



Informatica® Data Engineering Integration  
10.4.1

# Guía de ingesta masiva

Informatica Data Engineering Integration Guía de ingesta masiva

10.4.1

Junio 2020

© Copyright Informatica LLC 2018, 2020

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

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

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

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](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: 2020-08-28

# Tabla de contenido

<b>Prefacio .....</b>	<b>6</b>
Recursos de Informatica .....	6
Informatica Network. ....	6
Base de conocimiento de Informatica. ....	6
Documentación de Informatica. ....	6
Matrices de disponibilidad de producto de Informatica. ....	7
Informatica Velocity. ....	7
Catálogo de soluciones de Informatica. ....	7
Servicio internacional de atención al cliente de Informatica. ....	7
<b>Capítulo 1: Introducción a la ingesta masiva.....</b>	<b>8</b>
Resumen. ....	8
Caso de uso. ....	9
Arquitectura. ....	12
Clientes y herramientas. ....	12
servicios de aplicación. ....	13
Repositorios. ....	13
Conexiones. ....	13
Proceso. ....	14
Herramienta de ingesta masiva. ....	15
Iniciar sesión en la herramienta de ingesta masiva. ....	16
Interfaz del usuario de la herramienta de ingesta masiva. ....	16
<b>Capítulo 2: Preparar.....</b>	<b>19</b>
Resumen de preparación. ....	19
Creación de un servicio de ingesta masiva. ....	19
Configurar la supervisión. ....	21
Paso 1. Definir la configuración de supervisión. ....	22
Paso 2. Configurar informes y estadísticas. ....	23
Asignar la función de administrador. ....	24
<b>Capítulo 3: Crear.....</b>	<b>25</b>
Resumen de creación. ....	25
Datos incrementales. ....	26
Definición. ....	26
Origen. ....	27
Claves incrementales. ....	29
Destino. ....	31
Destino de Hive. ....	31
Destino de HDFS. ....	34

Opciones de carga incremental. . . . .	36
Transformaciones masivas. . . . .	38
Filtro Cláusulas. . . . .	41
Expresiones regulares. . . . .	42
Reemplazar criterios. . . . .	42
Reemplazo de transformación. . . . .	44
Opciones de Hive. . . . .	45
Crear una especificación de ingesta masiva. . . . .	47
Configurar la definición. . . . .	47
Configurar el origen. . . . .	47
Configurar un destino de Hive. . . . .	48
Configurar un destino de HDFS. . . . .	48
Configurar transformaciones masivas. . . . .	48
Configurar el reemplazo de transformación. . . . .	49
<b>Capítulo 4: Implementar. . . . .</b>	<b>50</b>
Resumen de implementación. . . . .	50
Implementar una especificación de ingesta masiva. . . . .	51
Volver a implementar una especificación de ingesta masiva. . . . .	52
Implementar en un archivo de almacenamiento de la aplicación. . . . .	52
Migrar una especificación de ingesta masiva. . . . .	53
<b>Capítulo 5: Run. . . . .</b>	<b>55</b>
Resumen de supervisión. . . . .	55
Tipos de carga. . . . .	55
Carga completa. . . . .	56
Carga incremental. . . . .	56
Ver instancias de ejecución. . . . .	57
Estado de instancia de ejecución. . . . .	57
Ejecutar una especificación de ingesta masiva. . . . .	58
<b>Capítulo 6: Supervisor. . . . .</b>	<b>60</b>
Resumen de supervisión. . . . .	60
Supervisar la herramienta de ingesta masiva. . . . .	60
Resumen Página. . . . .	61
Página Estadísticas de ejecución. . . . .	64
Supervisar estadísticas de ejecución. . . . .	67
Filtrar objetos de ingesta por estado. . . . .	69
Reiniciar trabajos de ingesta. . . . .	70
Supervisar la Herramienta del administrador. . . . .	72
Supervisar la aplicación y los trabajos de asignación de ingesta. . . . .	72
Cancelar trabajos de ingesta. . . . .	75
Anular trabajos de ingesta. . . . .	75

Solución de problemas de trabajos de ingesta masiva. . . . .	76
--	----

**Apéndice A: Referencia de comando infacmd mi. . . . . 77**

abortRun. . . . .	77
createService. . . . .	78
deploySpec. . . . .	82
exportSpec. . . . .	83
extendedRunStats. . . . .	84
getSpecRunStats. . . . .	86
listSpecRuns. . . . .	87
listSpecs. . . . .	88
restartMapping. . . . .	89
runSpec. . . . .	90

**Índice. . . . . 93**

# Prefacio

Utilice *Informatica® Guía de ingesta masiva de Data Engineering Integration* para conocer la interfaz de usuario de la herramienta de ingesta masiva y desplazarse por ella. Sepa cómo crear, implementar, ejecutar y supervisar trabajos de ingesta masiva.

## 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](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 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

## Introducción a la ingesta masiva

Este capítulo incluye los siguientes temas:

- [Resumen, 8](#)
- [Caso de uso, 9](#)
- [Arquitectura, 12](#)
- [Proceso, 14](#)
- [Herramienta de ingesta masiva, 15](#)

### Resumen

Utilice Informatica Mass Ingestion (la herramienta de ingesta masiva) para ingerir grandes cantidades de datos de una base de datos relacional en un destino de HDFS o Hive.

La herramienta de ingesta masiva simplifica el proceso de ingesta de datos, ya que proporciona un asistente que se puede utilizar para crear una especificación de ingesta masiva. Una especificación de ingesta masiva es una configuración que puede diseñar para especificar los datos que desea ingerir y la manera en que quiere hacerlo.

El asistente le indica los pasos que puede utilizar para configurar cada parte de la especificación. Esto incluye el origen relacional y el destino de Hive o HDFS, así como cualquier parámetro que quiera configurar en el origen, como un parámetro para filtrar determinadas columnas o enmascarar los datos para proteger la información privada.

Cuando la especificación de ingesta masiva se ejecuta, la herramienta de ingesta masiva utiliza Data Engineering Integration para ejecutar el trabajo de ingesta en un clúster de Hadoop. La especificación acaba con la necesidad de tener que crear y ejecutar asignaciones de forma manual, y es capaz de ingerir todos los datos de una sola vez. A medida que los esquemas de la base de datos relacional vayan evolucionando, la especificación puede dar cabida e ingerir solo los datos incrementales.

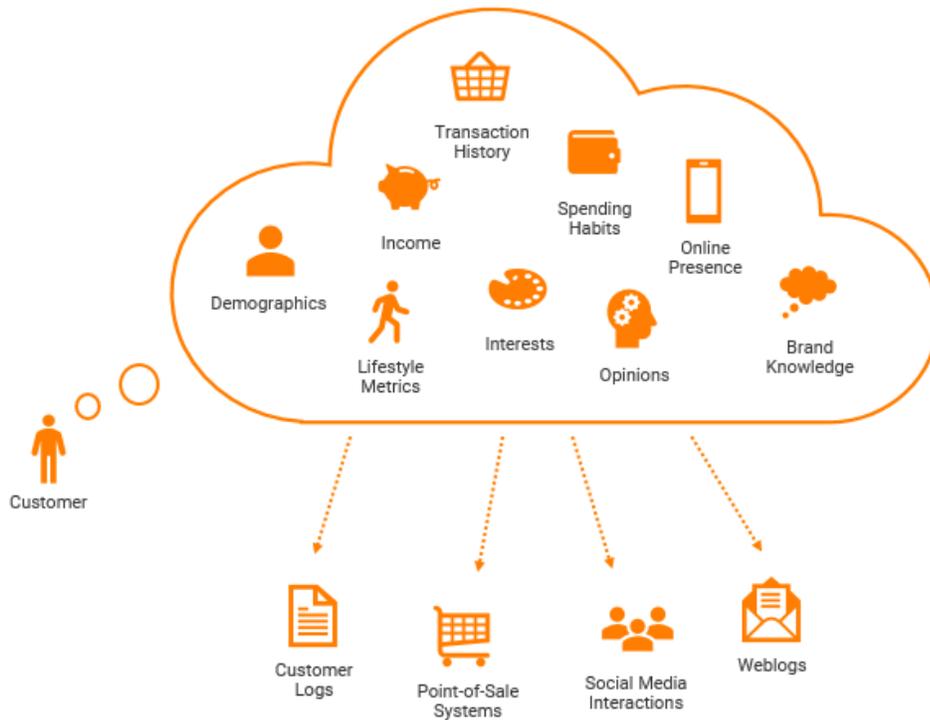
# Caso de uso

Trabaja en un grupo de TI de un banco comercial. Su equipo está a cargo de un nuevo proyecto para personalizar el programa de premios que se ofrece a los clientes que abren cuentas corrientes y de ahorro en el banco.

Su idea es recopilar y analizar datos sobre sus clientes para saber en qué tipos de premios están interesados. Por ejemplo, un cliente puede estar interesado en ahorrar dinero en productos de alimentación, mientras que otro puede estarlo en paquetes de viajes.

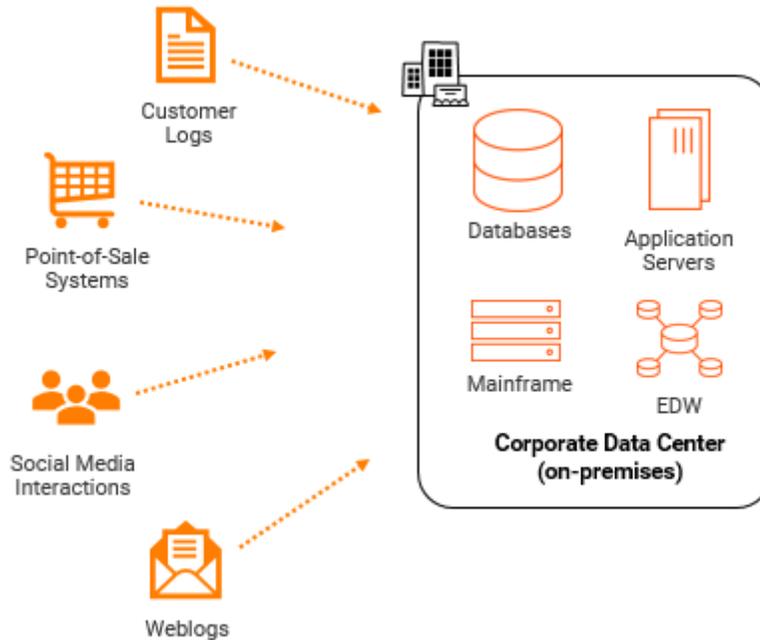
Así pues, recopila datos de los clientes como datos demográficos, indicadores de estilo de vida, ingresos, historial de transacciones, hábitos de consumo, presencia en línea, intereses, opiniones y conocimiento de marca. Los datos se obtienen a través de distintos medios: registros de clientes en los archivos del banco, sistemas de puntos de venta en compañías asociadas con el banco, interacciones en las redes sociales y blogs de clientes en Internet.

En la siguiente imagen se muestran los tipos de datos que recopila y el medio que usa para recopilarlos:



Cuando los datos se recopilan, se almacenan en el centro de datos corporativo del banco, compuesto por varias bases de datos relacionales.

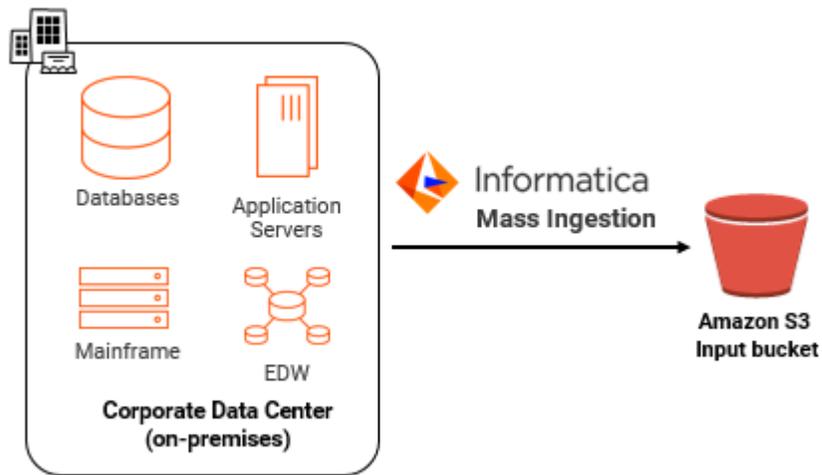
En la siguiente imagen se muestra la forma en que los datos se pueden almacenar:



Antes de que su analista de datos pueda comenzar a trabajar con los datos, es necesario ingerirlos desde las bases de datos relacionales a cubos de Amazon S3, pero no dispone ni del tiempo ni de los recursos necesarios para ingerir cantidades de datos tan grandes. Tendrá que desarrollar varias asignaciones y conjuntos de parámetros para ingerir los datos y asegurarse de que se ingieren correctamente. Además, tiene que procurar no ingerir información confidencial del cliente, como los números de tarjetas de crédito. Aparte de todo esto, tiene que mantener las asignaciones cuando los esquemas relacionales cambien.

En lugar de crear y ejecutar asignaciones de forma manual, puede utilizar la herramienta de ingesta masiva para crear una especificación de ingesta masiva que ingiera todos los datos de una sola vez. Solamente tiene que especificar el origen, el destino y cualquier parámetro que quiera aplicar a las tablas de origen. Cuando implemente y ejecute la especificación, el motor de Spark ingiere todos los datos en Amazon S3.

En la siguiente imagen se muestra la forma en que la ingesta masiva puede enlazar los datos que el banco almacena en sus bases de datos relacionales con los cubos de Amazon S3:

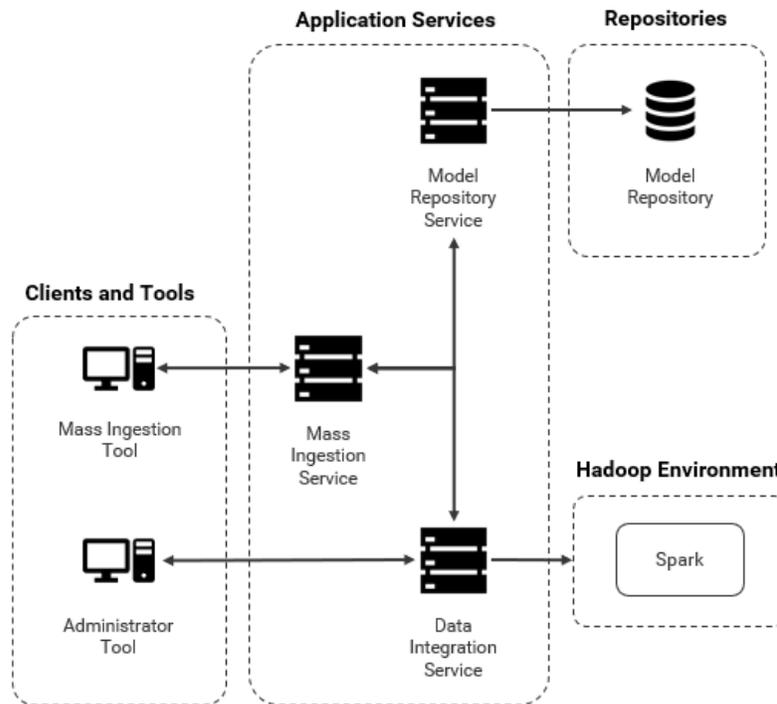


La ingesta masiva supone un ahorro de tiempo y de recursos sustancial, y su analista de datos tendrá más tiempo para analizar datos y desarrollar un nuevo sistema para el programa de premios del banco.

# Arquitectura

La arquitectura de ingesta masiva incluye los componentes que crean, implementan, ejecutan y supervisan una especificación de ingesta masiva.

La siguiente imagen muestra los componentes de ingesta masiva:



## Clientes y herramientas

Los componentes de ingesta masiva constan de los siguientes clientes y herramientas:

### Informatica Administrator

Herramienta del administrador. Cliente donde se puede crear y configurar un servicio de ingesta masiva. El cliente también se puede utilizar para supervisar los trabajos de asignación de ingesta específicos que el motor de Spark ejecuta para ingerir los datos en una especificación de ingesta masiva.

### Informatica Mass Ingestion

Herramienta de ingesta masiva. Aplicación web donde se diseñan e implementan especificaciones de ingesta masiva. Utilice la aplicación web para configurar, implementar, ejecutar y supervisar una especificación de ingesta masiva.

## servicios de aplicación

Los componentes de ingesta masiva incluyen los siguientes servicios de aplicación:

### Servicio de ingesta masiva

Administra y valida las especificaciones de ingesta masiva que se crean en la herramienta de ingesta masiva. El servicio de ingesta masiva implementa especificaciones y programa su ejecución. Cuando una especificación se ejecuta, el servicio de ingesta masiva genera estadísticas del trabajo de ingesta.

### Servicio de repositorio de modelos

Administra el repositorio de modelos. El servicio de repositorio de modelos almacena las especificaciones de ingesta masiva que se crean en el Repositorio de modelos.

**Nota:** La ingesta masiva no es compatible con un servicio de repositorio de modelos que se integra con un sistema de control de versiones.

### Servicio de integración de datos

Administra las especificaciones de ingesta masiva implementadas. El servicio de integración de datos se conecta con el entorno de Hadoop para permitir que el motor de Spark ingiera los datos que hay configurados en una especificación de ingesta masiva.

## Repositorios

Los componentes de ingesta masiva incluyen los siguientes repositorios:

### Repositorio de modelos

Repositorio que almacena los metadatos de la especificación de ingesta masiva que cree en la herramienta de ingesta masiva.

El Repositorio de modelos almacena las especificaciones de ingesta masiva en el proyecto `INFA_MI_PROJECT`. Si almacena una especificación en un proyecto diferente, el servicio de ingesta masiva la pasará por alto. Para que el servicio de ingesta masiva lea todas las especificaciones, no mueva la especificación a un proyecto diferente.

