



Informatica®
10.4.0

Guía de integración de Data Quality para PowerCenter

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

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

Informatica, PowerCenter 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.

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

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

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

Fecha de publicación: 2020-02-07

Tabla de contenido

Prefacio	4
Recursos de Informatica	4
Informatica Network.	4
Base de conocimiento de Informatica.	4
Documentación de Informatica	5
Matrices de disponibilidad de producto de Informatica.	5
Informatica Velocity.	5
Catálogo de soluciones de Informatica.	5
Servicio internacional de atención al cliente de Informatica.	5
 Capítulo 1: Integración de Informatica Data Quality.....	6
Resumen de integración de Data Quality.	6
Características y funciones de Data Quality.	7
Programas de instalación de integración.	8
Exportación de objetos en PowerCenter.	8
Instalación de datos de referencia.	8
Reglas y directrices relativas a la integración de productos.	9
Comportamiento de las transformaciones de Data Quality en PowerCenter.	9
 Capítulo 2: Instalación.....	11
Visión general del proceso de instalación.	11
Requisitos del sistema.	11
Antes de la instalación.	12
Instalación del complemento Servidor.	12
Instalación en Windows.	12
Instalación en UNIX.	13
Instalación del complemento Cliente.	13
Registro del complemento con PowerCenter.	14
 Capítulo 3: Requisitos de los datos de referencia.....	15
Visión general de los datos de referencia.	15
Requisitos de los datos de tablas de referencia.	16
Requisitos de los datos de referencia de direcciones.	17
Requisitos de los datos de llenado de identidad.	18
Requisitos del modelo probabilístico y del modelo clasificador.	19
Ubicaciones de archivo del modelo probabilístico y del modelo clasificador en PowerCenter.	20
Asignación de memoria del servicio de integración de PowerCenter.	21
 Índice.....	22

Prefacio

Lea la *Guía de usuario de integración de Informatica Data Quality para PowerCenter®* para obtener directrices sobre cómo configurar y usar objetos de calidad de datos con PowerCenter. Puede exportar asignaciones con transformaciones de calidad de datos, junto con los archivos y objetos de datos de referencia, desde un entorno de Informatica Data Quality a un entorno de PowerCenter. En la guía se indican los requisitos previos correspondientes a los objetos de calidad de datos y, asimismo, se describen los pasos necesarios para instalar el componente de integración en versiones de productos compatibles.

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
- Ver 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 pares

Base de conocimiento de Informatica

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

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

Documentación de Informatica

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

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

Matrices de disponibilidad de producto de Informatica

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

Informatica Velocity

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

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

Catálogo de soluciones de Informatica

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

Servicio internacional de atención al cliente de Informatica

Puede ponerse en contacto con un centro de atención global por teléfono o a través del 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 encontrar recursos de soporte en línea en Informatica Network, visite <https://network.informatica.com> y seleccione la opción eSupport.

CAPÍTULO 1

Integración de Informatica Data Quality

Este capítulo incluye los siguientes temas:

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

Resumen de integración de Data Quality

Los usuarios de Informatica pueden crear y ejecutar asignaciones en Informatica Data Quality. Los usuarios guardan las asignaciones en el repositorio de modelos. Los usuarios pueden exportar las asignaciones y cualquier objeto de datos de referencia relacionado o cualquier archivo que se vaya a usar en PowerCenter.

Como usuario de PowerCenter, puede importar las asignaciones en el repositorio de PowerCenter y ejecutar las asignaciones en las sesiones de PowerCenter. Utilice las asignaciones para analizar y mejorar el contenido y la estructura de los datos.

Puede utilizar las asignaciones que los usuarios han creado en Data Quality 10.4 con las siguientes versiones de PowerCenter:

- PowerCenter 10.4.0.
- PowerCenter 10.2 con la última revisión instalada.
- PowerCenter 10.1.1 con la última revisión instalada.
- PowerCenter 10.1.

