  
Elimina una fila de datos del modelo probabilístico.
7. Administrar etiquetas  
Abre el cuadro de diálogo **Administrar etiquetas**. Utilice el cuadro de diálogo para agregar o eliminar valores de etiqueta del modelo probabilístico.
8. Asignar etiqueta  
Asigna una etiqueta a uno o más valores de datos de referencia que seleccione. Puede utilizar la opción de asignar una etiqueta a todas las instancias de un valor de datos de referencia del modelo.
9. Editar propiedades

Muestra las propiedades del modelo probabilístico.

10. Compilar

Compila el modelo probabilístico.

11. Campo Buscar

Busca filas en el modelo que contienen el valor de datos de referencia que ha introducido. Utilice la flecha hacia arriba y la flecha hacia abajo para desplazarse a las filas que contienen el valor.

12. Flechas de avance y retroceso

Avanza y retrocede por las filas de valores de datos del modelo.

13. Número total de valores etiquetados

Indica el número de valores de datos de referencia que utilizan una etiqueta.

14. Campo Etiqueta

Muestra un valor de etiqueta que puede aplicar al valor de datos de referencia que seleccione.

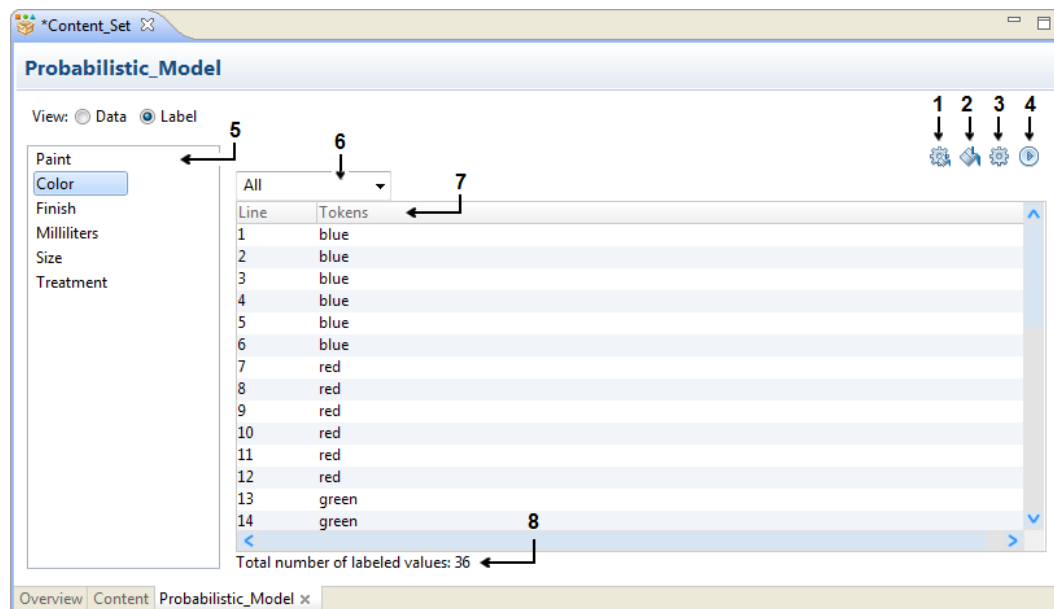
15. Menú Etiqueta

Muestra una lista de opciones que puede utilizar para asignar una etiqueta a uno o más valores de datos de referencia. Para abrir el menú, haga clic con el botón derecho en un valor de datos de referencia del editor de datos de referencia.

## Vista Etiqueta del modelo probabilístico

La vista Etiqueta enumera las etiquetas que define en el modelo probabilístico. Cuando seleccione una etiqueta, la vista Etiqueta muestra los valores de datos que ha asignado a la etiqueta en cada fila.

La siguiente imagen muestra las opciones del modelo probabilístico que puede utilizar cuando selecciona la vista Datos:



La vista Etiqueta incluye las opciones siguientes:

1. Administrar etiquetas

Abre el cuadro de diálogo **Administrar etiquetas**. Utilice el cuadro de diálogo para agregar o eliminar valores de etiqueta del modelo probabilístico.