**Nota:** El proyecto no se podrá ver en Developer tool, ni tampoco las especificaciones almacenadas en el proyecto. Utilice la línea de comandos para ver la lista de especificaciones almacenadas en el proyecto.

## Conexiones

Cuando se ejecutan trabajos de ingesta masiva, los componentes de ingesta masiva utilizan las siguientes conexiones:

### JDBC

Una conexión de JDBC tiene acceso a las tablas de la base de datos relacional del trabajo de ingesta masiva.

La conexión de origen que se usa en un trabajo de ingesta masiva debe ser una conexión de JDBC. Por ejemplo, para acceder a un esquema de Oracle, debe configurar una conexión de JDBC que use un controlador de Oracle para conectarse a una base de datos de Oracle. No se puede utilizar una conexión de Oracle.

### Sqoop

Cuando se configura una conexión de JDBC con argumentos de Sqoop, las tareas se reparten entre JDBC y Sqoop. JDBC se utiliza para importar metadatos desde una base de datos relacional, mientras que Sqoop lee los datos.

Si utiliza una carga incremental para ingerir datos utilizando una conexión de Sqoop, el servicio de ingesta masiva hace uso del modo de importación incremental de Sqoop. Cuando el servicio de ingesta masiva configura el filtro de datos incrementales, el filtro se inserta en el origen de Sqoop.

Si utiliza una conexión de Sqoop, tenga en cuenta las siguientes limitaciones:

- Una tabla de origen no se puede ingerir con una conexión de Sqoop si la tabla contiene caracteres especiales en los metadatos de tabla.
- No se pueden ingerir tipos de datos Blob con una conexión de Sqoop.

#### **Hadoop**

Una conexión de Hadoop permite que el servicio de integración de datos inserte los trabajos de ingesta masiva en el entorno de Hadoop donde los trabajos se ejecutan en el motor de Spark.

#### **Hive**

Una conexión de Hive accede a los datos de Hive y permite que el trabajo de ingesta masiva escriba datos de Hive en un destino de Hive.

#### **HDFS**

Una conexión de HDFS accede a los datos en el clúster de Hadoop para permitir que un trabajo de ingesta masiva escriba datos de archivo sin formato en el clúster.

Para obtener más información sobre las propiedades de conexión, consulte el apéndice "Conexiones" en la guía *Informatica Guía del usuario de Data Engineering Integration*.

## Proceso

El proceso de ingesta masiva incorpora los componentes de la arquitectura de ingesta masiva que crean, implementan, ejecutan y supervisan una especificación de ingesta masiva.

El proceso de ingesta masiva incluye las siguientes tareas:

#### **Crear**

Cree una especificación de ingesta masiva en la herramienta de ingesta masiva. El servicio de ingesta masiva valida y se conecta al servicio de repositorio de modelos para almacenar la especificación en un Repositorio de modelos.

Tras crear la especificación, esta se puede migrar entre Repositorios de modelos.

#### **Implementar**

Implemente la especificación de ingesta masiva en un servicio de integración de datos y especifique una conexión de Hadoop. El servicio de ingesta masiva procesa e implementa la especificación en un servicio de integración de datos.

La especificación de ingesta masiva también se puede implementar en un archivo de almacenamiento de la aplicación para guardar la información relativa a la especificación como una aplicación. Si implementa la especificación en un archivo de almacenamiento de aplicación, puede importar la aplicación al Repositorio de modelos e implementar la aplicación en un servicio de integración de datos.

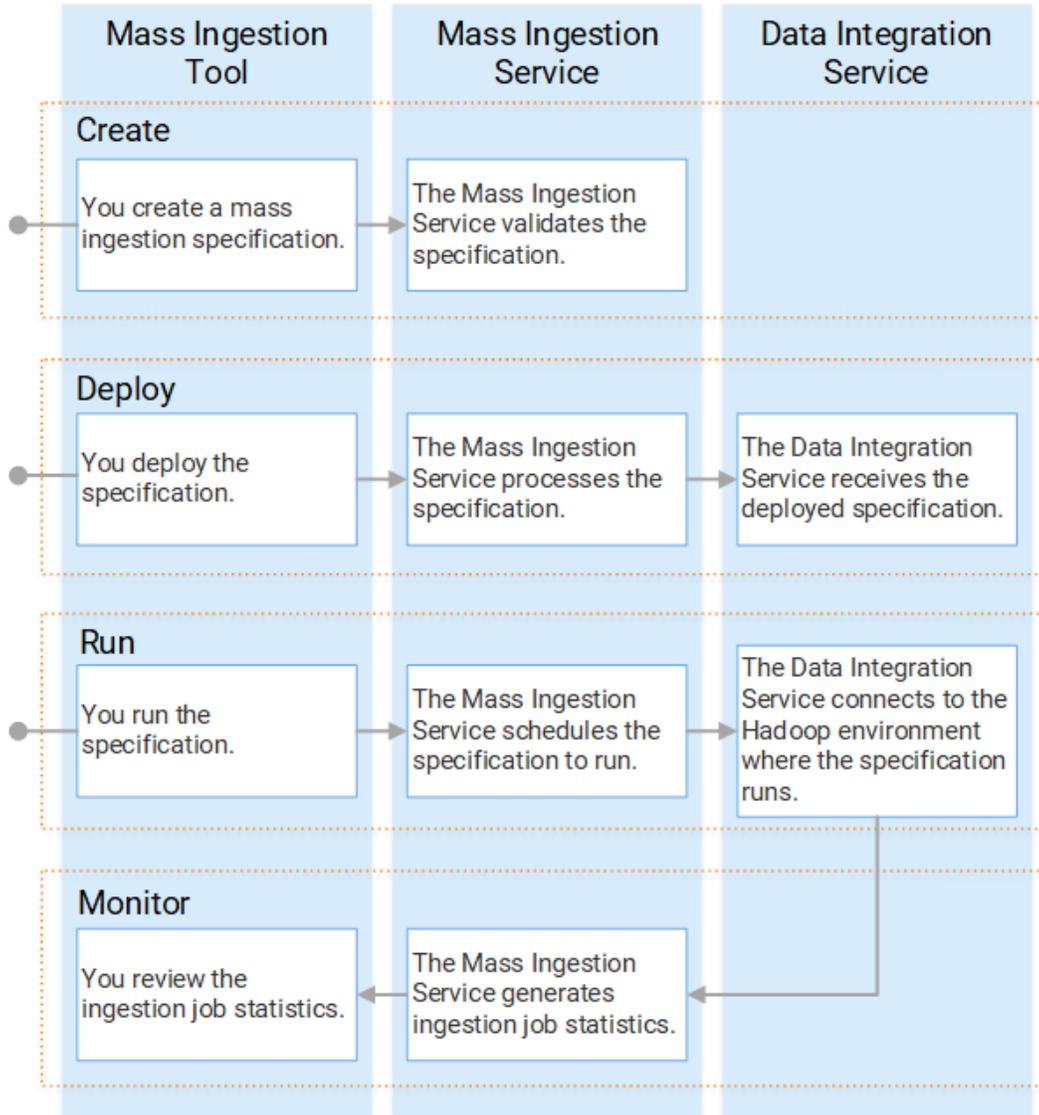
#### **Run**

Ejecute la especificación de ingesta masiva para ingerir datos en un destino de HDFS o Hive. El servicio de ingesta masiva programa la especificación que se va a ejecutar y el servicio de integración de datos inserta la especificación en el motor de Spark en el entorno de Hadoop.

## Supervisar

El servicio de ingesta masiva genera estadísticas de trabajo de ingesta. Estas estadísticas se pueden supervisar en la herramienta de ingesta masiva o en la Herramienta del administrador.

El siguiente diagrama ilustra al detalle el proceso de ingesta masiva cuando se crea, implementa, ejecuta y supervisa una especificación de ingesta masiva:



## Herramienta de ingesta masiva

La herramienta de ingesta masiva es una aplicación de Internet que proporciona una interfaz para diseñar especificaciones de ingesta masiva y ejecutar trabajos de ingesta masiva.

Cuando se diseña una especificación de ingesta masiva en la herramienta de ingesta masiva, hay que especificar la manera en la que quiere que se ingieran los datos de origen. Puede utilizar la herramienta de

ingesta masiva para implementar y ejecutar la especificación de ingesta masiva que va a llevar a cabo el trabajo de ingesta. También puede supervisar el trabajo de ingesta para revisar estadísticas de ingesta.

## Iniciar sesión en la herramienta de ingesta masiva

Después de que el administrador del sistema configure el servicio de ingesta masiva, puede iniciar sesión en la herramienta de ingesta masiva. Para obtener la URL de la herramienta de ingesta masiva, póngase en contacto con el administrador del sistema de Informática.

1. Abra un navegador web.
2. En el campo de dirección, introduzca la siguiente URL que le ha proporcionado el administrador del sistema de Informática:

```
http://<host>:<puerto>/mi/login
```

3. Si ha configurado la seguridad LDAP en la Herramienta del administrador, la página de inicio de sesión de la herramienta de ingesta masiva le pedirá que seleccione un dominio de seguridad. Seleccione el dominio de seguridad nativo o un dominio de seguridad LDAP configurado.

Si ha configurado la autenticación SAML en el dominio, se le llevará al proveedor de seguridad de terceros en cuestión para que inicie sesión.

4. Si el dominio no está configurado para la autenticación SAML, introduzca su nombre de usuario y contraseña en la página de inicio de sesión. Haga clic en **Iniciar sesión**.

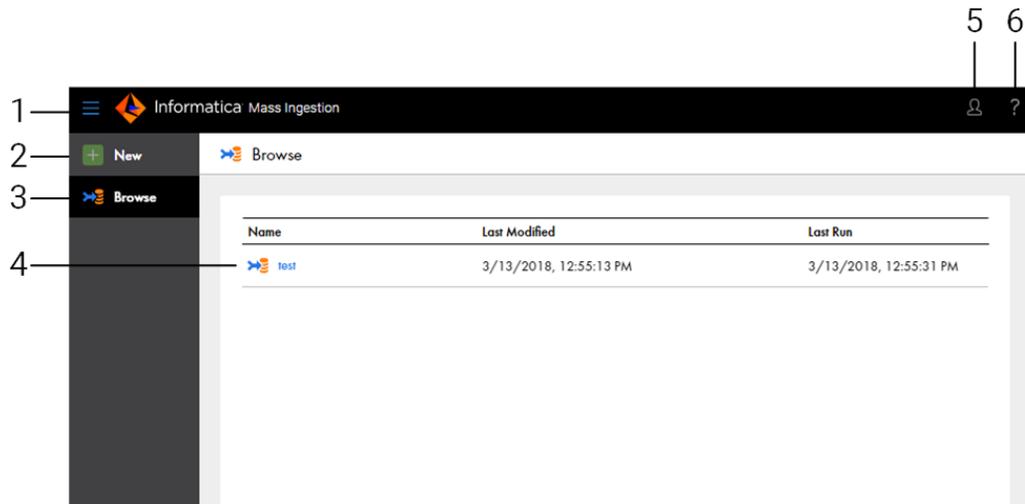
Aparece la página de inicio de la herramienta de ingesta masiva.

## Interfaz del usuario de la herramienta de ingesta masiva

Utilice la herramienta de ingesta masiva para diseñar e implementar trabajos de ingesta masiva.

Si bien se pueden crear varias especificaciones de ingesta masiva en la herramienta de ingesta masiva, solamente se puede trabajar en una especificación de ingesta masiva cada vez. Para trabajar en la herramienta de ingesta masiva, hay que acceder al entorno de trabajo de dicha herramienta.

En la siguiente imagen se muestra el entorno de trabajo de la herramienta de ingesta masiva:



1. Menú. Vea las opciones de navegación para crear una especificación de ingesta masiva, o bien examine las especificaciones de ingesta masiva.
2. Nuevo. Cree una especificación de ingesta masiva.
3. Examinar. Examine las especificaciones de ingesta masiva creadas.
4. Especificación de ingesta masiva. Revise el nombre de la especificación de ingesta masiva, así como la hora a la que se modificó por última vez y la hora a la que se ejecutó por última vez.
5. Usuario. Vea el usuario que ha iniciado sesión actualmente o que ha cerrado sesión en la herramienta de ingesta masiva.
6. Ayuda. Acceda a la ayuda de la herramienta de ingesta masiva.

## Navegación

Puede desplazarse a otras páginas desde el entorno de trabajo de la herramienta de ingesta masiva.

Puede desplazarse a las siguientes páginas:

### Examinar

Muestra las especificaciones de ingesta masiva que se hayan creado. Para ir a la página Examinar, seleccione Examinar en el menú.

### Crear

Muestra el asistente donde se puede crear una especificación de ingesta masiva. Para ir al asistente de creación, seleccione Nuevo en el menú.

### Visión general

Muestra detalles sobre una especificación de ingesta masiva. Para ir a la página Resumen, seleccione una especificación de ingesta masiva en la página Examinar.

En la página Resumen, puede acceder a las siguientes vistas:

- Vista de resumen. Revise las propiedades generales de especificación de ingesta masiva.
- Vista Implementar. Implemente, vuelva a implementar y ejecute una especificación de ingesta masiva.
- Vista Historial de ejecución. Examine las instancias de ejecución de una especificación de ingesta masiva.

### **Estadísticas de ejecución**

Muestra información detallada sobre el trabajo de ingesta de una especificación de ingesta masiva. Para ir a la página Estadísticas de ejecución, desplácese a la página Resumen. En la página Resumen, seleccione una instancia de ejecución de especificación de ingesta masiva.

En la página Estadísticas de ejecución, puede acceder a las siguientes vistas:

- Vista Propiedades del trabajo. Revise las propiedades generales de especificación de ingesta masiva de la instancia de ejecución.
- Vista Objetos de ingesta. Examine las tablas de origen que se ingieren en la especificación de ingesta masiva.
- Vista Estado de ingesta. Vea un resumen del trabajo de ingesta masiva configurado en la especificación de ingesta masiva.
- Vista Estadísticas de ingesta. Revise las estadísticas de trabajo de ingesta de una tabla de origen específica.

### **Ayuda**

Muestra la ayuda de la herramienta de ingesta masiva.

# CAPÍTULO 2

## Preparar

Este capítulo incluye los siguientes temas:

- [Resumen de preparación, 19](#)
- [Creación de un servicio de ingesta masiva, 19](#)
- [Configurar la supervisión, 21](#)
- [Asignar la función de administrador, 24](#)

## Resumen de preparación

Para poder utilizar la herramienta de ingesta masiva, un administrador debe configurar el servicio de ingesta masiva en la Herramienta del administrador.

El administrador debe realizar las siguientes tareas de configuración:

- Crear el servicio de ingesta masiva en el dominio de Informatica
- Configurar la supervisión en el servicio de ingesta masiva
- Asignar la función de administrador a un usuario o grupo para el servicio de ingesta masiva

## Creación de un servicio de ingesta masiva

Al crear un servicio de ingestión masiva, debe asociar un servicio de repositorio de modelos con un servicio de ingesta masiva. Un servicio de repositorio de modelos no se puede asociar a más de un servicio de ingesta masiva.

**Nota:** Debe crear el servicio de ingesta masiva en un dominio que utilice la autenticación nativa. Si crea el servicio de ingesta masiva en un dominio que utilice autenticación LDAP o Kerberos, no podrá iniciar sesión en la Herramienta de ingesta masiva.

1. En la Herramienta del administrador, haga clic en la ficha **Administrar**.
2. Haga clic en la vista **Servicios y nodos**.
3. En el navegador del dominio, seleccione el dominio.
4. Haga clic en **Acciones > Nuevo > Servicio de ingesta masiva**.  
Aparecerá el asistente **Nuevo servicio de ingesta masiva**.

5. En la página **Nuevo servicio de ingesta masiva: paso 1 de 3**, introduzca las siguientes propiedades:

Propiedad	Descripción
Nombre	Nombre del servicio. No se aplica la distinción entre mayúsculas y minúsculas al nombre, el cual debe ser único en el dominio. Este nombre no puede tener más de 128 caracteres ni empezar por @. Además, no puede contener espacios ni los siguientes caracteres especiales: ` ~ % ^ * + = { } \ ; : ' " / ? . , < >   ! ( ) [ ]
Descripción	Descripción del servicio. La descripción no puede tener más de 765 caracteres.
Ubicación	El dominio y la carpeta en los que se crea el servicio. Haga clic en <b>Examinar</b> para elegir una carpeta diferente. Puede mover el servicio una vez lo haya creado.
Licencia	Objeto de licencia que permite utilizar el servicio.
Nodo	Nodo en que se ejecuta este servicio.

6. Haga clic en **Siguiente**.
7. En la página **Nuevo servicio de ingesta masiva: paso 2 de 3**, introduzca las siguientes propiedades:

Propiedad	Descripción
servicio de repositorio de modelos	Servicio de repositorio de modelos para asociar con el servicio.
Nombre de usuario	El nombre de usuario que el servicio usa para acceder al servicio de repositorio de modelos. Introduzca el usuario del repositorio de modelos que ha creado.
Contraseña	La contraseña del usuario del repositorio de modelos.

8. Haga clic en **Siguiente**.  
Aparecerá la página **Nuevo servicio de ingesta masiva: paso 3 de 3**.
9. En la página **Nuevo servicio de ingesta masiva: paso 3 de 3**, introduzca las siguientes propiedades:

Propiedad	Descripción
Puerto HTTP	Número exclusivo de puerto HTTP del proceso del servicio de ingesta masiva cuando este usa el protocolo HTTP. El valor predeterminado es 9050.
Habilitar la seguridad de la capa de transporte (TLS)	Permite al protocolo de seguridad de la capa de transporte cifrar las conexiones que se establecen entre el servicio de ingesta masiva y los componentes externos. Si habilita el protocolo TLS, debe especificar un puerto HTTPS y un archivo de almacén de claves. No especifique un puerto HTTP.
Puerto HTTPS	Número exclusivo de puerto HTTPS del proceso del servicio de ingesta masiva cuando este usa el protocolo HTTPS. Si configura el número de puerto HTTPS, también debe configurar el archivo de almacén de claves que contiene las claves y los certificados necesarios.