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

Características y funciones de Data Quality

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

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

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

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

Programas de instalación de integración

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

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

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

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

Exportación de objetos en PowerCenter

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

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

Instalación de datos de referencia

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

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

El proceso de exportación no incluye archivos de datos de referencia de direcciones ni archivos de llenado de identidad. Si el mapplet o la asignación que exporta lee archivos de datos de referencia de dirección o archivos de datos de llenado de identidad, instale los archivos en el equipo del servicio de integración de PowerCenter.

Reglas y directrices relativas a la integración de productos

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

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

- No instale el complemento de integración si la versión de producto del repositorio de PowerCenter es la misma que la versión de producto del repositorio de modelos.
- No se pueden ejecutar dos complementos de integración de Data Quality simultáneamente en un solo equipo de PowerCenter. PowerCenter utiliza el complemento instalado más recientemente. Por ejemplo, si instala el complemento de integración de Data Quality 10.2 en un equipo con PowerCenter 10.1.1, no podrá seguir importando y ejecutando objetos de calidad de datos procedentes de un Repositorio de modelos de la versión 10.1.1. Puede seguir ejecutando cualquier objeto de calidad de datos que se haya importado en el Repositorio de PowerCenter antes de instalar el complemento 10.2.
- No se pueden editar mapplets o asignaciones de calidad de datos en un repositorio de PowerCenter.

Comportamiento de las transformaciones de Data Quality en PowerCenter

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

Conversión a mapplets

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

- Consolidación
- Excepción
- Generador de claves
- Coincidencia

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

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

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

Consideraciones sobre las transformaciones de decisión

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

asignación o un mapplet con una transformación de decisión que utilice un parámetro del sistema, reemplace el parámetro con un valor adecuado.

Consideraciones sobre la transformación de coincidencia

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

Transformaciones multiestrategia

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

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

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

CAPÍTULO 2

Instalación

Este capítulo incluye los siguientes temas:

- [Visión general del proceso de instalación, 11](#)
- [Requisitos del sistema, 11](#)
- [Antes de la instalación, 12](#)
- [Instalación del complemento Servidor, 12](#)
- [Instalación del complemento Cliente, 13](#)
- [Registro del complemento con PowerCenter, 14](#)

Visión general del proceso de instalación

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

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

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

Requisitos del sistema

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

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

Antes de la instalación

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

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

El programa de instalación escribe los archivos siguientes:

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

El programa de instalación añade los archivos al directorio siguiente:

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

Requisitos previos de los datos de referencia de direcciones

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

Instalación del complemento Servidor

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

Instalación en Windows

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

Cierre todas las aplicaciones antes de proceder con la instalación.

1. Inicie sesión en el equipo con una cuenta de usuario del sistema.
2. Extraiga el archivo Servidor de integración para Windows.
3. Acceda al directorio en el que ha extraído los archivos y ejecute `install.exe`.
4. Seleccione un idioma y haga clic en **Aceptar**.
5. En la pantalla de **Bienvenida**, haga clic en **Siguiente**.
6. Compruebe que no hay ningún servicio de PowerCenter ejecutándose y haga clic en **Aceptar**.
7. Especifique la ruta de acceso al directorio de instalación de PowerCenter. El programa de instalación asigna una ruta predeterminada a este directorio. Para seleccionar una ruta distinta, haga clic en **Elegir**.
8. Haga clic en **Siguiente**.
9. Revise el resumen previo a la instalación. Haga clic en **Anterior** para solucionar cualquier posible error.
10. Haga clic en **Instalar**.

11. Revise el resumen posterior a la instalación.

Para obtener más información sobre las tareas del proceso de instalación y ver las propiedades de la configuración de los componentes instalados, consulte los archivos de registro de la instalación.

12. Haga clic en **Listo**.

Instalación en UNIX

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

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

Instalación del complemento Cliente

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

1. Cierre el resto de aplicaciones.
2. Extraiga el archivo Cliente de integración.
3. Abra el directorio en el que ha extraído los archivos.
4. Ejecute `install.exe`.
5. En la pantalla de **Bienvenida**, haga clic en **Siguiente**.
El programa de instalación muestra un mensaje solicitándole que cierre las aplicaciones cliente de PowerCenter antes de continuar con la instalación.
6. Compruebe que no hay ninguna aplicación cliente de PowerCenter ejecutándose y haga clic en **Aceptar**.
7. Especifique la ruta de acceso al directorio Cliente de PowerCenter. El programa de instalación asigna una ruta predeterminada a este directorio. Para seleccionar una ruta distinta, haga clic en **Elegir**.
8. Haga clic en **Siguiente**.
9. Revise el resumen previo a la instalación. Haga clic en **Anterior** para solucionar cualquier posible error.