2. Asignar etiqueta  
Asigna una etiqueta a uno o más valores de datos de referencia que seleccione.  
Puede asignar una etiqueta a un solo valor de datos o puede asignar una etiqueta a varios valores en una única operación.
3. Editar propiedades  
Muestra las propiedades del modelo probabilístico.
4. Compilar  
Compila el modelo probabilístico.
5. Lista de valores de etiqueta  
Muestra las etiquetas que puede asignar a los valores de datos de referencia del modelo.
6. Filtro de asignación  
Filtra la lista de valores de datos de referencia que utilizan la etiqueta seleccionada. Las opciones de filtro muestran u ocultan los valores de datos de referencia según el método que se utiliza para asignar la etiqueta a los valores de datos.  
Al aplicar un filtro, el número total de valores con etiqueta de la vista Etiqueta refleja el número de valores que cumplen la condición del filtro.
7. Columna de valor de datos de referencia  
Muestra los valores de datos de referencia que utilizan la etiqueta actual.
8. Número total de valores etiquetados  
Indica el número de valores de datos de referencia que utilizan la etiqueta actual.

## Datos de referencia del modelo probabilístico

Los valores de datos de referencia de un modelo probabilístico representan los tipos de datos de entrada que puede conectar a una transformación en una asignación.

Puede añadir, editar y eliminar filas de datos de referencia en Developer tool. Puede pegar datos del portapapeles y puede importar datos desde un origen de datos. Después de agregar los valores de datos de referencia, asigne una etiqueta a cada valor de datos de cada fila.

## Datos de etiqueta del modelo probabilístico

Los valores de etiqueta de un modelo probabilístico representan los tipos de información que los valores de datos de referencia pueden contener. Al añadir filas de datos de referencia a un modelo, asigne una etiqueta a cada valor en cada fila. Las etiquetas que añade al modelo aparecen en la vista Etiqueta y en las opciones del menú de la vista Datos.

Puede asignar cualquier etiqueta en el modelo a cualquier valor de datos de referencia. Si el mismo valor tiene distintos significados en diferentes filas de datos de referencia, puede asignar una etiqueta diferente a cada valor en cada fila.

El rango de valores de etiquetas puede corresponder al rango de puertos de entrada que la transformaciones de etiquetador o de analizador leen durante análisis probabilísticos. El modelo probabilístico debe contener

al menos un valor de etiqueta que la transformación puede aplicar a los valores de datos en cada puerto de entrada.

Por ejemplo, un almacén puede almacenar datos de inventario en un archivo separado por comas que define ocho columnas. Diseñará una asignación que analice los datos de inventario a una tabla de base de datos. Creará un modelo probabilístico con un valor de etiqueta para cada columna de datos. Al ejecutar la asignación, la transformación de analizador escribe cada valor de los datos de entrada en la columna correcta de la tabla de destino.

La tabla siguiente muestra las columnas de datos de inventario y los valores de etiqueta que puede crear en un modelo probabilístico:

Nombre de columna de inventario	Nombre de etiqueta
Product_Name	Product_Name
Quantity	Quantity
Location	Location
Barcode	Barcode
SKU	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

**Nota:** Puede utilizar los nombres de columna de entrada u otros nombres. No es necesario que los nombres coincidan.

## Etiqueta de desbordamiento

Cuando una transformación no puede aplicar una etiqueta a un valor de datos de entrada, la transformación trata el valor de datos como datos de desbordamiento. La transformación de etiquetador aplica una etiqueta de desbordamiento a cualquier valor de datos que no pueda identificar. La transformación de analizador escribe cualquier valor de datos que no pueda identificar en un puerto de desbordamiento.

La siguiente tabla muestra cómo una transformación de analizador utiliza un puerto de desbordamiento para analizar elementos de datos de direcciones que un modelo probabilístico no reconoce:

Datos de entrada	Puerto Street_Name	Puerto Street_Descriptor	Puerto de desbordamiento
Park Place	Park	Place	Sin datos de desbordamiento
Park Avenue	Park	Avenue	Sin datos de desbordamiento
Madison Avenue	Madison	Avenue	Sin datos de desbordamiento
Central Park	Central	Park	Sin datos de desbordamiento
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

La transformación de analizador también escribe valores en un puerto de desbordamiento cuando el número de valores de entrada es mayor que el número de etiquetas del modelo. Antes de usar un modelo probabilístico en una transformación, revise los datos de entrada y compruebe que el modelo contiene el número correcto de valores de etiqueta.

## Propiedades del modelo probabilístico

Puede revisar las propiedades generales y las propiedades avanzadas de un modelo probabilístico.

Para abrir el editor de propiedades, seleccione la opción **Editar propiedades** en la vista Datos o en la vista Etiqueta.

Las propiedades generales muestran el nombre del modelo probabilístico, cualquier descripción del modelo y el nombre del archivo de datos del modelo. Las propiedades avanzadas muestran las propiedades computacionales que utiliza Developer tool para compilar el modelo probabilístico.