Propiedad	Descripción
Archivo de almacén de claves	Ruta de acceso y nombre de archivo del archivo de almacén de claves que contiene las claves y los certificados que se necesitan si utiliza conexiones HTTPS para el servicio de ingesta masiva.  Para crear un archivo de almacén de claves, se usa una utilidad keytool. La utilidad keytool genera y almacena pares de claves privadas o públicas y los certificados asociados en un archivo de almacén de claves. Puede utilizar el certificado autofirmado o un certificado firmado por una autoridad de certificación.
Contraseña del almacén de claves	La contraseña del archivo de almacén de claves.

10. Para habilitar el servicio de ingesta masiva, seleccione **Habilitar servicio**.
11. Haga clic en **Finalizar**.  
El dominio crea el servicio de ingesta masiva. Si seleccionó **Habilitar servicio**, el dominio habilita el servicio de ingesta masiva.
12. En el **navegador del dominio**, seleccione el servicio de ingesta masiva.
13. Haga clic en la URL para acceder a la Herramienta de ingesta masiva.

## Configurar la supervisión

Para supervisar una especificación de ingesta masiva, debe configurar el dominio para mostrar estadísticas e informes sobre los objetos del dominio.

Al configurar la supervisión en el dominio, los servicios de integración de datos que administran las especificaciones de ingesta masiva implementada almacenan estadísticas e informes en un Repositorio de modelos. Estas estadísticas incluyen información de historial sobre las especificaciones implementadas.

Si no configura la supervisión, no podrá supervisar el estado de ingesta en la herramienta de ingesta masiva. En la Herramienta del administrador, algunas vistas de las fichas Administrar y Supervisar no tienen contenido y los trabajos de asignación de ingesta no se puede supervisar. El gráfico de flujo de trabajo está vacío, y las notificaciones desaparecen al actualizar la página.

Para configurar estadísticas e informes de supervisión, realice las tareas siguientes:

1. Definir la configuración de supervisión. Configure un repositorio de modelos para almacenar estadísticas de tiempo de ejecución para objetos que los servicios de integración de datos ejecutan.
2. Configure informes y vistas de estadísticas. Seleccione qué estadísticas aparecen en las vistas **Estadísticas e Informes**.

## Paso 1. Definir la configuración de supervisión

Defina la configuración de supervisión del dominio. Al definir la configuración de supervisión, hay que especificar un Repositorio de modelos donde almacenar las estadísticas en tiempo de ejecución relativas a la especificación de ingesta masiva implementada en el servicio de integración de datos.

Cree el contenido del repositorio de modelos antes de configurar la configuración de supervisión. Si crea el contenido después de definir la configuración de supervisión, debe reciclar el servicio de repositorio de modelos tras la creación del contenido.

1. En la herramienta del administrador, haga clic en la ficha **Administrar** > vista **Servicios y nodos**.
2. Haga clic en la vista **Configuración de supervisión**.  
Se mostrará la configuración de supervisión actual.
3. Haga clic en **Editar** para cambiar la configuración de supervisión.
4. Edite las siguientes opciones:

Opción	Descripción
Servicio de repositorio de modelos	Nombre del repositorio de modelos que almacena la información del historial. El repositorio de modelos no debe estar integrado con un sistema de control de versiones.
Nombre de usuario	Nombre de usuario para acceder al servicio de repositorio de modelos.
Contraseña	Contraseña del nombre de usuario para acceder al servicio de repositorio de modelos.
Modificar contraseña	Modifique la contraseña del servicio de repositorio de modelos.
Dominio de seguridad	Nombre del dominio de seguridad al que pertenece el usuario del repositorio de modelos.
Conservar datos históricos de resumen	Número de días que el repositorio de modelos guarda datos de promedio. Si la purga está deshabilitada, el repositorio de modelos guarda los datos indefinidamente. El valor predeterminado es 180. El valor mínimo es 0. El valor máximo es 366.
Conservar datos históricos detallados	Número de días que el repositorio de modelos guarda datos por minuto. Si la purga está deshabilitada, el repositorio de modelos guarda los datos indefinidamente. El valor predeterminado es 14. El valor mínimo es 1. El valor máximo es 14.
Purgar estadísticas cada	Intervalo, en días, en el que el servicio de repositorio de modelos purga datos anteriores a los valores configurados en las opciones de <b>conservación de datos históricos</b> . El valor predeterminado es 1 día.
Hora	Hora del día en que el servicio de repositorio de modelos purgará las estadísticas. El valor predeterminado es 1:00 a. m.
Número máximo de registros ordenables	Número máximo de registros que se pueden ordenar en la ficha <b>Supervisar</b> . Si el número de registros de la ficha <b>Supervisar</b> es mayor que este valor, solo podrá ordenar por <b>Hora de inicio</b> y <b>Hora de finalización</b> . El valor predeterminado es 3000.

Opción	Descripción
Retardo máximo para las notificaciones de actualización	Tiempo máximo, en segundos, durante el cual el servicio de integración de datos almacena en la memoria intermedia las estadísticas antes de almacenarlas en el repositorio de modelos y de mostrarlas en la ficha <b>Supervisor</b> . Si el servicio de integración de datos se cierra inesperadamente antes de que almacene las estadísticas en el repositorio de modelos, las estadísticas se perderán. El valor predeterminado es 10.
Mostrar milisegundos en el campo de fecha y hora	Incluye los milisegundos para los campos de fecha y hora en la ficha <b>Supervisión</b> .

5. Haga clic en **Aceptar**.
6. Haga clic en **Guardar**.

Para aplicar la configuración, debe reiniciar todos los servicios de integración de datos.

## Paso 2. Configurar informes y estadísticas

De forma predeterminada, las estadísticas de trabajo de ingesta de una especificación de ingesta masiva están vacías en la herramienta de ingesta masiva. Para ver las estadísticas y los informes, debe ajustar la configuración de informes y estadísticas en el dominio. Esta configuración se aplica a todos los servicios de integración de datos del dominio.

Antes de configurar las estadísticas y los informes, debe especificar un servicio de repositorio de modelos en la Configuración de supervisión y habilitarlo.

1. En la Herramienta del administrador, haga clic en la ficha **Supervisor**.
2. En el menú **Acción**, seleccione **Configuración de informes y estadísticas**.
3. En la ficha **Estadísticas**, configure los intervalos de tiempo que desee usar para las estadísticas y, a continuación, seleccione la frecuencia con la que deben actualizarse las estadísticas asignadas a cada intervalo de tiempo.
4. Seleccione el intervalo de tiempo predeterminado que aparecerá para todas las estadísticas.
5. Haga clic en la ficha **Informes**.
6. Habilite los intervalos de tiempo que desee usar para los informes y, a continuación, seleccione la frecuencia con la que deben actualizarse los informes asignados a cada intervalo de tiempo.
7. Seleccione el intervalo de tiempo predeterminado que aparecerá para todos los informes y, a continuación, haga clic en **Aceptar**.
8. Haga clic en **Seleccionar informes**.
9. Añada los informes que desee ejecutar al cuadro **Informes seleccionados**.
10. Organice los informes en el orden en que desea verlos en la ficha **Supervisor**.
11. Haga clic en **Aceptar** para cerrar la ventana **Seleccionar informes**.
12. Haga clic en **Aceptar** para aceptar la configuración y cerrar la ventana **Configuración de informes y estadísticas**.

# Asignar la función de administrador

Asigne la función de administrador a un usuario o grupo para el servicio de ingesta masiva. La función de administrador asignada al usuario o grupo también determina las tareas que ese usuario o grupo puede realizar en la herramienta de ingesta masiva.

1. En la Herramienta del administrador, haga clic en la ficha **Seguridad**.
2. En el navegador, seleccione un usuario o grupo.
3. Haga clic en la ficha **Privilegios**.
4. Haga clic en Editar.

Se abre el cuadro de diálogo **Editar roles y privilegios**.

5. Para asignar la función de administrador, expanda el servicio de ingesta masiva en la ficha **Funciones**.
6. Para conceder la función de administrador, selecciónela para asignarla al usuario o grupo para el dominio o servicio de aplicación.
7. Para revocarla, anule las funciones asignadas al usuario o grupo.
8. Haga clic en **Aceptar**.

# CAPÍTULO 3

## Crear

Este capítulo incluye los siguientes temas:

- [Resumen de creación, 25](#)
- [Definición, 26](#)
- [Origen, 27](#)
- [Destino, 31](#)
- [Transformaciones masivas, 38](#)
- [Reemplazo de transformación, 44](#)
- [Crear una especificación de ingesta masiva, 47](#)

## Resumen de creación

Cree una especificación de ingesta masiva en la herramienta de ingesta masiva.

Cuando se crea una especificación de ingesta masiva, hay que configurar el trabajo que ingiere datos desde una base de datos relacional en un destino de Hive o de HDFS.

Para crear la especificación de ingesta masiva, se usa un asistente que ayuda a definir el trabajo de ingesta. En este asistente se pueden configurar diferentes propiedades en cada una de las siguientes páginas:

1. Definición. Tipo de especificación que desea crear.
2. Origen. Datos de origen que quiere ingerir.
3. Destino. Destino de Hive o HDFS en el que quiere ingerir datos.
4. Transformaciones masivas. Parámetros que se aplican en todas las tablas de origen que se ingieren.
5. Reemplazo de transformación. Parámetros que se aplican a tablas de origen concretas que se ingieren. Estos parámetros reemplazan a los parámetros configurados como transformaciones masivas.

A medida que configura la especificación de ingesta masiva, el servicio de ingesta masiva va validando las entradas. Cuando la especificación se guarde, se almacenará en el Repositorio de modelos.

## Datos incrementales

Puede configurar una especificación de ingesta masiva para ingerir datos incrementales, que son datos que se han modificado entre distintas ejecuciones de la especificación.

Para ingerir datos incrementales, habilite la carga incremental en la especificación y configure las siguientes opciones de carga incremental:

- Clave incremental. Clave que el motor de Spark utiliza para obtener datos incrementales.
- Modo incremental. Modo que el motor de Spark utiliza para ingerir datos incrementales.

Si se ingieren datos en un destino de Hive, también se pueden propagar los cambios de esquema para dar cabida a los desfases de esquema en la base de datos relacional.

Si no habilita la carga incremental, solo podrá utilizar cargas completas para ingerir datos. Una carga completa ingiere todos los datos en el destino cada vez que la especificación se ejecuta.

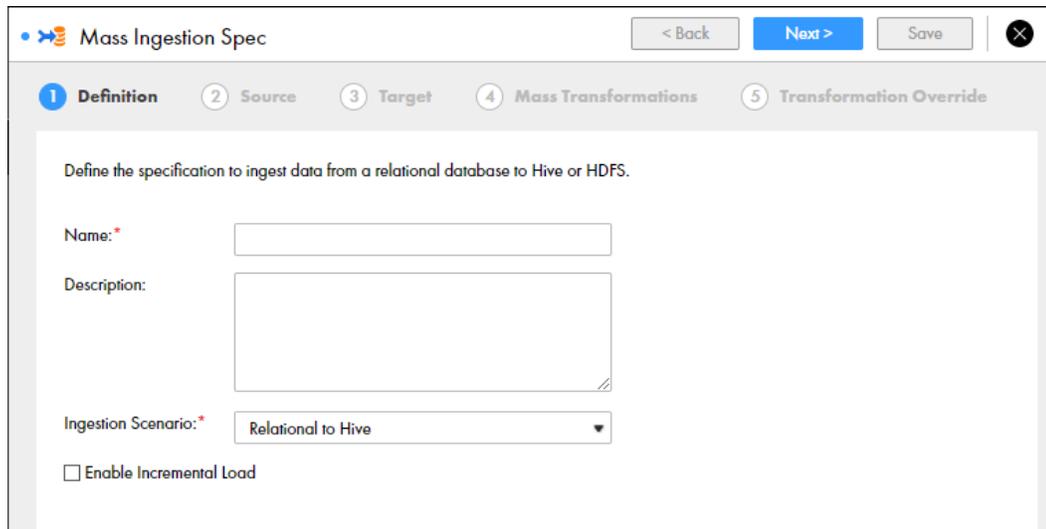
## Definición

La definición describe el tipo de especificación de ingesta masiva que desea crear.

En la definición hay que especificar un nombre y una descripción y seleccionar el escenario de la especificación. El escenario determina si la especificación ingiere datos desde una base de datos relacional en un destino de Hive o en un destino de HDFS. El escenario que se especifique define las propiedades que se pueden configurar en el resto de la especificación.

También se puede habilitar la carga incremental. Si habilita la carga incremental, puede ingerir datos incrementales cuando ejecute la especificación. Después de habilitar la carga incremental, deberá configurar las opciones de carga incremental cuando configure el origen relacional y el destino de Hive o de HDFS.

La siguiente imagen muestra la página Definición de la especificación de ingesta masiva:



The screenshot shows a web interface for configuring a 'Mass Ingestion Spec'. The title bar includes a logo, the text 'Mass Ingestion Spec', and navigation buttons: '< Back', 'Next >', 'Save', and a close button. Below the title bar is a progress indicator with five steps: 1. Definition (active), 2. Source, 3. Target, 4. Mass Transformations, and 5. Transformation Override. The main content area contains the following fields:

- A heading: 'Define the specification to ingest data from a relational database to Hive or HDFS.'
- A 'Name:' field with an asterisk and an empty text input box.
- A 'Description:' field with an empty text area.
- An 'Ingestion Scenario:' field with an asterisk and a dropdown menu currently showing 'Relational to Hive'.
- An 'Enable Incremental Load' checkbox, which is currently unchecked.

En la siguiente tabla se describen las propiedades que se pueden configurar en la definición:

Propiedad	Descripción
Nombre	Obligatorio. Nombre de la especificación de ingesta masiva. La distinción entre mayúsculas y minúsculas no se aplica a este nombre, que además debe ser único. Este nombre no se puede editar después de crear la especificación de ingesta masiva.
Descripción	Opcional. Descripción de la especificación.
Escenario de ingesta	Obligatorio. Escenario donde tiene lugar la ingesta de datos. Se pueden ingerir datos desde una base de datos relacional en un destino de Hive o en un destino de HDFS. El escenario de ingesta no se puede editar después de crear la especificación de ingesta masiva.
Habilitar carga incremental	Opcional. Permite ingerir datos utilizando una carga incremental cuando la especificación ejecuta.

## Origen

El origen describe los datos de origen que se quieren ingerir desde una base de datos relacional.

Para definir el origen, seleccione una conexión de JDBC y un esquema para encontrar la base de datos que desea utilizar. Luego, especifique las tablas que quiera ingerir.

Si habilita la carga incremental en la definición de la especificación de ingesta masiva, debe configurar las opciones de carga incremental en el origen para definir una clave incremental.

En la siguiente imagen se muestra la página Origen de la especificación de ingesta masiva:

The screenshot shows the 'Mass Ingestion Spec' application window. The 'Source' tab is active, showing a configuration screen for data ingestion. It includes dropdown menus for 'Source Connection' and 'Source Schema', and a table selection interface with 'Available' and 'Selected' columns. Below the table selection, there are 'Incremental Load Options' with radio buttons for 'Key Type' (Timestamp or ID) and a text input for 'Key Column'.

En la siguiente tabla se describen las propiedades que se pueden configurar para definir el origen:

Propiedad	Descripción
Conexión de origen	Obligatorio. Conexión de JDBC utilizada para hallar el esquema de origen. Si se producen cambios en las conexiones de JDBC disponibles, actualice el navegador o cierre sesión y vuelva a iniciarla en la herramienta de ingesta masiva.
Esquema de origen	Obligatorio. Esquema que define dónde se almacenan las tablas de origen.
Tablas de origen	Obligatorio. Tablas que se van a ingerir. Seleccione las tablas en la columna Disponible y muévalas a la columna Seleccionado.
Tipo de clave	Obligatorio si habilita la carga incremental. Es el tipo de clave incremental. Seleccione Marca de tiempo o ID. El valor predeterminado es Marca de tiempo.
Columna de claves	Obligatorio si habilita la carga incremental. Nombre de columna de la clave incremental. Si la clave es una marca de tiempo, la columna de claves debe almacenar datos de fecha/hora. Si la clave es un ID, la columna de claves debe almacenar datos numéricos.

## Claves incrementales

Si habilita la carga incremental en la definición de la especificación de ingesta masiva, debe configurar una clave incremental.

Una clave incremental es una columna en una tabla de origen que el motor de Spark usa para obtener datos incrementales de un origen. La clave incremental puede ser una marca de tiempo o un ID, según la manera en que su organización administre las actualizaciones de registros.

### Claves de marca de tiempo

Si la clave incremental es una marca de tiempo, la columna de clave incremental debe almacenar datos de fecha/hora, y la marca de tiempo debe indicar la última vez que la fila de datos se modificó.

Cuando la especificación de ingesta masiva se ejecuta, el motor de Spark obtiene las filas de la tabla de origen con una marca de tiempo más reciente que la marca de tiempo más antigua que se ingirió previamente. Si la marca de tiempo de una fila de la tabla es más reciente que la marca de tiempo más antigua ingerida, el motor de Spark obtiene las filas asociadas a la marca de tiempo como datos incrementales.