10. Haga clic en **Instalar**.
11. Revise el resumen posterior a la instalación.
Para obtener más información sobre las tareas del proceso de instalación y ver las propiedades de la configuración de los componentes instalados, consulte los archivos de registro de la instalación.
12. Haga clic en **Listo**.

Registro del complemento con PowerCenter

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

El programa de instalación del servicio de integración escribe el archivo de complemento

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

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

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

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

CAPÍTULO 3

Requisitos de los datos de referencia

Este capítulo incluye los siguientes temas:

- [Visión general de los datos de referencia, 15](#)
- [Requisitos de los datos de tablas de referencia, 16](#)
- [Requisitos de los datos de referencia de direcciones, 17](#)
- [Requisitos de los datos de llenado de identidad, 18](#)
- [Requisitos del modelo probabilístico y del modelo clasificador, 19](#)

Visión general de los datos de referencia

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

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

Archivos de datos de referencia

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

Archivos de datos de referencia de direcciones

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

Archivos de llenado de identidad

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

Conjuntos de contenido

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

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

Requisitos de los datos de tablas de referencia

Antes de ejecutar una asignación que lea las tablas de referencia en PowerCenter, copie los datos de la columna para las tablas en la máquina de los servicios de PowerCenter.

Al exportar la asignación desde el repositorio de modelos, la operación de exportación escribe los datos de columna de la tabla de referencia en uno o más archivos. La operación de exportación agrega los archivos a un archivo comprimido en una ubicación que especifique. Extraiga el archivo comprimido y busque los archivos de datos de columna. Los nombres de archivo finalizan en `.dic`. Cópielos en el equipo de los servicios de PowerCenter.

Las ubicaciones de los archivos de datos de referencia en la estructura de directorios de PowerCenter deben corresponder a las ubicaciones de la tabla de referencia en el repositorio de modelos.

Use el siguiente formato de directorio para determinar las ubicaciones correctas para los archivos de datos de la tabla de referencia:

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

Reglas y directrices para la exportación de datos de la tabla de referencia

Considere las siguientes reglas y directrices al exportar una asignación que lea los datos de la tabla de referencia:

- Para exportar la asignación desde el repositorio de modelos, elija la opción **Informática > PowerCenter** en el asistente de exportación. Al elegir la opción **PowerCenter**, la operación de exportación escribe en el sistema de archivos tanto los datos de la tabla de referencia administrados como los datos de la tabla de referencia no administrados.
- La operación de exportación especifica la primera columna en cada tabla de referencia como la columna válida.
- PowerCenter lee archivos de datos de referencia en formato UTF-8.

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

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

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

Requisitos de los datos de referencia de direcciones

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

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

Tipo de datos de referencia	Uso
Lote	Operaciones generales de validación de direcciones.
CAMEO	Se usa para añadir datos demográficos de clientes a los registros de direcciones.
Certificado	Se usa para comprobar que los registros de direcciones cumplen los estándares de certificación que define un proveedor de servicios de correo. Los siguientes países definen estándares de certificación: <ul style="list-style-type: none">- Australia. Certifica el correo de acuerdo con el estándar del sistema de aprobación de coincidencia de direcciones (AMAS).- Canadá. Certifica el correo de acuerdo con el estándar del programa evaluación y reconocimiento de software (SERP).- Francia. Certifica el correo de acuerdo con el estándar del servicio de administración de direcciones nacional (SNA).- Nueva Zelanda. Certifica el correo de acuerdo con el estándar SendRight.- Estados Unidos. Certifica el correo de acuerdo con el estándar del sistema de soporte de precisión de codificación (CASS).
Lista de sugerencias	Se usa para encontrar versiones válidas alternativas de un registro de dirección incompleto. Use los datos de la lista de sugerencias cuando configure una asignación de validación de direcciones para procesar registros de direcciones únicos en tiempo real.
Geocode	Se usa para añadir coordenadas de latitud y de longitud a los registros de direcciones.
Suplementario	Se usa para identificar el área geográfica o el área de población que contiene una dirección. La transformación del validador de direcciones puede añadir datos suplementarios a los registros de direcciones de los siguientes países: <ul style="list-style-type: none">- Japón. Identifica el distrito urbano que contiene la dirección.- Serbia. Identifica el código postal a nivel de calle para la dirección.- Reino Unido. Añade un identificador de punto de entrega para la dirección.- Estados Unidos. Identifica el centro de la población que contiene la dirección.