El elemento básico en la compilación de modelos probabilísticos es *n-gram*. Un *n-gram* es una serie de letras que siguen o preceden a otras letras para completar una palabra. Cuando se ejecuta una asignación, la transformación de etiquetador o analizador crea varios *n-gram* para cada valor en la columna de datos de referencia del modelo probabilístico. La transformación compara los valores de los datos de entrada con los valores de los datos de referencia y los *n-gram*. Las propiedades avanzadas de un modelo probabilístico determinan cómo administra el modelo probabilístico los *n-gram* y otras características.

**Nota:** Los valores predeterminados de las propiedades avanzadas representan la configuración preferida para el análisis probabilístico y la compilación del modelo probabilístico. Si edita una propiedad avanzada, es posible que afecte negativamente a la exactitud del análisis probabilístico. No edite las propiedades avanzadas a menos que entienda los efectos de los cambios que va a efectuar.

### TEMAS RELACIONADOS

- [“Reglas y directrices de los modelos probabilísticos y de clasificador” en la página 40](#)

## Configuración del modelo probabilístico

Los pasos para configurar un modelo probabilístico comienzan con el tipo de análisis que desea realizar. Utilice un modelo probabilístico en una transformación de etiquetador para identificar los tipos de información en cada valor de una cadena de entrada. Utilice un modelo probabilístico en una transformación de analizador para analizar los valores de datos de una cadena de entrada en distintos puertos de salida.

Puede utilizar el mismo modelo probabilístico para etiquetar y analizar datos. Cuando utilice el modelo en una transformación de etiquetador, la transformación crea un solo puerto de salida para cada puerto de entrada que seleccione. Cuando utilice el modelo en una transformación de analizador, la transformación crea un puerto de salida para cada tipo de datos de entrada que identifica.

Para crear un modelo probabilístico, realice las tareas siguientes:

1. Identifique los valores de datos de referencia y los valores de etiqueta que agregar al modelo.  
Puede utilizar un fragmento de los datos que desea analizar. Cree un objeto de datos en el repositorio de modelos que lea el fragmento de datos.
2. Cree un conjunto de contenido y agregue un modelo probabilístico a ese conjunto.

3. Añada los valores de datos de referencia al modelo.

4. Añada los valores de etiqueta al modelo.

Puede importar los datos del objeto de datos del repositorio de modelos. También puede introducir una única fila de datos de referencia o una única etiqueta.

Para utilizar el modelo probabilístico para analizar datos, compruebe que el modelo contiene un valor de etiqueta para cada puerto de salida que la transformación debe crear.

5. Asigne una etiqueta para hacer referencia a cada valor de datos de cada fila.

Puede asignar una etiqueta a varios valores de datos de referencia en una única operación.

6. Compile el modelo.

Después de compilar el modelo probabilístico, puede utilizar el modelo en una transformación.

## Crear un modelo probabilístico vacío

Puede crear un objeto del modelo probabilístico que no contenga datos de referencia ni datos de etiqueta. Cree el modelo vacío y añada o importe datos al modelo.

1. En Object Explorer, abra o cree un conjunto de contenido.

2. Seleccione la vista Contenido.

3. Seleccione **Modelos probabilísticos** y haga clic en **Añadir**.

Se abrirá el asistente del modelo probabilístico.

4. Seleccione la opción **Modelo probabilístico**.

Haga clic en **Siguiente**.

5. Especifique un nombre para el modelo probabilístico.

También puede escribir una descripción para el modelo.

6. Haga clic en **Finalizar**.

## Cómo crear un modelo probabilístico a partir de un objeto de datos

Puede utilizar un objeto de datos como origen de datos para los datos del modelo probabilístico.

Un modelo probabilístico funciona de forma óptima cuando utiliza los datos de entrada de las transformaciones de etiquetador o de analizador como origen para los datos de referencia del modelo.

1. En Object Explorer, abra o cree un conjunto de contenido.

2. Seleccione la vista **Contenido**.

3. Seleccione **Modelos probabilísticos** y haga clic en **Añadir**.

Se abrirá el asistente del modelo probabilístico.

4. Seleccione la opción **Modelo probabilístico de los objetos de datos**.

Haga clic en **Siguiente**.

5. Especifique un nombre para el modelo probabilístico.

También puede escribir una descripción para el modelo.

6. Examine el repositorio de modelos y seleccione el objeto de datos que contiene los datos que importar.

No seleccione un objeto de datos de medios sociales.

Haga clic en **Siguiente**.



7. Revise las columnas del objeto de datos y seleccione una o más columnas que agregar al modelo. Puede añadir columnas de datos de referencia y una columna de etiquetas en la misma operación.
  - Para importar una columna de datos como datos de referencia, seleccione el nombre de columna y haga clic en **Datos**.