Por ejemplo, se puede haber ingerido la siguiente tabla de origen en la ejecución previa de la especificación:

<b>EmpLastName</b>	<b>LastModified</b>
'Basquez'	01/27/2017 02:43:05
'Savage'	03/15/2014 07:16:20
'Greene'	12/13/2012 09:42:11

Observe que la marca de tiempo más antigua es 01/27/2017 02:43:05.

En la siguiente tabla se muestran los datos que hay actualmente en la tabla de origen:

<b>EmpLastName</b>	<b>LastModified</b>
'Basquez'	10/22/2018 04:20:57
'Savage'	03/15/2014 07:16:20
'Greene'	12/13/2012 09:42:11
'Caldwell'	09/13/2018 04:24:26

Como la marca de tiempo más antigua ingerida es 01/27/2017 02:43:05, el motor de Spark obtiene las filas de la tabla de origen con una marca de tiempo más reciente que 01/27/2017 02:43:05.

En la tabla de origen actual se muestran dos marcas de tiempo que son más recientes: 10/22/2018 04:20:57 y 09/13/2018 04:24:26, con lo cual las filas asociadas a esas marcas de tiempo son datos incrementales.

Cuando la especificación se ejecuta, el motor de Spark ingiere las siguientes filas de datos:

<b>EmpLastName</b>	<b>LastModified</b>
'Basquez'	10/22/2018 04:20:57
'Caldwell'	09/13/2018 04:24:26

## Claves de ID

Si la clave incremental es un ID, la columna de clave incremental debe almacenar datos numéricos, y el ID debe indicar que se ha añadido una nueva fila de datos a la tabla de origen.

Cuando la especificación de ingesta masiva se ejecuta, el motor de Spark obtiene las filas de la tabla de origen con un ID que es mayor al valor de ID máximo de las filas que se han ingerido previamente. Si el valor de ID de una fila de la tabla es superior al ID máximo ingerido, el motor de Spark obtiene las filas asociadas al ID como datos incrementales.

Por ejemplo, se puede haber ingerido la siguiente tabla de origen en la ejecución previa de la especificación:

<b>EmpID</b>	<b>EmpLastName</b>
481530	'Basquez'
481531	'Savage'
481532	'Greene'

Tenga en cuenta que el valor de ID máximo es 481532.

En la siguiente tabla se muestran los datos que hay actualmente en la tabla de origen:

<b>EmpID</b>	<b>EmpLastName</b>
481530	'Basquez'
481531	'Savage'
481532	'Greene'
481533	'Caldwell'
481534	'Galloway'

Como los ID 481533 y 481534 son más grandes que el ID máximo (481532) que se ha ingerido previamente, las filas asociadas a estos ID son datos incrementales.

Cuando la especificación se ejecuta, el motor de Spark ingiere las siguientes filas de datos:

<b>EmpID</b>	<b>EmpLastName</b>
481533	'Caldwell'
481534	'Galloway'

# Destino

El destino hace referencia al destino de Hive o HDFS donde desea ingerir los datos de las tablas de origen.

El destino se define según el escenario que se seleccione en la página Definición. Utilice la página Destino para configurar las propiedades del destino.

Si habilita la carga incremental en la especificación de ingesta masiva, debe configurar opciones de carga incremental en el destino. Utilice las opciones de carga incremental para especificar un modo de ingerir datos en el destino y si quiere propagar los cambios de esquema en el origen.

## Destino de Hive

Configure un destino de Hive para ingerir datos de origen en tablas de destino de Hive.

Cuando la especificación de ingesta masiva se configura para ingerir datos en un destino de Hive, hay que configurar una conexión de Hive y propiedades de Hive para definir el destino.

Puede ingerir datos en una tabla de Hive interna o externa. Las tablas de Hive internas se administran en Hive, mientras que las externas son tablas sin administrar. Puede especificar una ubicación externa de una tabla externa de Hive como Amazon S3, Microsoft Azure Data Lake Store o HBase.

Si habilita una carga incremental en la definición de la especificación de ingesta masiva, debe configurar las opciones de carga incremental para que el destino de Hive seleccione un modo para ingerir los datos. Además, puede optar por propagar los cambios de esquema en el origen.

En la siguiente imagen se muestra la página Destino de un destino de Hive:

The screenshot shows the 'Mass Ingestion Spec' configuration interface. At the top, there are navigation buttons: '< Back', 'Next >', and 'Save'. Below these are five numbered steps: 1 Definition, 2 Source, 3 Target (highlighted), 4 Mass Transformations, and 5 Transformation Override. The main content area is titled 'Configure the Hive target where you want to ingest the source tables. Configure partition and cluster properties for specific target tables when you configure table parameters.' The form includes the following fields and options:

- Target Connection:** A dropdown menu with 'Choose a connection' selected.
- Target Schema:** A text input field containing 'default'.
- Target Table Prefix:** An empty text input field.
- Target Table Suffix:** An empty text input field.
- Options:** Radio buttons for 'Hive Options' (selected) and 'DDL Query'.
- Storage Format:** A dropdown menu with 'Text' selected.
- External Table:** An unchecked checkbox.
- External Location:** An empty text input field.
- Incremental Load Options:** A section header with a help icon.
- Mode:** Radio buttons for 'Append' (selected) and 'Overwrite'.
- Propagate schema changes on the source:** An unchecked checkbox.

En la siguiente tabla se describen las propiedades que se pueden configurar para definir el destino de Hive:

Propiedad	Descripción
Conexión de destino	Obligatorio. Conexión de Hive utilizada para hallar el destino de almacenamiento de Hive. Si se realizan cambios en las conexiones Hive disponibles, actualice el navegador o cierre sesión y vuelva a iniciarla en la herramienta de ingesta masiva.
Esquema de destino	Obligatorio. Esquema que define las tablas de destino.
Prefijo de tabla de destino	Prefijo añadido a los nombres de las tablas de destino. Introduzca una cadena. Puede introducir caracteres alfanuméricos y de subrayado. El prefijo no distingue mayúsculas de minúsculas.
Sufijo de tabla de destino	Sufijo añadido a los nombres de las tablas de destino. Introduzca una cadena. Puede introducir caracteres alfanuméricos y de subrayado. El prefijo no distingue mayúsculas de minúsculas.
Opciones de Hive	Seleccione esta opción para configurar la ubicación de destino de Hive.
Consulta DDL	Seleccione esta opción para configurar una consulta DDL personalizada que defina el modo en que los datos de las tablas de origen se cargan en las tablas de destino.
Formato de almacenamiento	Obligatorio. Formato de almacenamiento de las tablas de destino. Puede seleccionar Clúster predeterminado, Texto, Avro, Parquet u ORC. El valor predeterminado es Clúster predeterminado. Si selecciona Clúster predeterminado, la especificación utiliza el formato de almacenamiento predeterminado en el clúster de Hadoop.
Tabla externa	Seleccione esta opción si la tabla es externa.
Ubicación externa	Ubicación externa del destino de Hive. Las tablas se escriben de forma predeterminada en el directorio de almacén de Hive predeterminado. Por cada origen que se ingiere se crea un subdirectorío en la ubicación externa especificada. Por ejemplo, puede especificar <code>/temp</code> . Una tabla de origen con el nombre <code>PRODUCT</code> se ingerirá en la ubicación externa <code>/temp/PRODUCT/</code> .
Modo	Obligatorio si habilita la carga incremental. Seleccione Anexar o Sobrescribir. El modo de anexión anexa los datos incrementales al destino. El modo de sobrescritura sobrescribe los datos en el destino por los datos incrementales. El valor predeterminado es Anexar.
Propagar cambios de esquema en el origen	Opcional. Si se añaden columnas nuevas a las tablas de origen o se modifican columnas existentes, los cambios se propagan a las tablas de destino.

Configure las propiedades de la partición y del clúster para las tablas de destino específicas al configurar el reemplazo de transformación.

Cuando ingiera en un destino de Hive, considere las siguientes directrices:

- Si una tabla de origen se ingiere en un destino de Hive y el nombre de esa tabla de origen contiene una palabra clave reservada en Hive, los datos de la tabla de origen se ingieren en una tabla de destino que tendrá un nombre generado de manera aleatoria.
- Una tabla de origen no se puede ingerir en Hive si los metadatos de tabla utilizan caracteres UTF-8. Para solucionar este problema, configure el almacenamiento metastore de Hive para el procesamiento de datos UTF-8.

- No se puede ingerir una tabla de origen en un archivo Avro ubicado en un destino de Hive si la tabla de origen contiene una columna con un tipo de datos de marca de tiempo o si la carga incremental está configurada con una clave de marca de tiempo. Para ingerir datos de marca de tiempo en un archivo Avro, el controlador Hive JDBC de terceros debe tener una versión de Hive superior a 1.1.
- Si se ejecuta una carga completa para ingerir datos en un destino de Hive de una ubicación externa, todas las filas de la tabla de origen se añaden a la tabla de Hive de destino. Por ejemplo, si la tabla de origen contiene 500 filas y se ejecuta una carga completa dos veces, la tabla de Hive contendrá 1000 filas. Para restablecer la tabla, debe borrar los datos en la ubicación externa.

## Consulta DDL

Cuando se configura una especificación de ingesta masiva para ingerir datos en un destino de Hive, se puede configurar una consulta DDL personalizada que defina cómo se van a cargar los datos de las tablas de origen en las tablas de destino.

Puede definir la consulta DDL para personalizar la tabla de destino o especificar más parámetros. La tabla de destino contiene las columnas que se han definido en la consulta DDL.

Para definir una consulta DDL, utilice marcadores de posición e instrucciones SQL. Utilice los marcadores de posición para obtener el nombre de la tabla, la lista de columnas y los nombres de columna. El servicio de integración de datos sustituye los marcadores de posición por los valores reales en tiempo de ejecución de acuerdo con las tablas que se ingieren. Los marcadores de posición se deben incluir entre llaves. Por ejemplo, {INFA\_TABLE\_NAME}.

Puede utilizar los siguientes marcadores de posición:

### **INFA\_TABLE\_NAME**

Obtiene el nombre de la tabla de destino en tiempo de ejecución.

### **INFA\_COLUMN\_LIST**

Obtiene una lista de columnas de la tabla de destino en tiempo de ejecución.

Por ejemplo, se podría ingerir una tabla `CUSTOMER`. Para definir cómo desea ingerir esta tabla en el destino, puede introducir la siguiente consulta DDL:

```
CREATE TABLE {INFA_TABLE_NAME} ({INFA_COLUMN_LIST}) CLUSTERED BY (LAST_NAME) INTO 10
BUCKETS STORED AS TEXT
```

En tiempo de ejecución, el servicio de integración de datos sustituye {INFA\_TABLE\_NAME} por `CUSTOMER`, así como {INFA\_COLUMN\_LIST} por la lista de columnas que aparecen en la tabla `CUSTOMER`. Puede que el servicio de integración de datos expanda la consulta DDL a la siguiente consulta:

```
CREATE TABLE CUSTOMER (FIRST_NAME STRING, LAST_NAME STRING, EMAIL STRING, GENDER STRING,
CREDIT_CARD DECIMAL (38,0), CREDIT_CARD_TYPE STRING, STATE, STRING, USSTATE STRING, CITY
STRING) CLUSTERED BY (LAST_NAME) INTO 10 BUCKETS STORED AS TEXT
```

**Nota:** No se puede utilizar un marcador de posición para especificar las columnas de partición y agrupado por columnas. Cuando especifique las columnas de partición y agrupado por columnas, introduzca el nombre de la columna en la consulta DDL.

La siguiente imagen muestra la opción para configurar una consulta DDL de un destino de Hive:

Mass Ingestion Spec

< Back Next > Save

1 Definition 2 Source 3 Target 4 Mass Transformations 5 Transformation Override

Configure the Hive target where you want to ingest the source tables. Configure partition and cluster properties for specific target tables when you configure table parameters.

Target Connection:\* Choose a connection

Target Schema:\* default

Target Table Prefix: Target Table Suffix:

Hive Options  DDL Query

DDL Query: CREATE TABLE {{INFA\_TABLE\_NAME}} ({{INFA\_COLUMN\_LIST}})

Incremental Load Options

Mode:\*  Append  Overwrite

Propagate schema changes on the source

## Destino de HDFS

Configure un destino de HDFS para ingerir datos de origen en un archivo sin formato en HDFS.

Cuando la especificación de ingesta masiva se configura para ingerir datos en un destino de HDFS, hay que configurar una conexión de HDFS y un directorio de ingesta para definir el destino.

Si habilita una carga incremental en la definición de la especificación de ingesta masiva, debe configurar las opciones de carga incremental para que el destino de HDFS seleccione un modo de ingerir los datos.

En la siguiente imagen se muestra la página Destino de un destino de HDFS:

En la siguiente tabla se describen las propiedades que se pueden configurar para definir el destino de HDFS:

Propiedad	Descripción
Conexión de destino	Obligatorio. Conexión de HDFS utilizada para hallar el destino de almacenamiento de HDFS. Si se realizan cambios en las conexiones HDFS disponibles, actualice el navegador o cierre sesión y vuelva a iniciarla en la herramienta de ingesta masiva.
Prefijo de tabla de destino	Prefijo añadido a los nombres de los archivos de destino. Introduzca una cadena. Puede introducir caracteres alfanuméricos y de subrayado. El prefijo no distingue mayúsculas de minúsculas.
Sufijo de tabla de destino	Sufijo añadido a los nombres de los archivos de destino. Introduzca una cadena. Puede introducir caracteres alfanuméricos y de subrayado. El prefijo no distingue mayúsculas de minúsculas.
Directorio de ingesta	Obligatorio. Directorio de destino en HDFS. Por cada origen que se ingiere, se crea un subdirectorio en el directorio de ingesta. Si el directorio especificado ya existe, se reemplazará. Por ejemplo, puede especificar <code>/temp</code> . Una tabla de origen con el nombre <code>PRODUCT</code> se ingerirá en el directorio <code>/temp/PRODUCT/</code> .
Compresión	Obligatorio. Formato de archivo comprimido que almacena los archivos de destino. Puede seleccionar Ninguno, Gzip, Bzip2, LZ0, Snappy o Personalizado. Si elige Personalizado, introduzca el códec de compresión. El valor predeterminado es ninguno.
Códec de compresión	Si selecciona una compresión personalizada, introduzca el nombre de clase completo que implementa la interfaz <code>CompressionCodec</code> de Hadoop.

Propiedad	Descripción
Delimitadores	Delimitadores utilizados para separar datos en los archivos de destino. Puede seleccionar coma, punto y coma, espacio, tabulación u otro cualquiera. Si selecciona Otro, puede definir un delimitador personalizado.
Otro delimitador	Obligatorio si elige Otro como delimitador. Introduzca un delimitador personalizado.
Modo	Obligatorio si habilita la carga incremental. Seleccione Anexar o Sobrescribir. El modo de anexión anexa los datos incrementales al destino. El modo de sobrescritura sobrescribe los datos en el destino por los datos incrementales. El valor predeterminado es Anexar.

**Nota:** Cuando el servicio de integración de datos almacena archivos temporales que se ingieren en un destino de HDFS, anexa un ID único al nombre de archivo original. El nombre de archivo resultante puede tener una longitud máxima de 255 caracteres.

## Códec de compresión

Cuando se configura una especificación de ingesta masiva para ingerir datos en un directorio de destino de HDFS, puede configurar un códec de compresión para escribir los datos ingeridos en un archivo comprimido.

Puede seleccionar una de las siguientes opciones de compresión:

- Gzip
- Bzip2
- LZO
- Snappy
- Personalizado

Si especifica un códec de compresión personalizado, debe especificar el nombre de clase completo que implementa la interfaz `CompressionCodec` de Hadoop.

## Opciones de carga incremental

Si habilita la carga incremental en la definición de la especificación de ingesta masiva, debe configurar opciones de carga incremental en el destino.

Las opciones de carga incremental del destino incluyen un modo incremental. El motor de Spark utiliza el modo incremental para determinar cómo cargar datos incrementales en el destino.

Si se ingieren datos en un destino de Hive, también puede configurar las opciones de carga incremental para propagar los cambios de esquema al origen. Propagar cambios de esquema permite los desfases de esquema en la base de datos relacional como, por ejemplo, las columnas nuevas que se han añadido a las tablas de origen.

## Modos incrementales

Si habilita la carga incremental en una especificación de ingesta masiva, debe seleccionar un modo de ingerir los datos.

Puede seleccionar el modo de anexión o el modo de sobrescritura. El modo de anexión anexa los datos incrementales a los datos en el destino. El modo de sobrescritura sobrescribe los datos en el destino por los datos incrementales. El modo de sobrescritura se puede usar si el destino es un área de almacenamiento provisional, y no el almacén de datos operativo final.

Por ejemplo, el destino puede contener los siguientes datos de una ejecución previa de la especificación:

<b>EmpID</b>	<b>EmpLastName</b>
481530	'Basquez'
481531	'Savage'
481532	'Greene'

En la siguiente tabla se muestran los datos incrementales que el motor de Spark ingiere en la ejecución actual de la especificación:

<b>EmpID</b>	<b>EmpLastName</b>
481533	'Caldwell'
481534	'Galloway'

Si utiliza el modo de anexión, los datos incrementales se anexan al destino. Cuando el trabajo de ingesta se complete, el destino contendrá los siguientes datos:

<b>EmpID</b>	<b>EmpLastName</b>
481530	'Basquez'
481531	'Savage'
481532	'Greene'
481533	'Caldwell'
481534	'Galloway'

Si utiliza el modo de sobrescritura, los datos incrementales sobrescriben el destino. Cuando el trabajo de ingesta se complete, el destino contendrá los siguientes datos:

<b>EmpID</b>	<b>EmpLastName</b>
481533	'Caldwell'
481534	'Galloway'

## Propagar cambios de esquema

Cuando se ingieren datos en un destino de Hive, los cambios de esquema en el origen se pueden propagar para permitir los desfases de esquema en la base de datos relacional.