El servicio de integración de PowerCenter lee la ruta de acceso a los datos de referencia de direcciones en el archivo de configuración AD50.cfg.

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

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

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

Ubicación de datos de referencia

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

Datos de claves de licencia

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

Valores de carga previa

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

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

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

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

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

Requisitos de los datos de llenado de identidad

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

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

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

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

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

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

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

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

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

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

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

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

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

```
./
```

Propiedades de IDQTx.cfg

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

IdentityCacheDir

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

La propiedad tiene el siguiente valor predeterminado:

```
./identityCache
```

IdentityIndexDir

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

La propiedad tiene el siguiente valor predeterminado:

```
./identityIndex
```

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

Requisitos del modelo probabilístico y del modelo clasificador

Los modelos probabilístico y clasificador son archivos de datos de referencia que identifican los tipos de información de una cadena de datos.

Antes de ejecutar una asignación que lea un archivo de modelo clasificador o modelo probabilístico, realice las siguientes tareas:

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

Ubicaciones de archivo del modelo probabilístico y del modelo clasificador en PowerCenter

Antes de ejecutar una asignación que lea un modelo probabilístico o un modelo clasificador en PowerCenter, copie el archivo del modelo en el equipo de servicios de PowerCenter.

Al exportar la asignación desde el repositorio de modelos, la operación de exportación agrega cada modelo a un archivo comprimido en una ubicación que especifique. Extraiga el archivo comprimido y busque los archivos del modelo que la asignación lee. Los nombres de archivos de datos del modelo probabilístico y del modelo clasificador finalizan en `.ner` y `.classifier` respectivamente. Copie los archivos en un directorio que el servicio de integración de PowerCenter pueda leer.

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

```
[PowerCenter_Installation]/server/bin
```

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

La siguiente tabla describe los archivos de propiedad:

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

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

```
[PowerCenter_Installation]/server/bin
```

Reglas y directrices para exportación del modelo probabilístico y modelo clasificador

Considere las siguientes reglas y directrices al exportar una asignación que lea los datos de la tabla de referencia:

- Al exportar una asignación que lea datos del modelo probabilístico o del modelo clasificador para usar en PowerCenter, seleccione la opción **Informática > Archivo de metadatos de objeto de exportación** en el asistente de exportación para crear el archivo comprimido.

La operación de exportación también exporta la asignación como un archivo XML. Puede descartar el archivo XML.

- Seleccione la opción para exportar modelos entrenados solamente. Un modelo probabilístico o clasificador entrenado contiene los datos lógicos que la asignación lee y no contiene los datos de origen para el modelo. Copie los archivos del modelo entrenado en el equipo de los servicios de PowerCenter.
- Exporte la asignación del repositorio de modelos en el repositorio de PowerCenter en una operación independiente. Para exportar la asignación del repositorio de modelos al repositorio de PowerCenter, seleccione la opción **Informática > PowerCenter**.

Al seleccionar la opción **PowerCenter**, la operación de exportación puede copiar los datos de referencia desde tablas de referencia administradas y tablas de referencia no administradas.

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

[PowerCenter_Installation]/server/bin

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

Asignación de memoria del servicio de integración de PowerCenter

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

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

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

INDICE

I
Informatica Data Quality [6](#)

Integración de Data Quality [6](#)