Puede seleccionar varias columnas de datos. Developer tool fusiona el contenido de las columnas que seleccione en una única columna.
  - Para importar una columna de datos como valores de etiqueta, seleccione el nombre de columna y haga clic en **Etiqueta**.

Al importar datos de referencia y valores de etiqueta, Developer tool asigna la etiqueta de cada fila a la cadena de datos de referencia de la misma fila. Puede obtener una vista previa de los datos antes de seleccionar las columnas. Puede cambiar las asignaciones de etiquetas después de crear el modelo.

Haga clic en **Siguiente**.
8. Seleccione el número de filas que quiera importar desde el origen de datos.

De forma predeterminada, Developer tool importará todas las filas del origen de datos. Si especifica un número, el modelo cuenta las filas desde el inicio del conjunto de datos.
9. Especifique los delimitadores de los valores de datos que va a importar.

Puede especificar distintos delimitadores para valores de datos de referencia y valores de etiqueta. El delimitador predeterminado es un carácter de espacio.
10. Haga clic en **Finalizar** y guarde el modelo.

Después de crear el modelo probabilístico, compruebe las asignaciones de etiqueta y compile el modelo.

## Añadir datos de un origen de datos a un modelo probabilístico

Puede importar varias filas de valores de datos de referencia y de etiqueta a un modelo probabilístico en una única operación.

1. Abra el conjunto de contenido en el que se encuentra el modelo probabilístico.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Haga clic en **Anexar datos**.

Se abrirá el asistente del modelo probabilístico.
4. Examine el repositorio de modelos y seleccione el objeto de datos que contiene los datos que importar.

No seleccione un objeto de datos de medios sociales.

Haga clic en **Siguiente**.
5. Revise las columnas del objeto de datos y seleccione una o más columnas que agregar al modelo. Puede añadir columnas de datos de referencia y una columna de etiquetas en la misma operación.
  - Para importar una columna de datos como datos de referencia, seleccione el nombre de columna y haga clic en **Datos**.

Puede seleccionar varias columnas de datos. Developer tool fusiona el contenido de las columnas que seleccione en una única columna.
  - Para importar una columna de datos como valores de etiqueta, seleccione el nombre de columna y haga clic en **Etiqueta**.

Al importar datos de referencia y valores de etiqueta, Developer tool asigna la etiqueta de cada fila a la cadena de datos de referencia de la misma fila. Puede obtener una vista previa de los datos antes de seleccionar las columnas. Puede cambiar las asignaciones de etiquetas después de crear el modelo.

Haga clic en **Siguiente**.

6. Seleccione el número de filas que quiera importar desde el origen de datos.  
De forma predeterminada, Developer tool importará todas las filas del origen de datos. Si especifica un número, el modelo cuenta las filas desde el inicio del conjunto de datos.
7. Especifique los delimitadores de los valores de datos que va a importar.  
Puede especificar distintos delimitadores para valores de datos de referencia y valores de etiqueta. El delimitador predeterminado es un carácter de espacio.
8. Haga clic en **Finalizar** y guarde el modelo.

## Agregar una fila de datos de referencia a un modelo probabilístico

Utilice la vista Datos para agregar una fila vacía a un modelo probabilístico.

1. Abra el conjunto de contenido en el que se encuentra el modelo.  
Seleccione el nombre del modelo y haga clic en **Editar**.
2. Seleccione la vista Datos.
3. Para agregar una fila vacía al modelo, haga clic en **Nuevo**.
4. Seleccione la fila que ha agregado e introduzca uno o más valores de datos de referencia en la fila.
5. Guarde el modelo probabilístico.

Después de guardar el modelo, asigne una etiqueta a cada valor de la fila. También puede compilar el modelo.

## Cómo añadir una etiqueta a un modelo probabilístico

Puede agregar una única etiqueta a un modelo probabilístico. Añada una etiqueta para cada tipo de información que los valores de datos del modelo representen. Si utiliza el modelo probabilístico en una transformación de analizador, añada una etiqueta para cada puerto de salida que espera que cree la transformación.

1. Abra el conjunto de contenido que contiene el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. En las vistas Datos o Etiqueta, haga clic en **Administrar etiquetas**.  
Aparecerá el cuadro de diálogo **Administrar etiquetas**.
4. En el cuadro de diálogo **Administrar etiquetas**, haga clic en **Nuevo**.  
Aparecerá una etiqueta en la primera fila vacía del cuadro de diálogo.
5. Edite el nombre de la etiqueta. Si lo desea, puede actualizar el color de la etiqueta.
6. Haga clic en **Aceptar** para añadir la etiqueta al modelo.
7. Guarde el modelo probabilístico.

Después de añadir la etiqueta, debe asignarla a un valor de datos como mínimo.