El desfase de esquema sucede cuando el esquema de una tabla de origen cambia, por ejemplo, al añadirle columnas nuevas o cambiar una ya existente en la tabla. Revisar los esquemas de las tablas de destino para ver si coinciden con los esquemas de las tablas de origen cada vez que se va a realizar una carga de datos puede ser una tarea muy laboriosa.

Para dar cabida a posibles desfases de esquema en las tablas de origen, puede configurar cargas incrementales para propagar los cambios de esquema del origen al destino. Si las tablas de origen tienen columnas nuevas o cambios en columnas existentes, esta opción propaga los cambios a las tablas de destino.

En la siguiente imagen se muestra la propiedad que se puede utilizar para propagar cambios de esquema:

The screenshot shows the 'Mass Ingestion Spec' configuration window, currently on the 'Target' step (indicated by a blue circle with the number 3). The window has a title bar with '< Back', 'Next >', and 'Save' buttons, and a close button. Below the title bar is a progress indicator with five steps: 1 Definition, 2 Source, 3 Target, 4 Mass Transformations, and 5 Transformation Override. The main content area contains the following configuration options:

- Target Connection: \* Choose a connection (dropdown menu)
- Target Schema: \* default (text input)
- Target Table Prefix: ? (text input)
- Target Table Suffix: ? (text input)
- Mode:  Hive Options  DDL Query
- Storage Format: \* Text (dropdown menu)
- External Table
- External Location: ? (text input)
- Incremental Load Options ?
- Mode: \*  Append  Overwrite
- Propagate schema changes on the source (highlighted with a red box)

Para obtener más información sobre cómo propagar cambios de esquema, vea la *Guía del usuario de Informática PowerExchange for Hive*. La ingesta masiva hace uso de la estrategia de esquema de destino APPLYNEWSHEMA.

## Transformaciones masivas

Configure transformaciones masivas para definir cualquier parámetro que desee aplicar a todas las tablas de origen que se ingieran desde una base de datos relacional.

Para establecer parámetros para tablas específicas, configure el reemplazo de transformación.

En la siguiente imagen se muestra la página Transformaciones masivas de la especificación de ingesta masiva:

En la siguiente tabla se describen los parámetros que se pueden configurar:

Propiedad	Descripción
Filtrar por	<p>Filtra filas de la tabla de destino según el criterio de una columna de la tabla. Introduzca una cláusula de filtro para determinar los criterios, por ejemplo, <code>STATE='California'</code>. En la cláusula de filtro se puede utilizar cualquier función de lenguaje de transformación. La cláusula de filtro debe evaluar como TRUE o FALSE.</p> <p>Para establecer criterios correspondientes a varias columnas, utilice los operadores AND y OR. Por ejemplo, utilice el operador AND para especificar una cláusula de filtro como <code>STATE='California' AND STATUS='Single'</code>.</p>
Eliminar columnas	<p>Elimina columnas en la tabla de destino. Introduzca cada columna como una cadena y separe los nombres de columna con una coma, o especifique una expresión regular.</p> <p>Por ejemplo, si introduce <code>COL1</code>, la columna <code>COL1</code> se eliminará en la tabla de destino.</p> <p>Si introduce una expresión regular como <code>.*SSN.*</code>, las columnas cuyos nombres contienen "SSN" se eliminarán.</p>
Acortar	<p>Acorta valores de columna en la tabla de destino para eliminar espacios antes y después de los valores. Introduzca cada columna como una cadena y separe los nombres de columna con una coma, o especifique una expresión regular.</p> <p>Por ejemplo, si introduce <code>COL1</code>, los valores en la columna <code>COL1</code> se acortarán en la tabla de destino.</p> <p>Si introduce una expresión regular como <code>.*SSN.*</code>, los valores en las columnas cuyos nombres contienen "SSN" se acortarán.</p>
Convertir en mayúsculas	<p>Cambia a mayúsculas los valores de columna en la tabla de destino. Introduzca cada columna como una cadena y separe los nombres de columna con una coma, o especifique una expresión regular.</p> <p>Por ejemplo, si introduce <code>COL1</code>, los valores en la columna <code>COL1</code> cambiarán a mayúsculas en la tabla de destino.</p> <p>Si introduce una expresión regular como <code>.*SSN.*</code>, los valores en las columnas cuyos nombres contienen "SSN" cambiarán a mayúsculas.</p>

Propiedad	Descripción
Convertir en minúsculas	<p>Cambia a minúsculas los valores de columna en la tabla de destino. Introduzca cada columna como una cadena y separe los nombres de columna con una coma, o especifique una expresión regular.</p> <p>Por ejemplo, si introduce COL1, los valores en la columna COL1 cambiarán a minúsculas en la tabla de destino.</p> <p>Si introduce una expresión regular como <code>. *SSN. *</code>, los valores en las columnas cuyos nombres contienen "SSN" cambiarán a minúsculas.</p>
Reemplazar columnas	<p>Reemplaza los valores de columna en la tabla de destino. Introduzca cada columna como una cadena y separe los nombres de columna con una coma, o especifique una expresión regular.</p> <p>Por ejemplo, si introduce COL1, los valores en la columna COL1 se reemplazarán en la tabla de destino.</p> <p>Si introduce una expresión regular como <code>. *SSN. *</code>, los valores en las columnas cuyos nombres contienen "SSN" se reemplazarán.</p> <p>Si especifica columnas para reemplazarlas, debe especificar los criterios de reemplazo.</p>
Reemplazar criterios	<p>Obligatorio si especifica columnas para reemplazarlas. Determina la forma de reemplazar los valores de columna en la tabla de destino. Puede seleccionar <code>Patrón</code> o <code>Cadena entera</code>.</p> <p>Si selecciona <code>Cadena entera</code>, todos los valores en las columnas que especifique se reemplazan por el valor que configure.</p> <p>Si selecciona <code>Patrón</code>, introduzca el patrón que quiera reemplazar y, luego, introduzca el valor por el que quiera reemplazar ese patrón.</p>

Propiedad	Descripción
Patrón	<p>Obligatorio si configura los criterios de reemplazo usando el método del patrón. Determina el patrón que se va a reemplazar. El patrón debe ser una expresión regular. Por ejemplo, si desea reemplazar los valores en las columnas que contienen números de la Seguridad Social, puede introducir el siguiente patrón:</p> <pre>^\d(3)-?\d(2)-?\d(4)\$</pre> <p>El patrón reemplazará todos los valores que coincidan con él. En la columna que contiene números de la Seguridad Social, el patrón reemplaza el número de la Seguridad Social completo.</p> <p>Para enmascarar solo los primeros cinco dígitos del número de la Seguridad Social, puede introducir este patrón:</p> <pre>^\d(3)-?\d(2)\$</pre>
Valor	<p>Obligatorio si configura columnas para reemplazarlas. Reemplaza el patrón o la cadena entera de acuerdo a los criterios configurados.</p> <p>Si selecciona <i>Cadena entera</i> como criterio de reemplazo, el valor que especifique reemplazará los valores en todas las columnas que se hayan especificado para el reemplazo. Por ejemplo, si reemplaza COL1 y COL2 e introduce el valor XXX, todos los valores de las columnas COL1 y COL2 se reemplazarán por el valor XXX.</p> <p>Si selecciona <i>Patrón</i> como criterio de reemplazo, el valor que especifique reemplazará todos los valores que coincidan con el patrón.</p> <p>Por ejemplo, si quiere enmascarar el número entero de la Seguridad Social de acuerdo con el patrón:</p> <pre>^\d(3)-?\d(2)-?\d(4)\$</pre> <p>Deberá introducir el valor XXX-XX-XXXX. Todos los valores en la columna SSN aparecerán como XXX-XX-XXXX.</p> <p>Puede que también quiera enmascarar solo los primeros cinco dígitos del número de la Seguridad Social de acuerdo con el patrón:</p> <pre>^\d(3)-?\d(2)\$</pre> <p>Introduzca el valor XXX-XX. Los primeros cinco dígitos de cada número de la Seguridad Social aparecerán como XXX-XX. Por ejemplo, si el número original de la Seguridad Social es 123-45-6789, el valor reemplazado será XXX-XX-6789.</p>

Si configura parámetros para una columna de tabla que no existe en todas las tablas, las tablas donde la columna no exista no se ingerirán. Deberá reconfigurar los parámetros de cada tabla donde la columna no exista cuando configure el reemplazo de transformación.

## Filtro Cláusulas

Las cláusulas de filtro determinan la manera en que las filas de la tabla de destino se filtran según los criterios de una columna de la tabla.

Para diseñar una cláusula de filtro, puede especificar los nombres de columna en la tabla de destino y cualquiera de las funciones de lenguaje de transformación que usen nombres de columna como argumentos. Para establecer varios criterios, utilice los operadores AND y OR. La cláusula de filtro debe evaluar como TRUE o FALSE.

Por ejemplo, podría ingerir la siguiente tabla de origen:

EMPLOYEEID	PHONENUMBER
607014	(630) 4468851

EMPLOYEEID	PHONENUMBER
620368	(904) 3854084
698107	(549) 5694371
621861	(904) 9062721

Para filtrar las filas que se ingieren en el destino, puede utilizar la siguiente cláusula de filtro para obtener solo las filas con el código de área (904):

```
RTRIM(PHONENUMBER, REG_EXTRACT(PHONENUMBER, '.*{[0-9]{7}}$')) = '(904)'
```

Las filas en la tabla se filtran por el código de área (904). La siguiente tabla se ingiere en el destino:

EMPLOYEEID	PHONENUMBER
620368	(904) 3854084
621861	(904) 9062721

Para obtener más información sobre las funciones de lenguaje de transformación, vea el capítulo "Funciones" de la *Referencia del lenguaje de transformación de Informatica*.

## Expresiones regulares

Una expresión regular describe un rango o un patrón de valores.

Puede utilizar una expresión regular para especificar las columnas que quiera parametrizar en la especificación de ingesta masiva. Utilice una expresión regular cuando las columnas de diferentes tablas de origen se denominen de distinta forma, pero contengan la misma información. Si elige reemplazar las columnas, puede utilizar también una expresión regular para especificar el patrón en los criterios de reemplazo.

Por ejemplo, puede que quiera eliminar las columnas que contengan números de la Seguridad Social. Todos los nombres de las columnas contienen SSN, pero dichos nombres tienen distintos prefijos dependiendo de la tabla de origen donde aparece una columna. Para especificar todas las variaciones en los nombres de columna, puede utilizar una expresión regular como `.*SSN`.

## Reemplazar criterios

Los criterios de reemplazo determinan cómo se van a reemplazar los valores de columna en la tabla de destino.

Configure los criterios de reemplazo para reemplazar los valores en columnas de acuerdo a una cadena entera o a un patrón.

### Reemplazar una cadena entera

Si configura los criterios de reemplazo para reemplazar una cadena entera, todos los valores de las columnas que especifique se reemplazarán según los nuevos valores que se configuren. Por ejemplo, configure los criterios para reemplazar los valores en las columnas `COL1` y `COL2`, y configure el nuevo valor como `XXX`. Todos los valores en las columnas `COL1` y `COL2` se reemplazarán por el valor `XXX`.

En la siguiente imagen se muestran las opciones para configurar criterios de reemplazo según una cadena entera:

Mass Ingestion Spec

< Back Next > Save

1 Definition 2 Source 3 Target 4 Mass Transformations 5 Transformation Override

Define the changes that you want to apply to all source tables. You can override values for individual source tables on the next screen.

Filter By:

Drop Columns:

Trim:

Convert to Uppercase:

Convert to Lowercase:

Replace Columns:

⚠ Columns COL1 will be replaced.

Replace Criteria: \*  Value:

## Reemplazar un patrón

Si configura los criterios de reemplazo para reemplazar un patrón, introduzca el patrón y un nuevo valor que reemplace los valores que coincidan con el patrón. El patrón que introduzca debe ser una expresión regular.

Por ejemplo, desea enmascarar todos los valores en una columna que contiene direcciones IP. Para hallar coincidencias con una dirección IP, puede introducir el siguiente patrón:

```
^(\d{1,2}|\d\d\d|2[0-4]\d|25[0-5])\.  
(\d{1,2}|\d\d\d|2[0-4]\d|25[0-5])\  
(\d{1,2}|\d\d\d|2[0-4]\d|25[0-5])\  
(\d{1,2}|\d\d\d|2[0-4]\d|25[0-5])$
```

Todos los valores que coincidan con el patrón se reemplazan de acuerdo con el nuevo valor que configure. Si configura el nuevo valor como `xxx`, todos los valores que coincidan con el patrón se reemplazarán por `xxx`.

En la siguiente imagen se muestran las opciones para configurar los criterios de reemplazo basados en un patrón:

Mass Ingestion Spec

< Back Next > Save

1 Definition 2 Source 3 Target 4 Mass Transformations 5 Transformation Override

Define the changes that you want to apply to all source tables. You can override values for individual source tables on the next screen.

Filter By:

Drop Columns:

Trim:

Convert to Uppercase:

Convert to Lowercase:

Replace Columns:

⚠ Columns COL1 will be replaced.

Replace Criteria: \*  Previous string or pattern Value:

## Reemplazo de transformación

Configure el reemplazo de transformación para reemplazar las transformaciones masivas que se aplican a todas las tablas de origen en una especificación de ingesta masiva. Cuando configure el reemplazo, puede establecer parámetros para aplicarlos a tablas específicas que se ingieren desde la base de datos relacional.

El reemplazo de transformación se puede configurar de una de las siguientes maneras:

- Antes de configurar el reemplazo de transformación, configure transformaciones masivas para aplicar parámetros a todas las tablas de origen. Luego, vaya a cada tabla y edite el parámetro aplicado.
- Borre las transformaciones masivas y configure parámetros específicos para cada tabla.

En la siguiente imagen se muestra la página Reemplazo de transformación de un destino de HDFS:

Mass Ingestion Spec

< Back Next > Save

1 Definition 2 Source 3 Target 4 Mass Transformations 5 Transformation Override

Define the changes that you want to apply to individual source tables. The parameters configured on this page override the mass transformation applied to a source table.

Ingestion Objects (1)

Sources	Incremental Column	Filter By	Drop Columns	Trim	Convert to Uppercase	Convert to Lowercase	Replace Columns
EMPLOYEE	<input type="text"/>						

Se pueden editar los siguientes parámetros de cada tabla de origen:

- Columna incremental. Si la carga incremental está habilitada, edite la columna que determina los datos incrementales en las tablas de origen.
- Filtrar por. Filtra las columnas de las tablas de origen.

- Eliminar columnas. Elimina columnas de las tablas de origen.
- Reemplazar columnas. Reemplaza los valores en las columnas de las tablas de origen.
- Acortar. Acorta los espacios iniciales y finales de los valores en las columnas de las tablas de origen.
- Convertir en mayúsculas. Cambia a mayúsculas los valores en las columnas de las tablas de origen.
- Convertir en minúsculas. Cambia a minúsculas los valores en las columnas de las tablas de origen.
- Hive. Si la especificación ingiere datos en un destino de Hive, especifique las propiedades de Hive de las tablas de destino.

Cuando se configuran parámetros de una tabla de origen, no se pueden configurar diferentes parámetros para columnas distintas en la tabla. Si se configuran parámetros de una columna que no existe en la tabla, dicha tabla no se podrá ingerir.

## Opciones de Hive

Cuando se configura un reemplazo de transformación para un destino de Hive, se pueden configurar más parámetros para especificar la forma en que los datos se cargan en las tablas de destino de Hive.

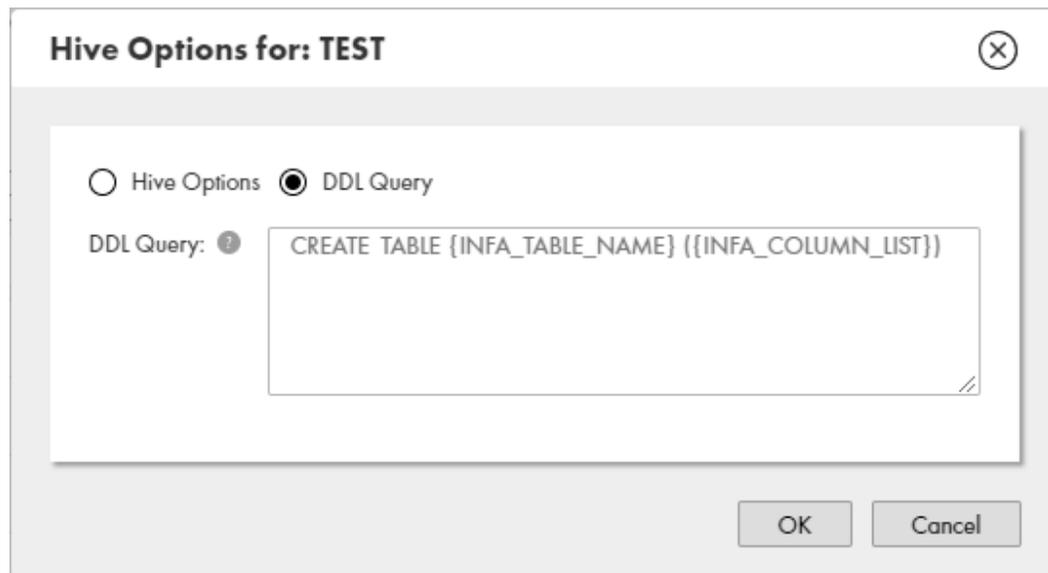
En la siguiente imagen se muestra el cuadro de diálogo que aparece cuando se editan las propiedades de la tabla de destino de Hive:

En la siguiente tabla se describen las propiedades de la tabla de destino de Hive:

Propiedad	Descripción
Opciones de Hive	Seleccione esta opción para configurar la ubicación de destino de Hive.
Consulta DDL	Seleccione esta opción para configurar una consulta DDL personalizada que defina el modo en que los datos de las tablas de origen se cargan en las tablas de destino.
Formato de almacenamiento	Obligatorio. Formato de almacenamiento de las tablas de destino. Puede seleccionar Texto, Avro, Parquet u ORC. El valor predeterminado es Texto.

Propiedad	Descripción
Ubicación externa	Ubicación externa del destino de Hive. Introduzca una ubicación externa para especificar una ubicación distinta al almacén depósito de Hive predeterminado. Por cada origen que se ingiere se crea un subdirectorio en la ubicación externa especificada. Por ejemplo, puede especificar /temp. Una tabla de origen con el nombre PRODUCT se ingerirá en la ubicación externa /temp/PRODUCT/.
Clave de partición	Clave de partición de la tabla de Hive de destino.
Agrupado por	Clave de clúster de la tabla de Hive de destino.
Número de cubos	Obligatorio si se especifica una clave de clúster. Número de cubos entre el que se va a dividir la tabla de Hive de destino.

En la siguiente imagen se muestra el cuadro de diálogo que aparece cuando se edita la consulta DDL en las propiedades de la tabla de destino de Hive:



Para editar la consulta DDL, utilice instrucciones SQL y los siguientes marcadores de posición:

**INFA\_TABLE\_NAME**

Obtiene el nombre de la tabla de destino en tiempo de ejecución.

**INFA\_COLUMN\_LIST**

Obtiene una lista de columnas de la tabla de destino en tiempo de ejecución.

# Crear una especificación de ingesta masiva

Cree una especificación de ingesta masiva en la herramienta de ingesta masiva.

1. En la herramienta de ingesta masiva, haga clic en **Nuevo** para crear una especificación de ingesta masiva.  
Se abre la página **Definición** relativa a la nueva especificación de ingesta masiva.
2. Para diseñar la especificación de ingesta masiva, configure las propiedades en cada página del asistente.
3. Haga clic en **Guardar**.  
Se abre la página **Resumen** de la especificación de ingesta masiva.

## Configurar la definición

Configure una definición para determinar el tipo de especificación de ingesta masiva que desea crear.

1. En la página **Definición**, introduzca el nombre y la descripción de la especificación de ingesta masiva.
2. Elija uno de los siguientes escenarios de ingesta:
  - Relacional con Hive. Seleccione este escenario para ingerir datos de una base de datos relacional en un destino de Hive.
  - Relacional con HDFS. Seleccione este escenario para ingerir datos de una base de datos relacional en un destino de HDFS.
3. Si lo desea, seleccione **Habilitar carga incremental** si tiene previsto ejecutar cargas incrementales de los datos.  
Si no habilita la carga incremental, solo podrá utilizar cargas completas para ingerir datos.

## Configurar el origen

Configure el origen para especificar la base de datos relacional y las tablas de origen relacionales que desee ingerir.

1. En la página **Origen**, seleccione la conexión de origen y el esquema de origen.  
Las tablas de origen que hay disponibles en el esquema aparecen debajo de la columna **Seleccionado**.
2. Para añadir tablas de origen a la especificación de ingesta masiva, arrastre las tablas de origen desde la columna **Disponible** hasta la columna **Seleccionado**.
3. Si ha habilitado la carga incremental en la definición de la especificación de ingesta masiva, configure la clave incremental utilizando las siguientes propiedades:
  - a. Tipo de clave. Seleccione **Marca de tiempo** o **ID**. El valor predeterminado es Marca de tiempo.
  - b. Columna de claves. Introduzca el nombre de la columna que almacena la clave incremental. Si la clave es una marca de tiempo, la columna debe almacenar datos de fecha/hora. Si la clave es un ID, la columna debe almacenar datos numéricos.

## Configurar un destino de Hive

Si la especificación de ingesta masiva ingiere datos en un destino de Hive, configure dicho destino para especificar cómo quiere que los datos se ingieran en las tablas de destino.

1. Seleccione la conexión de destino.
2. Introduzca el esquema de destino.
3. Si lo desea, introduzca un prefijo y un sufijo para añadirlos a las tablas de destino.
4. Especifique una de las siguientes opciones:
  - Seleccione **Opciones de Hive**.
    1. Seleccione el formato de almacenamiento. El valor predeterminado es texto.
    2. Si el destino es un destino de Hive externo, seleccione **Tabla externa** e introduzca la ubicación externa.
  - Seleccione **Consulta DDL** e introduzca una consulta DDL personalizada.
5. Si ha habilitado la carga incremental en la definición de la especificación de ingesta masiva, configure las siguientes opciones de carga incremental:
  - Modo. Seleccione **Anexar** o **Sobrescribir**. El valor predeterminado es Anexar.
  - Propagar cambios de esquema en el origen. Seleccione esta opción si desea actualizar las columnas en las tablas de origen o añadir a ellas columnas nuevas.

## Configurar un destino de HDFS

Si la especificación de ingesta masiva ingiere datos en un destino de HDFS, configure el destino de HDFS para especificar cómo quiere que los datos se ingieran en los archivos sin formato de HDFS.

1. Seleccione la conexión de destino.
2. Si lo desea, introduzca un prefijo y un sufijo para añadirlos a los archivos de destino.
3. Especifique el directorio de ingesta.
4. Seleccione el formato de compresión. Si elige utilizar una compresión personalizada, introduzca el códec de compresión.
5. Seleccione el delimitador. Si selecciona **Otro**, introduzca un delimitador personalizado.
6. Si ha habilitado la carga incremental en la definición de la especificación de ingesta masiva, configure la siguiente opción de carga incremental:
  - Modo. Seleccione **Anexar** o **Sobrescribir**. El valor predeterminado es Anexar.

## Configurar transformaciones masivas

Si lo desea, configure una transformación masiva para aplicar parámetros en todas las tablas de origen que se ingieran.

1. Para **Filtrar por**, introduzca una cláusula de filtro.
2. Para **Eliminar columnas**, indique las columnas que quiera eliminar o introduzca una expresión regular.
3. Para **Acortar**, indique las columnas que quiera acortar o introduzca una expresión regular.
4. Para **Convertir en mayúsculas**, indique las columnas que quiera convertir en mayúsculas o introduzca una expresión regular.

5. Para **Convertir en minúsculas**, indique las columnas que quiera convertir en minúsculas o introduzca una expresión regular.
6. Para **Reemplazar columnas**, indique las columnas que quiera reemplazar o introduzca una expresión regular.
  - a. Para **Reemplazar criterios**, seleccione **Cadena entera** o **Patrón**. El valor predeterminado es Cadena entera.
  - b. Si selecciona Cadena entera, introduzca la nueva cadena.
  - c. Si selecciona Patrón, introduzca la cadena o el patrón anterior. Luego, especifique la nueva cadena.

## Configurar el reemplazo de transformación

Si lo desea, configure el reemplazo de transformación para reemplazar los parámetros que se aplican a tablas de origen específicas.

**Nota:** En algunos parámetros, puede hacer clic en flecha junto al cuadro de texto para seleccionar columnas de una lista de columnas que hay en la tabla.

1. Si ha habilitado la carga incremental en la definición de la especificación de ingesta masiva, edite la columna de clave incremental de una tabla de origen en la columna **Columna incremental**.
2. En la columna **Filtrar por**, edite la cláusula de filtro.
3. En la columna **Eliminar columnas**, edite la lista de columnas o la expresión regular.
4. En la columna **Acortar**, edite la lista de columnas o la expresión regular.
5. En la columna **Convertir en mayúsculas**, edite la lista de columnas o la expresión regular.
6. En la columna **Convertir en minúsculas**, edite la lista de columnas o la expresión regular.
7. En la columna **Reemplazar columnas**, haga clic en el lápiz.
  - a. Para **Reemplazar**, edite la lista de columnas o la expresión regular.
  - b. Para **Reemplazar criterios**, seleccione **Cadena entera** o **Patrón**.
  - c. Si selecciona Cadena entera, edite la nueva cadena.
  - d. Si selecciona Patrón, edite la cadena o el patrón previo, o edite la nueva cadena.
8. Si el destino es un destino de Hive, haga clic en el lápiz en la columna **Hive** y elija una de las siguientes opciones:
  - Seleccione **Opciones de Hive** e introduzca la clave de la partición, la clave del clúster y el número de cubos.
  - Seleccione **Consulta DDL** y edite la consulta DDL personalizada.

# CAPÍTULO 4

## Implementar

Este capítulo incluye los siguientes temas:

- [Resumen de implementación, 50](#)
- [Implementar una especificación de ingesta masiva, 51](#)
- [Volver a implementar una especificación de ingesta masiva, 52](#)
- [Implementar en un archivo de almacenamiento de la aplicación, 52](#)
- [Migrar una especificación de ingesta masiva, 53](#)

## Resumen de implementación

Implemente una especificación de ingesta masiva en un servicio de integración de datos para preparar la ejecución de la especificación e ingerir datos en el destino.

Antes de implementar una especificación, debe especificar un servicio de integración de datos y una conexión de Hadoop. El servicio de integración de datos utiliza la conexión de Hadoop para conectarse al entorno de Hadoop cuando se ejecuta la especificación implementada. Tras implementar la especificación de ingesta masiva, puede volver a implementarla utilizando una conexión de Hadoop diferente, o bien volver a implementarla en otro servicio de integración de datos.

Cuando la especificación de ingesta masiva se implementa, se aísla de los cambios en la herramienta de ingesta masiva. Por ejemplo, implementa la especificación de ingesta masiva en un servicio de integración de datos. Tras comprobar la salida de ingesta, edita la especificación de ingesta masiva. Los cambios que realice en la especificación de ingesta masiva no afectarán a la especificación implementada.

# Implementar una especificación de ingesta masiva

Implemente una especificación de ingesta masiva en la herramienta de ingesta masiva.

**Nota:** Si realiza cambios en las conexiones de Hadoop o los servicios de integración de datos disponibles, actualice el navegador, o cierre sesión y vuelva a iniciarla en la herramienta de ingesta masiva.

1. Vaya a una especificación de ingesta masiva en la herramienta de ingesta masiva.

Se abre la página **Resumen** de la especificación de ingesta masiva.

En la siguiente imagen se muestra la vista **Implementar**:

MI

Summary

Name: MI  
Description  
Last modified time: 3/15/2018, 12:06:53 PM  
Last modified by: Administrator

Deploy

Deploy on: Select a Data Integration Ser

Execution History (0)

Start Time	Service Name	Load Type	Status
No data to display			

2. En la vista **Implementar**, seleccione un servicio de integración de datos.

Aparece la opción para seleccionar una conexión de Hadoop.

3. Seleccione una conexión de Hadoop.

En la siguiente imagen se muestra la opción para seleccionar una conexión de Hadoop en la vista **Implementar**:

MI

Summary

Name: MI  
Description  
Last modified time: 3/15/2018, 12:06:53 PM  
Last modified by: Administrator

Deploy

Deploy on: DIS

Hadoop Connection: Choose a Hadoop connection

Deploy

Execution History (0)

Start Time	Service Name	Load Type	Status
No data to display			

4. Haga clic en **Implementar** para implementar la especificación de ingesta masiva.

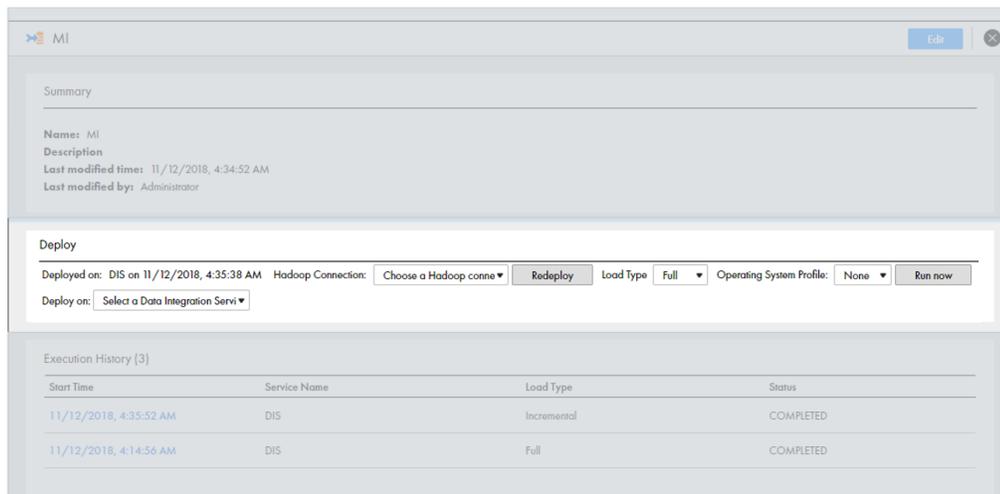
## Volver a implementar una especificación de ingesta masiva

Vuelva a implementar una especificación de ingesta masiva en un servicio de integración de datos distinto o una conexión de Hadoop diferente.

**Nota:** Si realiza cambios en las conexiones de Hadoop o los servicios de integración de datos disponibles, actualice el navegador, o cierre sesión y vuelva a iniciarla en la herramienta de ingesta masiva.

1. Vaya a una especificación de ingesta masiva en la herramienta de ingesta masiva.  
Se abre la página **Resumen** de la especificación de ingesta masiva.
2. Para volver a implementar la especificación de ingesta masiva en otra conexión de Hadoop, seleccione un servicio de integración de datos en la vista **Implementar**.

En la siguiente imagen se muestra la vista **Implementar** en la página **Resumen**, donde se puede volver a implementar la especificación:



**Nota:** Si no se selecciona un servicio de integración de datos, la especificación de ingesta masiva se vuelve a implementar en el mismo servicio de integración de datos en una conexión de Hadoop distinta.

3. Seleccione una conexión de Hadoop.
4. Haga clic en **Volver a implementar** para volver a implementar la especificación de ingesta masiva.

## Implementar en un archivo de almacenamiento de la aplicación

Un archivo de almacenamiento de la aplicación contiene los objetos y los metadatos de una aplicación en formato XML. Cuando se implementa una especificación de ingesta masiva en un archivo de

almacenamiento de la aplicación, toda la información sobre la especificación se guarda como una aplicación en el archivo XML.

Este archivo tiene una extensión .iar. Una especificación de ingesta masiva solamente se puede implementar en un archivo de almacenamiento de la aplicación a través de la línea de comandos.

Se recomienda crear un archivo de almacenamiento de la aplicación para realizar lo siguiente:

- Implementar la aplicación. Si su organización solo permite a los administradores implementar aplicaciones en un Servicio de integración de datos, un administrador puede implementar una aplicación de un archivo de almacenamiento en un Servicio de integración de datos. El administrador puede utilizar la Herramienta del administrador o el comando `infacmd dis deployApplication`.
- Importar la aplicación a un repositorio de modelos. Puede utilizar Developer tool para importar la aplicación a un Repositorio de modelos, o bien un administrador puede importar la aplicación a un Repositorio de modelos utilizando el comando `infacmd tools importObjects`.
- Guardar el archivo de almacenamiento de aplicación en otro sistema. Por ejemplo, si el repositorio de modelos no está integrado con un sistema de control de versiones, un administrador puede proteger el archivo de almacenamiento en un sistema de control de versiones.
- Hacer una copia de seguridad de la especificación. Es posible que quiera hacer una copia de seguridad de la especificación de ingesta masiva para evitar pérdidas de datos debido a problemas de hardware o software. Si necesita recuperar la especificación, puede importar el archivo de almacenamiento de la aplicación a un Repositorio de modelos.

Para obtener más información sobre `infacmd`, consulte la *Referencia de comandos de Informatica*.

Para obtener más información sobre los archivos de almacenamiento de la aplicación, consulte la *Guía de Informatica Developer tool*.

## Migrar una especificación de ingesta masiva

Migre una especificación de ingesta masiva para moverla a un Repositorio de modelos en un entorno diferente. Por ejemplo, puede que desee migrar una especificación de ingesta masiva entre entornos de desarrollo, prueba y producción.

Cuando una especificación de ingesta masiva se migra, los nombres de conexión de origen y destino y los ID deben ser los mismos entre los entornos.

Dependiendo del estado de los repositorios de origen y de destino, usará un método u otro para migrar una especificación de ingesta masiva. El repositorio de origen es el Repositorio de modelos que contiene las especificaciones de ingesta masiva que desea migrar. El repositorio de destino es el Repositorio de modelos donde desea migrar las especificaciones de ingesta masiva.

Elija un método para migrar una especificación de ingesta masiva dependiendo de uno de los siguientes estados de repositorio:

- El repositorio de destino está vacío y desea migrar todo el contenido del Repositorio de modelos. Para migrar el contenido, haga una copia de seguridad de este en el repositorio de origen y restablezca el contenido en el repositorio de destino.
- El repositorio de destino está vacío y desea migrar solo las especificaciones de ingesta masiva. Antes de migrar las especificaciones de ingesta masiva, cree un servicio de ingesta masiva y asócielo al repositorio de destino. El servicio de ingesta masiva crea un proyecto `INFA_MI_PROJECT` en el repositorio de destino, donde puede importar las especificaciones de ingesta masiva. Para migrar una especificación de ingesta masiva, expórtela desde el proyecto `INFA_MI_PROJECT` en el repositorio de origen utilizando

infacmd tools exportObjects. Importe la especificación de ingesta masiva al proyecto `INFA_MI_PROJECT` en el repositorio de destino utilizando `infacmd tools importObjects`.