## Asignar una etiqueta a un valor de datos de referencia

Puede asignar una etiqueta a un solo valor de datos en una fila de datos de referencia.

Puede asignar distintas etiquetas al mismo valor de datos si el valor de datos aparece en distintas ubicaciones en la fila o en filas distintas.

1. Abra el conjunto de contenido en el que se encuentra el modelo.

2. Seleccione el nombre del modelo y haga clic en **Editar**.
  3. Seleccione la vista Datos.
  4. Busque un valor de datos que no tenga una etiqueta o que tenga una etiqueta incorrecta. Los valores de datos que utilizan una etiqueta están codificados por color.
  5. Seleccione la fila de datos que contiene el valor de datos.  
La fila aparecerá en el editor.
  6. Haga clic con el botón derecho sobre un valor de datos en el editor y seleccione una etiqueta en el menú contextual.  
Developer tool asigna la etiqueta al valor de datos.
  7. Guarde el modelo probabilístico.
- Después de guardar el modelo probabilístico, puede compilarlo.

## Asignar una etiqueta a varios valores de datos

Puede asignar una etiqueta a varios valores de datos de referencia en una única operación.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Haga clic en **Asignar etiqueta**.  
Se abrirá el cuadro de diálogo **Asignar una etiqueta a varios valores**.
4. Introduzca uno o más caracteres en el campo Buscar.  
Puede escribir caracteres comodín en el campo Buscar.
5. También puede seleccionar otros criterios de búsqueda.  
Puede seleccionar o desactivar las opciones siguientes:
  - Coincidir mayúsculas y minúsculas.  
Especifica que la operación de búsqueda distingue entre mayúsculas y minúsculas. No utilice caracteres comodín con esta opción.
  - Hacer coincidir la cadena completa. Especifica que la operación de búsqueda intenta encontrar una coincidencia completa entre los caracteres del valor de datos de referencia y los caracteres que introduzca. No utilice caracteres comodín con esta opción.
  - Omitir valores etiquetados.  
Especifica que la operación de búsqueda omite todos los valores de datos de referencia que utilizan una etiqueta.
6. Seleccione una etiqueta para asignarla a los valores de datos de referencia que coinciden con los criterios de búsqueda.  
También puede seleccionar la opción **Sin etiqueta**. Seleccione esta opción para eliminar la etiqueta de los valores de datos de referencia que incluyan los caracteres que introduzca.
7. Haga clic en **Iniciar**.  
Developer tool asigna la etiqueta a todos los valores de datos de referencia que coinciden con los criterios de búsqueda definidos.

**Nota:** Para ver los valores de datos de referencia que ha etiquetado en una única operación, utilice el filtro **Asignado en masa** en la vista Etiqueta.

## Eliminar las filas de un modelo probabilístico

Puede eliminar una o más filas de datos de referencia de un modelo probabilístico en una única acción.

1. Abra el conjunto de contenido en el que se encuentra el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. En la vista Datos, seleccione una o más filas de datos de referencia.
4. Haga clic en **Eliminar**.

Developer tool elimina las filas que ha seleccionado del modelo de clasificador.

Para deshacer la operación, presione las teclas Ctrl + Z del teclado.

## Eliminar una etiqueta de un modelo probabilístico

Al eliminar un valor de etiqueta de un modelo, todos los valores de datos de referencia que utilizase la etiqueta permanecen en el modelo. Asigne otro valor de etiqueta a cada valor de datos de referencia.

1. Abra el conjunto de contenido que contiene el modelo.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. En las vistas Datos o Etiqueta, haga clic en **Administrar etiquetas**.
4. En el cuadro de diálogo **Administrar etiquetas**, seleccione un valor de etiqueta.
5. Haga clic en **Eliminar**.
6. Haga clic en **Aceptar** para eliminar la etiqueta.
7. Guarde el modelo probabilístico.

**Nota:** Una etiqueta es un elemento estructural de un modelo probabilístico. Si añade o quita una etiqueta después de añadir el modelo a una transformación, puede invalidar la operación que utiliza el modelo. Para utilizar el modelo que se ha actualizado, elimine y vuelva a crear la operación de transformación.

## Cómo compilar el modelo probabilístico

Al actualizar las asignaciones de etiquetas o los datos en un modelo probabilístico, puede compilar el modelo. Compile el modelo para actualizar la lógica del modelo con las asociaciones entre los valores de datos de referencia actuales y los valores de etiqueta actuales.

Antes de compilar el modelo probabilístico, compruebe que cada valor de etiqueta identifica al menos uno valor de datos de referencia.

- Para compilar el modelo, abra el modelo en Developer tool y haga clic en **Compilar**.