- El repositorio de destino incluye contenido de ingesta masiva y quiere actualizar las especificaciones de ingesta masiva en dicho repositorio de destino. Para migrar una especificación de ingesta masiva, expórtela desde el proyecto `INFA_MI_PROJECT` en el repositorio de origen utilizando `infacmd tools exportObjects`. Importe la especificación de ingesta masiva al proyecto `INFA_MI_PROJECT` en el repositorio de destino utilizando `infacmd tools importObjects`, con la opción [`<-ConflictResolution|-cr>`] para reemplazar o cambiar de nombre las especificaciones en el repositorio de destino.

Después de migrar una especificación de ingesta masiva a otro entorno, impleméntela en un servicio de integración de datos.

Para obtener más información sobre `infacmd tools`, vea el capítulo "infacmd tools" en la *Referencia de comandos de Informatica*.

Para obtener más información sobre cómo exportar e importar contenido en el Repositorio de modelos, vea el capítulo "Importar y exportar objetos" en la *Guía de Informatica Developer tool*.

# CAPÍTULO 5

## Run

Este capítulo incluye los siguientes temas:

- [Resumen de supervisión, 55](#)
- [Tipos de carga, 55](#)
- [Ver instancias de ejecución, 57](#)
- [Ejecutar una especificación de ingesta masiva, 58](#)

## Resumen de supervisión

Tras implementar una especificación de ingesta masiva, puede ejecutarla para ingerir los datos.

Cuando ejecute la especificación, puede seleccionar uno de los siguientes tipos de carga:

- **Carga completa.** Una carga completa ingiere todos los datos en el destino. Cuando se usa una carga completa, el motor de Spark elimina los datos existentes en el destino de HDFS o Hive y los reemplaza por los que se hayan configurado en la especificación.
- **Carga incremental.** Una carga incremental ingiere solo datos incrementales en el destino. El motor de Spark anexa los datos incrementales al destino o directamente sobrescribe el destino con los datos incrementales, dependiendo del modo que se haya configurado en la especificación.

Cada vez que la especificación se ejecuta, el servicio de ingesta masiva genera una nueva instancia de ejecución de la especificación. El servicio de integración de datos procesa la instancia y se conecta al entorno de Hadoop. En el entorno de Hadoop, el motor de Spark ejecuta los trabajos de asignación de ingesta individuales que ingieren datos de las tablas de origen al destino.

Si los trabajos de ingesta dejan de responder o tardan mucho tiempo en completarse, se pueden cancelar en la herramienta de supervisión.

Para obtener información sobre cómo cancelar trabajos de ingesta, vea ["Cancelar trabajos de ingesta" en la página 75](#) en el capítulo de supervisión.

## Tipos de carga

Ejecute una especificación de ingesta masiva usando una carga completa o una carga incremental. Puede decidir si quiere utilizar una carga completa o incremental dependiendo del estado de los datos en la base de datos relacional y en el destino.

## Carga completa

Utilice una carga completa para ingerir todos los datos de la especificación de ingesta masiva en el destino. Cuando se usa una carga completa, los datos existentes en el destino de Hive o de HDFS se borran y se reemplazan por los datos de las tablas de origen.

Puede que quiera ejecutar una carga completa por alguno de los siguientes motivos:

### **Como requisito previo para ejecutar cargas incrementales.**

Cuando cree una especificación de ingesta masiva, ejecute una carga completa inicial antes de comenzar a ejecutar cargas incrementales en los datos. La carga completa inicial permite que el motor de Spark cree una base de la que obtener los datos incrementales en las ejecuciones subsiguientes.

Una carga completa inicial también puede ayudar a los administradores a mantener registros de documentación propia. Por ejemplo, puede ejecutar una carga incremental utilizando el modo de sobrescritura como primera ejecución de la especificación, pero el motor de Spark carecerá de una base de la que obtener datos incrementales. Como resultado, el motor de Spark ingerirá todos los datos del origen y llevará a cabo una carga completa de facto. Los registros indicarían que un usuario ejecutó una carga incremental, pero puede que no esté del todo claro si se han ingerido todos los datos o solo los datos incrementales en el destino.

Si ejecuta una carga completa inicial seguida de cargas incrementales subsiguientes, el administrador podrá distinguir si el motor de Spark ha ingerido todos los datos o solo los datos incrementales en cada ejecución de la especificación.

### **Para actualizar la base de cargas incrementales.**

Ejecute una carga completa para actualizar el destino con instrucciones UPSERT y DELETE que se han ejecutado en la base de datos relacional.

Si ejecuta una carga incremental, el motor de Spark obtiene las filas que se han añadido a una tabla relacional usando instrucciones INSERT. El motor de Spark no puede obtener las filas que se han cambiado mediante instrucciones UPSERT y DELETE, de modo que es posible que una carga incremental de una base de datos relacional no constituya una representación exacta de los datos de origen.

## Carga incremental

Utilice una carga incremental para ingerir solo datos incrementales en el destino. Para poder utilizar una carga incremental, debe habilitar las cargas incrementales y configurar opciones de carga incremental en la especificación de ingesta masiva.

Puede que quiera utilizar una carga incremental porque una carga completa puede requerir una gran cantidad de tiempo y recursos, sobre todo cuando los datos no se han modificado en gran medida entre las ejecuciones de una especificación. Para disfrutar de una solución más rentable, puede utilizar una carga incremental para ingerir solo los datos incrementales cada vez que la especificación se ejecute.

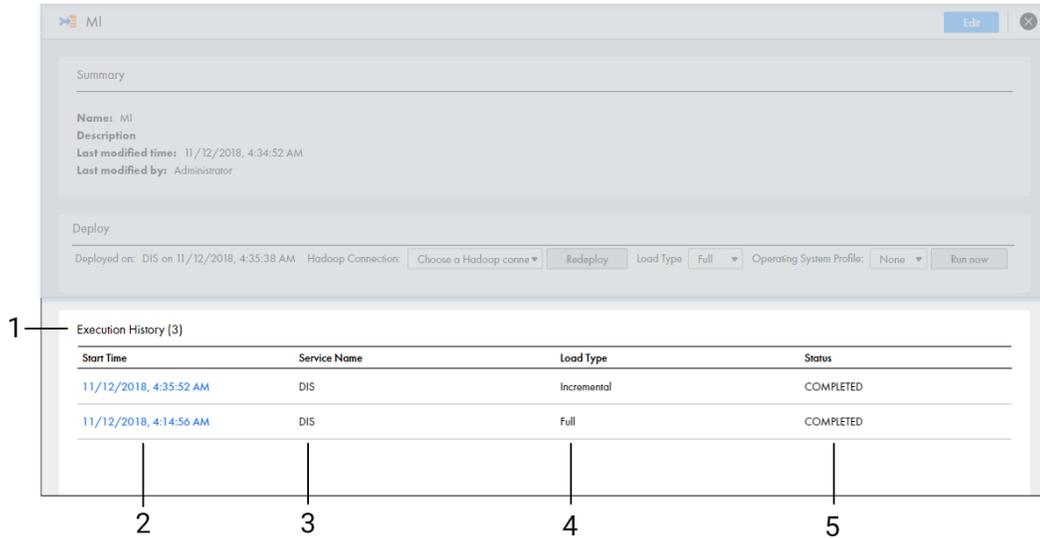
Ejecute una carga completa inicial antes de empezar a ejecutar cargas incrementales en una especificación. Una carga completa inicial permite que el motor de Spark conserve internamente los datos de origen iniciales y utilice esos valores conservados como base de la que obtener los datos incrementales de una carga incremental. Una vez que el motor de Spark conserve los datos de origen iniciales, podrá ingerir datos incrementales basados en los datos de origen iniciales en cualquier destino.

**Nota:** Si ejecuta una carga incremental usando el modo de anexión, el destino debe contener archivos o tablas existentes para que el motor de Spark pueda anexar datos incrementales a los datos existentes. Para crear archivos o tablas, ejecute una carga completa.

# Ver instancias de ejecución

Cada vez que se ejecuta una especificación de ingesta masiva, el servicio de ingesta masiva genera una nueva instancia de ejecución de la especificación. Las instancias de ejecución se pueden ver en la vista Historial de ejecución de la página Resumen de la herramienta de ingesta masiva.

En la siguiente imagen se muestran varias instancias de ejecución en la vista Historial de ejecución:



1. Vista Historial de ejecución. Muestra las instancias de ejecución de la especificación de ingesta masiva.
2. Hora de inicio. Muestra la hora de inicio de la instancia de ejecución.
3. Nombre de servicio. Muestra el servicio de integración de datos en el que está implementada la especificación de ingesta masiva.
4. Tipo de carga. Muestra el tipo de carga que el motor de Spark ha usado para ingerir los datos.
5. Estado. Muestra el estado de la instancia de ejecución.

También se puede usar la línea de comandos para mostrar las instancias de ejecución de una especificación de ingesta masiva.

Para obtener más información, vea ["listSpecRuns" en la página 87](#) en el apéndice "Referencia de comandos infacmd mi".

## Estado de instancia de ejecución

El estado de la instancia de ejecución puede aparecer como Completado, Error, En ejecución, Anulado, Cancelado, En cola o Desconocido.

En la siguiente tabla se describe el estado que puede aparecer en cada instancia de ejecución:

Estado	Descripción
Completado	La instancia de ejecución de la especificación de ingesta masiva se ha completado. Todos los datos se han ingerido en el destino correctamente.
Error	La instancia de ejecución de la especificación de ingesta masiva ha generado errores. Algunos datos no se han podido ingerir en el destino.
En ejecución	La instancia de ejecución de la especificación de ingesta masiva está en curso. Se están ingiriendo datos en el destino.

Estado	Descripción
Anulada	La instancia de ejecución de la especificación de ingesta masiva se ha anulado. La instancia de ejecución se ha anulado desde la línea de comandos.
Cancelado	La instancia de ejecución de la especificación de ingesta masiva se ha cancelado. Los trabajos para ingerir algunas tablas de origen en la especificación de ingesta masiva se han cancelado en el servicio de integración de datos.
En cola	La instancia de ejecución de la especificación de ingesta masiva se ha puesto en cola. Los trabajos para ingerir algunas tablas de origen en la especificación de ingesta masiva se han puesto en cola en el servicio de integración de datos.
Desconocido	El estado de la instancia de ejecución de la especificación de ingesta masiva es desconocido. El servicio de ingesta masiva no puede obtener el estado de la instancia de ejecución. El estado de la instancia de ejecución puede mostrar <b>Desconocido</b> en las siguientes circunstancias: <ul style="list-style-type: none"> <li>- No hay un Repositorio de modelos configurado en las opciones de supervisión del servicio de ingesta masiva.</li> <li>- El servicio de integración de datos está deshabilitado.</li> </ul>

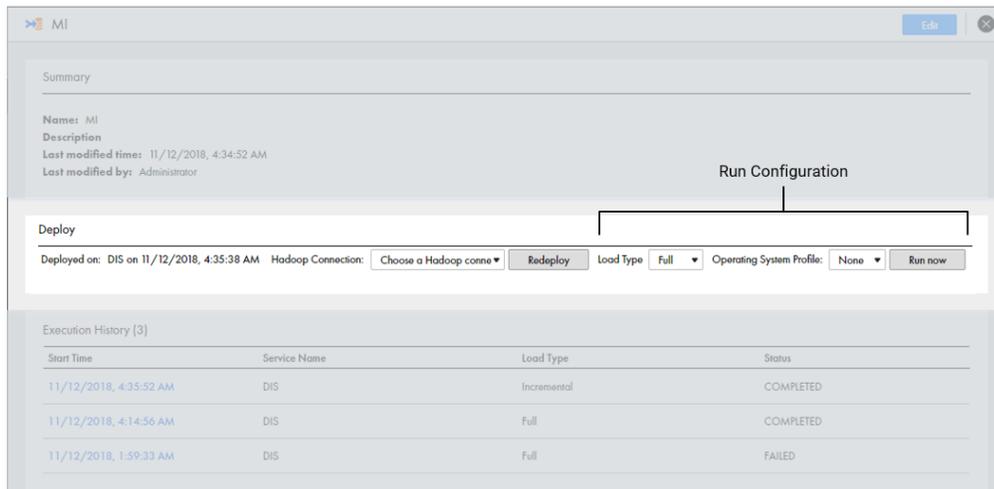
## Ejecutar una especificación de ingesta masiva

Ejecute una especificación de ingesta masiva para ingerir los datos de las tablas de origen en el destino de HDFS o Hive.

1. En la herramienta de ingesta masiva, vaya a una especificación de ingesta masiva.

Se abre la página **Resumen**.

En la siguiente imagen se muestra la configuración de ejecución en la vista **Implementar** de la página **Resumen**:



2. En la vista **Implementar**, seleccione el tipo de carga. Puede seleccionar **Completa** o **Incremental**. El valor predeterminado es **Completa**.

**Nota:** Si la carga incremental no está habilitada en la especificación, no se podrá utilizar. Para poder utilizar una carga incremental, edite la especificación. En ella, habilite la carga incremental y configure las opciones de carga incremental correspondientes.

3. Seleccione un perfil del sistema operativo. El valor predeterminado es ninguno.
4. Haga clic en **Ejecutar** para ejecutar la especificación de ingesta masiva.

En la vista **Historial de ejecución** se muestra una instancia de ejecución de la especificación de ingesta masiva.

# CAPÍTULO 6

## Supervisor

Este capítulo incluye los siguientes temas:

- [Resumen de supervisión, 60](#)
- [Supervisar la herramienta de ingesta masiva, 60](#)
- [Supervisar la Herramienta del administrador, 72](#)
- [Solución de problemas de trabajos de ingesta masiva, 76](#)

### Resumen de supervisión

Cuando ejecuta la especificación de ingesta masiva, puede comenzar a supervisar las estadísticas de trabajo. Supervise las estadísticas de trabajo en la herramienta de ingesta masiva o en la Herramienta del administrador.

### Supervisar la herramienta de ingesta masiva

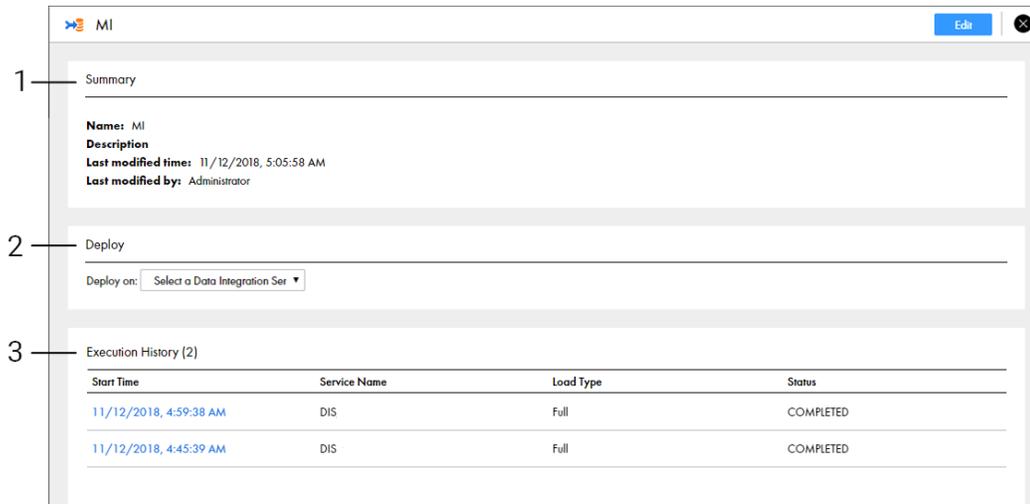
Puede supervisar una especificación de ingesta masiva en la herramienta de ingesta masiva. Supervise los detalles de la especificación de ingesta masiva y el historial de especificaciones en la página Resumen, y supervise estadísticas de trabajos de ingesta en la página Estadísticas de ejecución.

La página Resumen proporciona detalles generales sobre la especificación, el historial de implementación y el historial de ejecución. La página Estadísticas de ejecución proporciona estadísticas detalladas sobre el trabajo de ingesta de una instancia de ejecución de especificación.

## Resumen Página

En la página Resumen, revise detalles sobre una especificación de ingesta masiva, las opciones de ejecución e implementación y el historial de ejecución.

En la siguiente imagen se muestra la página Resumen:



1. Resumen. Muestra las propiedades generales de especificación de ingesta masiva.
2. Implementar. Muestra opciones para implementar la especificación de ingesta masiva.
3. Historial de ejecución. Muestra las instancias de ejecución de la especificación de ingesta masiva.

## Vista de resumen

La vista de resumen muestra las propiedades generales de especificación de ingesta masiva.

En la siguiente tabla se describen las propiedades de especificación que puede revisar:

Propiedad	Descripción
Nombre	Nombre de la especificación de ingesta masiva.
Descripción	Descripción de la especificación de ingesta masiva.
Hora de última modificación	Hora a la que la especificación de ingesta masiva se modificó por última vez.
Última modificación por	Último usuario que modificó la especificación de ingesta masiva.

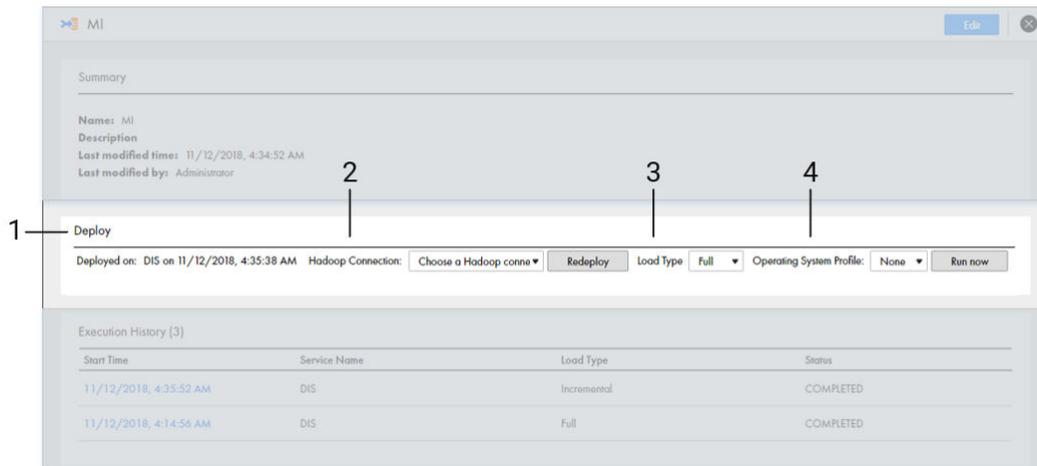
## Vista Implementar

La vista Implementar muestra opciones para implementar y ejecutar la especificación de ingesta masiva.