## Buscar filas de datos en un modelo probabilístico

Utilice la vista Datos para buscar las filas de datos de referencia que contienen un valor que ha introducido.

1. Abra el conjunto de contenido en el que se encuentra el modelo probabilístico.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. Seleccione la vista Datos.
4. Introduzca uno o más caracteres en el campo **Buscar**.

La vista Datos muestra la primera fila del modelo que contiene el valor que ha introducido.

5. Utilice la flecha hacia arriba o la flecha hacia abajo para desplazarse a otras filas que contienen el valor.

## Filtrar valores de datos de referencia por asignación de etiqueta

Utilice la vista Etiqueta para localizar los valores de datos de referencia que utilicen una etiqueta especificada. Filtre los resultados en función del método que haya utilizado para asignar la etiqueta.

1. Abra el conjunto de contenido en el que se encuentra el modelo probabilístico.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. En la vista Etiqueta, seleccione un valor de etiqueta.

El modelo probabilístico muestra una lista de los valores de datos de referencia que utilizan la etiqueta. El modelo también muestra el número de valores de datos que usan la etiqueta.

4. Aplique un filtro a la lista de valores de datos de referencia que utilizan la etiqueta.

Seleccione uno de estos filtros:

- Todos. Muestra los valores de datos de referencia que utilizan la etiqueta. Esta es la opción predeterminada.
- Asignado por el usuario. Muestra los valores de referencia de datos que se han seleccionado de forma individual al asignar la etiqueta.
- Asignado en masa. Muestra los valores de datos de referencia a los que ha asignado una etiqueta como parte de una operación de asignación en masa.

El modelo probabilístico muestra los valores de datos de referencia que cumplen la condición del filtro.

## Buscar valores de etiqueta no utilizados

Utilice la vista Etiqueta para buscar cualquier valor de etiqueta que no haya asignado a un valor de datos de referencia. Debe asignar cada etiqueta a al menos un valor de datos de referencia.

1. Abra el conjunto de contenido en el que se encuentra el modelo probabilístico.
2. Seleccione el nombre del modelo y haga clic en **Editar**.
3. En la vista Etiqueta, seleccione un valor de etiqueta.

El modelo probabilístico muestra una lista de los valores de datos de referencia que utilizan la etiqueta. El modelo también muestra el número total de valores de datos que usan la etiqueta.

Si el número total de valores de datos es cero, no ha asignado la etiqueta a ningún valor de datos de referencia en el modelo probabilístico.

## Operaciones de copiar y pegar

Puede copiar un modelo probabilístico de un conjunto de contenido a otro en un repositorio de modelos. Copie un modelo probabilístico para compartir recursos con otros usuarios de la herramienta Developer.

Puede copiar un modelo a otro conjunto de contenido, o importar un modelo en el conjunto de contenido actual. Puede importar varios modelos de varios conjuntos de contenido en el repositorio en una única operación.

Cuando copia un modelo, el servicio de administración de contenido crea una copia del archivo de datos del modelo en el equipo host de los servicios de Informática. Cada modelo utiliza un archivo de datos diferente.

## Cómo copiar un modelo probabilístico en otro conjunto de contenido

Puede copiar un modelo probabilístico de un conjunto de contenido a otro en un repositorio de modelos. Cuando copie un modelo probabilístico, especifique el objeto de modelo y los conjuntos de contenido de origen y destino.

1. Abra el conjunto de contenido que contiene el modelo probabilístico.

2. Seleccione un modelo probabilístico y haga clic en **Copiar en**.

3. Examine el repositorio de modelos y seleccione un conjunto de contenido.

Puede copiar el modelo probabilístico en un conjunto de contenido del proyecto actual o de otro proyecto.

4. Haga clic en **Aceptar**.

La herramienta Developer copia el modelo probabilístico en el conjunto de contenido seleccionado.

## Cómo importar un modelo probabilístico desde otro conjunto de contenido

Puede importar un modelo probabilístico de un conjunto de contenido a otro en un repositorio de modelos. Cuando importe un modelo probabilístico, especifique uno o más objetos de modelo y los conjuntos de contenido de origen y destino.

1. Abra el conjunto de contenido que contendrá el modelo probabilístico.

2. Seleccione un modelo probabilístico y haga clic en **Pegar desde**.

3. Examine el repositorio de modelos y seleccione un modelo de probabilístico.

Puede pegar el modelo probabilístico desde un conjunto de contenido en el proyecto actual o en otro proyecto.

4. Haga clic en **Aceptar**.

La herramienta Developer pega el modelo probabilístico en el conjunto de contenido actual.

## Copiar las filas de datos de referencia en el portapapeles

Puede copiar una o más filas de datos de referencia de un modelo probabilístico en el portapapeles. Puede pegar las filas a otro modelo probabilístico.

1. Abra el conjunto de contenido en el que se encuentra el modelo probabilístico.

2. Seleccione el nombre del modelo y haga clic en **Editar**.

3. En la vista Datos, seleccione una o más filas de datos de referencia.

4. Utilice las teclas Ctrl + C para copiar las filas en el portapapeles.

La operación copia los datos de referencia y los valores de etiqueta que ha asignado a los datos de referencia.

Puede utilizar las teclas Ctrl + V para pegar las filas en un editor de texto o en la vista Datos de otro modelo probabilístico.

## APÉNDICE A

# Datos de referencia e Informatica Data Engineering Integration

- [Datos de referencia e Informatica Data Engineering Integration, 71](#)

## Datos de referencia e Informatica Data Engineering Integration

Informatica Data Engineering Integration combina un dominio de Informatica y aplicaciones cliente de Informatica con un clúster de Hadoop. Puede insertar una asignación desde Developer tool hasta el clúster y ejecutar la asignación en los nodos del clúster.

La operación de inserción copia los datos de la tabla de referencia, los datos del conjunto de contenido y los datos de llenado de identidad que la asignación lee en el clúster. Una vez que la asignación se ejecuta, el clúster elimina los datos de referencia que la operación de inserción ha copiado con la asignación.

La operación de inserción no copia datos de referencia de validación de direcciones. Si inserta una asignación que realice validación de direcciones, deberá instalar los archivos de datos de referencia de validación de direcciones en cada nodo que ejecute la asignación. El clúster no elimina los archivos de datos de referencia de validación de direcciones después de que se ejecute la asignación de validación de direcciones.

### Datos de referencia para validación de direcciones

Cuando se ejecuta una asignación de validación de direcciones en un entorno Hadoop, los archivos de datos de referencia de direcciones deben encontrarse en cada uno de los nodos de datos en que se ejecuta la asignación. Informatica Data Engineering Integration se instala con un script de shell que se puede utilizar para instalar los archivos en los nodos de datos.

Utilice el script de shell para instalar los archivos de datos de referencia de direcciones en los nodos de datos en una sola operación. El script lee un archivo que contiene los nombres o direcciones IP de los nodos. El script copia los archivos de datos de referencia de direcciones en cada nodo que el archivo identifique.

El nombre del script de es `copyRefDataToComputeNodes.sh`.

Busque el script en el siguiente directorio de la instalación de Data Engineering Integration:

<Directorio de instalación de Informatica>/tools/dq/av

En la tabla siguiente se describen las opciones que utiliza el script:

Opción	Descripción
-n	El archivo que contiene la lista de nombres o direcciones IP de los nodos de datos en el clúster de Hadoop. Introduzca cada nombre o dirección IP de nodo en una línea separada del archivo. De forma predeterminada, el script lee el archivo desde el directorio \$BASEDIR/HadoopDataNodes, donde \$BASEDIR es la ubicación del script de shell.
-p	Un mensaje para confirmar que desea instalar los archivos de datos de referencia de direcciones. De forma predeterminada, el script muestra un mensaje para confirmar que desea copiar los archivos del directorio de origen en los directorios de destino de los nodos de datos. Si ejecuta el script de shell en un programa, puede deshabilitar el mensaje. El valor predeterminado de la opción es Y. Para deshabilitar el mensaje, defina el valor en N.
-s	El directorio de origen de los archivos de datos de referencia de direcciones que el script copia a los nodos. De forma predeterminada, el script lee los archivos del directorio /reference_data del equipo local. <b>Nota:</b> Los archivos de datos de referencia de direcciones utilizan la extensión de nombre de archivo .MD. El directorio de origen solo debe contener los archivos de datos de referencia de direcciones.
-t	El directorio de cada nodo en el que el script copia los archivos de datos de referencia de direcciones. De forma predeterminada, el script copia los archivos en el directorio /reference_data de cada nodo.
-u	El nombre del usuario que ejecuta el script. El usuario debe tener acceso de shell seguro sin contraseña a los nodos.

## Instalación de los archivos de datos de referencia de direcciones

Para instalar archivos de datos de referencia de direcciones en los nodos de datos de un clúster de Hadoop, ejecute el script de shell `copyRefDataToComputeNodes.sh`. También puede definir un trabajo para ejecutar el script de shell en una aplicación de programador de trabajos a los intervalos de tiempo que especifique.

Antes de ejecutar el script o definir el trabajo, revise los valores de opciones especificados para el script. Puede aceptar los valores predeterminados o actualizarlos.

### Instalación de los archivos de datos de referencia de direcciones en la línea de comandos

Para instalar los archivos en la línea de comandos, siga estos pasos:

1. En la línea de comandos, abra el directorio siguiente:  
`<Directorio de instalación de Informatica>/tools/dq/av`
2. Ejecute `copyRefDataToComputeNodes.sh`.  
Opcionalmente, introduzca uno o más valores para las opciones de script. Si no introduce un valor para una opción, el script se ejecuta con el valor predeterminado de la opción.  
De forma predeterminada, el script le solicita que confirme la instalación de los archivos. Para instalar los archivos, escriba Y.

### Instalación de los archivos de datos de referencia de direcciones con un trabajo programado

Puede definir un trabajo para ejecutar el script de shell a los intervalos de tiempo que especifique. Agregue el trabajo a una aplicación de programador de trabajos. Si define un trabajo para instalar los archivos, debe deshabilitar el mensaje para confirmar la instalación.

Para deshabilitar el mensaje, defina la siguiente opción en el script de shell:

`-p n`



# INDICE

## A

administrar columnas  
  tablas de referencia [23](#)  
administrar filas  
  tablas de referencia [24](#)

## C

conjuntos de caracteres [36](#)  
conjuntos de contenido  
  conjuntos de caracteres [36](#)  
  conjuntos de patrones [37](#)  
  conjuntos de tokens [39](#)  
  control de versiones [13](#), [22](#), [30](#)  
  expresiones regulares [38](#)  
  modelos de clasificador [37](#)  
  modelos probabilísticos [37](#)  
conjuntos de patrones [37](#)  
conjuntos de tokens [39](#)  
control de versiones  
  conjuntos de contenido [13](#), [30](#)  
  tablas de referencia [13](#)  
  tablas de referencia de Developer tool [30](#)  
  tablas de referencia de la Herramienta del analista [22](#)  
crear una tabla de referencia a partir de datos de columnas de perfil  
  tablas de referencia [17](#)  
crear una tabla de referencia a partir de patrones de columna  
  tablas de referencia [18](#)  
crear una tabla de referencia manualmente  
  tablas de referencia [16](#)

## D

Data Engineering Integration  
  instalación de los datos de referencia de direcciones [72](#)  
  requisitos de los datos de referencia [71](#)  
  script de instalación de datos de referencia de direcciones [71](#)

## E

Entorno Hadoop  
  instalación de los datos de referencia de direcciones [72](#)  
  requisitos de los datos de referencia [71](#)  
  script de instalación de datos de referencia de direcciones [71](#)  
exportar una tabla de referencia  
  tablas de referencia [25](#)  
expresiones regulares [38](#)

## H

Herramienta del analista  
  buscar y reemplazar valores de datos de referencia [25](#)

## I

importar una tabla de referencia  
  tablas de referencia [20](#)

## M

modelos de clasificador  
  en conjuntos de contenido [37](#)  
  reglas y directrices [41](#)  
modelos probabilísticos  
  en conjuntos de contenido [37](#)  
  reglas y directrices [41](#)

## P

privilegios  
  Servicio de administración de contenido [12](#)

## S

Servicio de administración de contenido  
  almacén de datos de referencia [11](#)  
  privilegios de la tabla de referencia [12](#)

## T

tablas de referencia  
  actualizar en la Herramienta del analista [27](#)  
  administradas y no administradas [11](#)  
  administrar columnas [23](#)  
  administrar filas [24](#)  
  almacén de datos de referencia [11](#)  
  buscar y reemplazar valores en la Herramienta del analista [25](#)  
  control de versiones [13](#), [22](#), [30](#)  
  crear una tabla de referencia a partir de columnas de perfil [17](#)  
  crear una tabla de referencia a partir de patrones de columna [18](#)  
  crear una tabla de referencia manualmente [16](#)  
  en análisis basado en patrones [12](#)  
  exportar una tabla de referencia [25](#)  
  importar una tabla de referencia [20](#)  
  Información general de Developer tool [31](#)  
  privilegios [12](#)  
  propiedades en la herramienta del analista [15](#)  
  propiedades en la herramienta del desarrollador [32](#)  
  Resumen de herramienta Analyst [14](#)

tablas de referencia (*continuado*)

Servicio de administración de contenido [11](#)

tablas de referencia administradas [11](#)

tablas de referencia no administradas [11](#)

visualizar tablas de traza de auditoría [27](#)

tablas de referencia administradas [11](#)

tablas de referencia no administradas

definición [11](#)

habilitar y deshabilitar las ediciones [26](#)

tablas de referencia no administradas (*continuado*)

sincronización con el repositorio de modelos [12](#)

## V

visualizar eventos de tabla de auditoría

tablas de referencia [27](#)