Para implementar la especificación, puede especificar un servicio de integración de datos y una conexión de Hadoop. Tras implementar una especificación de ingesta masiva, puede volver a implementarla utilizando una conexión de Hadoop diferente o en un servicio de integración de datos distinto.

El servicio de integración de datos utiliza la conexión de Hadoop para conectarse al entorno de Hadoop. Cuando la especificación de ingesta masiva se implemente, se puede ejecutar desde la vista Implementar utilizando opciones para seleccionar un tipo de carga y un perfil de sistema operativo.

La siguiente imagen muestra la vista Implementar tras haber implementado la especificación de ingesta masiva:



1. Implementado en. Muestra el nombre del servicio de integración de datos donde la especificación se ha implementado y la hora a la que se ha implementado.
2. Conexión de Hadoop. Opción para seleccionar una conexión de Hadoop para volver a implementar la especificación.
3. Tipo de carga. Opción para seleccionar un tipo de carga para definir cómo el motor de Spark ingiere los datos de origen cuando la especificación se ejecute. El valor predeterminado es Completa.
4. Perfil de sistema operativo. Opción para seleccionar un perfil de sistema operativo cuando la especificación se ejecute. El valor predeterminado es Ninguno.

## Vista Historial de ejecución

La vista Historial de ejecución muestra las instancias de ejecución de la especificación de ingesta masiva e información relativa a cada instancia de ejecución.

En la siguiente tabla se describe la información de instancia de ejecución que se puede ver en la vista Historial de ejecución:

Propiedad	Descripción
Hora de inicio	Hora a la que se ha iniciado la instancia de ejecución.
Nombre de servicio	Nombre del servicio de integración de datos en el que se implementa la especificación de ingesta masiva.
Tipo de carga	Tipo de carga que el motor de Spark ha usado para ingerir los datos en la especificación de ingesta masiva. El tipo de carga puede ser Completo o Incremental.
Estado	Estado de la instancia de ejecución. El estado puede ser Completado, Error, En ejecución, Anulado, Cancelado, En cola o Desconocido.

En la siguiente tabla se describe el estado que puede aparecer en cada instancia de ejecución:

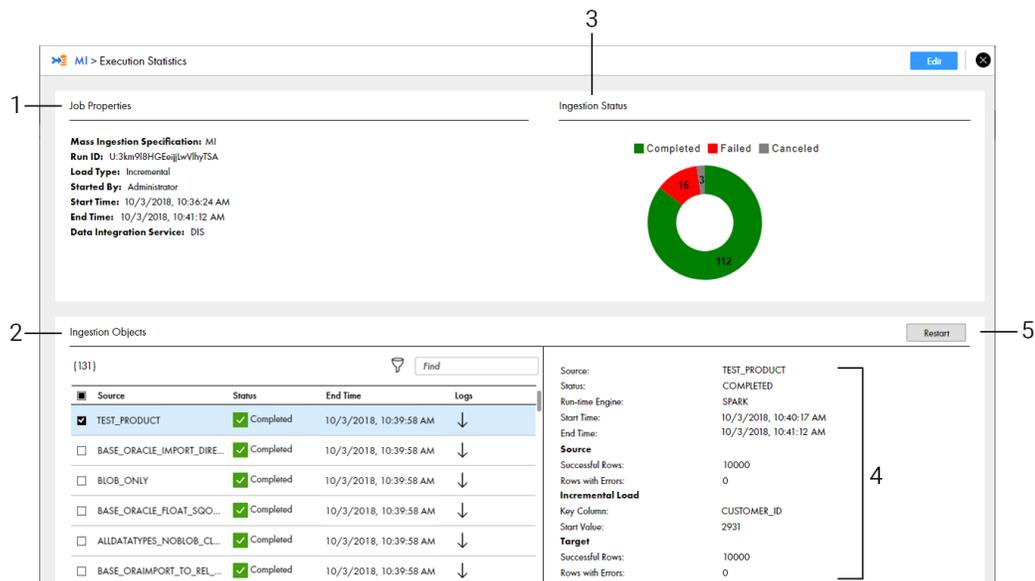
Estado	Descripción
Completado	La instancia de ejecución de la especificación de ingesta masiva se ha completado. Todos los datos se han ingerido en el destino correctamente.
Error	La instancia de ejecución de la especificación de ingesta masiva ha generado errores. Algunos datos no se han podido ingerir en el destino.
En ejecución	La instancia de ejecución de la especificación de ingesta masiva está en curso. Se están ingiriendo datos en el destino.
Anulada	La instancia de ejecución de la especificación de ingesta masiva se ha anulado. La instancia de ejecución se ha anulado desde la línea de comandos.
Cancelado	La instancia de ejecución de la especificación de ingesta masiva se ha cancelado. Los trabajos para ingerir algunas tablas de origen en la especificación de ingesta masiva se han cancelado en el servicio de integración de datos.
En cola	La instancia de ejecución de la especificación de ingesta masiva se ha puesto en cola. Los trabajos para ingerir algunas tablas de origen en la especificación de ingesta masiva se han puesto en cola en el servicio de integración de datos.
Desconocido	<p>El estado de la instancia de ejecución de la especificación de ingesta masiva es desconocido. El servicio de ingesta masiva no puede obtener el estado de la instancia de ejecución.</p> <p>El estado de la instancia de ejecución puede mostrar <b>Desconocido</b> en las siguientes circunstancias:</p> <ul style="list-style-type: none"> <li>- No hay un Repositorio de modelos configurado en las opciones de supervisión del servicio de ingesta masiva.</li> <li>- El servicio de integración de datos está deshabilitado.</li> </ul>

Si hace clic en la hora de inicio de una instancia de ejecución en la vista Historial de ejecución, aparece la página Estadísticas de ejecución de la ejecución de especificación en cuestión. En la página Estadísticas de ejecución puede seguir de cerca las estadísticas del trabajo de ingesta.

## Página Estadísticas de ejecución

En la página Estadísticas de ejecución puede ver información detallada sobre el trabajo de ingesta de una instancia de ejecución de la especificación de ingesta masiva.

En la siguiente imagen se muestra la página Estadísticas de ejecución:



1. Propiedades del trabajo. Muestra las propiedades generales del trabajo de ingesta.
2. Objetos de ingesta. Muestra las tablas de origen que el trabajo ingiere en el destino.
3. Estado de ingesta. Muestra una representación gráfica del trabajo de ingesta.
4. Estadísticas de ingesta. Muestra las estadísticas de trabajo de ingesta de una tabla de origen específica.
5. Reiniciar. Reinicia el trabajo de ingesta de los objetos de ingesta.

## Vista Propiedades del trabajo

La vista Propiedades del trabajo muestra las propiedades generales que describen la instancia de ejecución de la especificación de ingesta masiva.

En la siguiente tabla se describen las propiedades que se pueden ver en la vista Propiedades del trabajo:

Propiedad	Descripción
Especificación de ingesta masiva	Nombre de la especificación de ingesta masiva.
ID de ejecución	Número de identificador de ejecución (o ID de ejecución) de la instancia de ejecución de especificación de ingesta masiva.
Tipo de carga	Tipo de carga que el motor de Spark ha usado para ingerir los datos. El tipo de carga puede ser Completo o Incremental.
Iniciada por	Usuario que ha iniciado la ejecución de la especificación de ingesta masiva.
Hora de inicio	Hora a la que se ha iniciado el trabajo de ingesta.

Propiedad	Descripción
Hora de finalización	Hora a la que ha finalizado el trabajo de ingesta.
Servicio de integración de datos	Nombre del servicio de integración de datos en el que se implementa la especificación de ingesta masiva.

## Vista Objetos de ingesta

La vista Objetos de ingesta muestra las tablas de origen que forman parte del trabajo de ingesta.

Utilice la vista Objetos de ingesta para supervisar las tablas que se ingieren con correctamente y las que no. Si algunas tablas de origen generan errores de ingesta, reinicie el proceso de ingesta de dichas tablas.

En la siguiente tabla se describen las propiedades que se muestran en la vista Objetos de ingesta:

Propiedad	Descripción
Origen	Nombre de la tabla de origen configurada en la especificación de ingesta masiva.
Estado	Estado de ingesta de la tabla de origen. El estado puede ser Completado, Error, En ejecución, Cancelado, En cola, Desconocido o Realizando la operación.
Hora de finalización	Hora a la que se completa el trabajo de ingesta de la tabla de origen.
Registros	Opción de descarga del registro de ingesta de la tabla de origen.

En la siguiente tabla se describe el estado que puede aparecer en cada tabla de origen:

Estado	Descripción
Completado	El trabajo de ingesta de la tabla de origen se ha completado. Todos los datos en la tabla de origen se han ingerido en el destino correctamente.
Error	El trabajo de ingesta de la tabla de origen ha generado errores. Los datos de la tabla de origen no se han podido ingerir en el destino.
En ejecución	El trabajo de ingesta de la tabla de origen está en curso. Se están ingiriendo datos en el destino.
Anulada	El trabajo de ingesta de la tabla de origen se ha anulado. La instancia de ejecución de la especificación de ingesta masiva que contiene el trabajo de ejecución se ha anulado desde la línea de comandos.
Cancelado	El trabajo de ingesta de la tabla de origen se ha cancelado. La asignación implementada que lleva a cabo el trabajo de ingesta se ha cancelado en el servicio de integración de datos.
En cola	El trabajo de ingesta de la tabla de origen se ha puesto en cola. El servicio de ingesta masiva está esperando para programar la asignación implementada que lleva a cabo el trabajo de ingesta para ejecutarlo.

Estado	Descripción
Desconocido	<p>El estado del trabajo de ingesta de la tabla de origen es desconocido. El servicio de ingesta masiva no puede obtener el estado de la tabla de origen.</p> <p>El estado de la instancia de ejecución puede mostrar <i>Desconocido</i> en las siguientes circunstancias:</p> <ul style="list-style-type: none"> <li>- No hay un Repositorio de modelos configurado en las opciones de supervisión del servicio de ingesta masiva.</li> <li>- El servicio de integración de datos está deshabilitado.</li> </ul>
Realizando la operación	El servicio de ingesta masiva se encuentra obteniendo el estado de la tabla de origen del servicio de integración de datos.

## Vista Estado de ingesta

La vista Estado de ingesta muestra una representación gráfica del trabajo de ingesta masiva.

El gráfico resume el estado de ingesta que se muestra en la vista Objetos de ingesta. El estado de cada tabla de origen puede ser Completado, Error, En ejecución, Cancelado, En cola, Desconocido o Realizando la operación.

Teniendo como referencia el estado de ingesta, puede decidir si modificar la especificación de ingesta masiva. Si modifica la especificación y la vuelve a ejecutar, revise las estadísticas en una página Estadísticas de ejecución única de la instancia de ejecución generada.

## Vista Estadísticas de ingesta

La vista Estadísticas de ingesta muestra las estadísticas de trabajo de ingesta de una tabla de origen específica.

En la siguiente tabla se describen las propiedades que se muestran en la vista Estadísticas de ingesta:

Propiedad	Descripción
Origen	Nombre de la tabla de origen que se ingiere.
Estado	Estado de ingesta de la tabla de origen. El estado puede ser Completado, Error, En ejecución, Cancelado, En cola, Desconocido o Realizando la operación.
Motor de tiempo de ejecución	Motor de tiempo de ejecución en el entorno de Hadoop que lleva a cabo el trabajo de ingesta. El motor de tiempo de ejecución es Spark.
Hora de inicio	Hora a la que se inicia el trabajo de ingesta de la tabla de origen.
Hora de finalización	Hora a la que se completa el trabajo de ingesta de la tabla de origen.
Origen	Estadísticas de las filas de tabla de la base de datos de origen. Se puede ver número de filas que se ha ingerido correctamente y el número de filas que contiene errores.

Propiedad	Descripción
Carga incremental	<p>Detalles que describen los datos incrementales que se han ingerido. Los detalles se muestran si el motor de Spark ha usado una carga incremental para ingerir los datos. Puede ver el nombre de la columna que el motor de Spark ha usado para obtener datos incrementales en la tabla de origen y el valor inicial que determina el valor que el motor de Spark ha empleado para comenzar a ingerir los datos incrementales.</p> <p>Por ejemplo, si la clave incremental es una columna ID y el valor inicial es 2500, quiere decir que el motor de Spark ha ingerido las filas que tienen una ID igual o mayor a 2500.</p>
Target	<p>Estadísticas de las filas de tabla que se ingieren en el destino. Se puede ver número de filas que se ha ingerido correctamente y el número de filas que contiene errores.</p>

**Nota:** Para ver unas estadísticas determinadas, debe habilitar la supervisión de Spark. La supervisión de Spark puede afectar al rendimiento.

## Supervisar estadísticas de ejecución

Supervise las estadísticas de ejecución de una instancia de ejecución de especificación de ingesta masiva en la página Estadísticas de ejecución.

1. Vaya a una especificación de ingesta masiva en la herramienta de ingesta masiva.  
Se abre la página **Resumen**.
2. Seleccione una instancia de ejecución en la vista Historial de ejecución para ver las estadísticas del trabajo de ingesta.  
Se abre la página **Estadísticas de ejecución** de la instancia de ejecución que seleccione.

- En la página **Estadísticas de ejecución**, seleccione una tabla de origen en la vista Objetos de ingesta para ver estadísticas del trabajo que ingiere dicha tabla de origen.

Las estadísticas aparecen en la vista Estadísticas de ingesta. Puede ver las estadísticas de ingesta de una tabla individualmente.

Por ejemplo, en la siguiente imagen se muestran las diferentes tablas de origen que pueden aparecer en la vista Objetos de ingesta:

Job Properties

Mass Ingestion Specifications: test  
Run ID: U:2zHDO26EaxXCOpk51DGDw  
Lead Type: Full  
Started By: Administrator  
Start Time: 11/12/2018, 4:59:38 AM  
End Time: 11/12/2018, 5:02:04 AM  
Data Integration Service: DIS

Ingestion Status

Completed

Ingestion Objects

(2)

Source	Status	End Time	Logs
<input type="checkbox"/> TEST_PRODUCT	Completed	11/12/2018, 5:02:04 AM	↓
<input type="checkbox"/> TEST_OPPORTUNITY	Completed	11/12/2018, 5:02:04 AM	↓

Select a table to view job statistics.

En la siguiente imagen se muestran las estadísticas que pueden aparecer en la vista Estadísticas de ingesta de la primera tabla de origen en la vista Objetos de ingesta:

Job Properties

Mass Ingestion Specifications: test  
Run ID: U:2zHDO26EaxXCOpk51DGDw  
Lead Type: Full  
Started By: Administrator  
Start Time: 11/12/2018, 4:59:38 AM  
End Time: 11/12/2018, 5:02:04 AM  
Data Integration Service: DIS

Ingestion Status

Completed

Ingestion Objects

(2)

Source	Status	End Time	Logs
<input checked="" type="checkbox"/> TEST_PRODUCT	Completed	11/12/2018, 5:02:04 AM	↓
<input type="checkbox"/> TEST_OPPORTUNITY	Completed	11/12/2018, 5:02:04 AM	↓

Source: TEST\_PRODUCT  
Status: COMPLETED  
Run-time Engine: SPARK  
Start Time: 11/12/2018, 4:59:38 AM  
End Time: 11/12/2018, 5:02:04 AM

**Source**  
Successful Rows: 502  
Rows with Errors: 0

**Target**  
Successful Rows:  
Rows with Errors:

## Filtrar objetos de ingesta por estado

Filtre objetos de ingesta en la página Estadísticas de ejecución para encontrar las tablas específicas que quiera supervisar. Los objetos de ingesta se pueden filtrar por el estado de ingesta.

1. En la vista Objetos de ingesta de la página Estadísticas de ejecución, seleccione el icono de **filtro**.

En la siguiente imagen se muestra el icono de **filtro** de la vista Objetos de ingesta:

Job Properties

Mass Ingestion Specifications: test  
Started By: Administrator  
Start Times: 3/12/2018, 1:00:52 PM  
End Times:  
Data Integration Services: DIS

Ingestion Status

Failed Canceled Unknown

125

Filter

Ingestion Objects

(133)

Source	Status	End Time	Logs
<input type="checkbox"/> ALLDATAR1	Failed	3/12/2018, 1:08:44 PM	↓
<input type="checkbox"/> ALLDATAR2	Unknown		
<input type="checkbox"/> ALLDATAR3	Unknown		
<input type="checkbox"/> ALLDATATYPES	Unknown		
<input type="checkbox"/> ALLDATATYPES_1_SRC	Unknown		
<input type="checkbox"/> ALLDATATYPES_1_TGT	Unknown		
<input type="checkbox"/> ALLDATATYPES_2_SRC	Failed	3/12/2018, 1:09:18 PM	↓
<input type="checkbox"/> ALLDATATYPES_2_TGT	Unknown		

Select a table to view job statistics.

Restart

2. Seleccione **Estado** para filtrar los objetos de ingesta por el estado de ingesta.

En la siguiente imagen se muestra la opción **Estado** de la vista Objetos de ingesta:

Ingestion Objects

(133 of 133)

Find

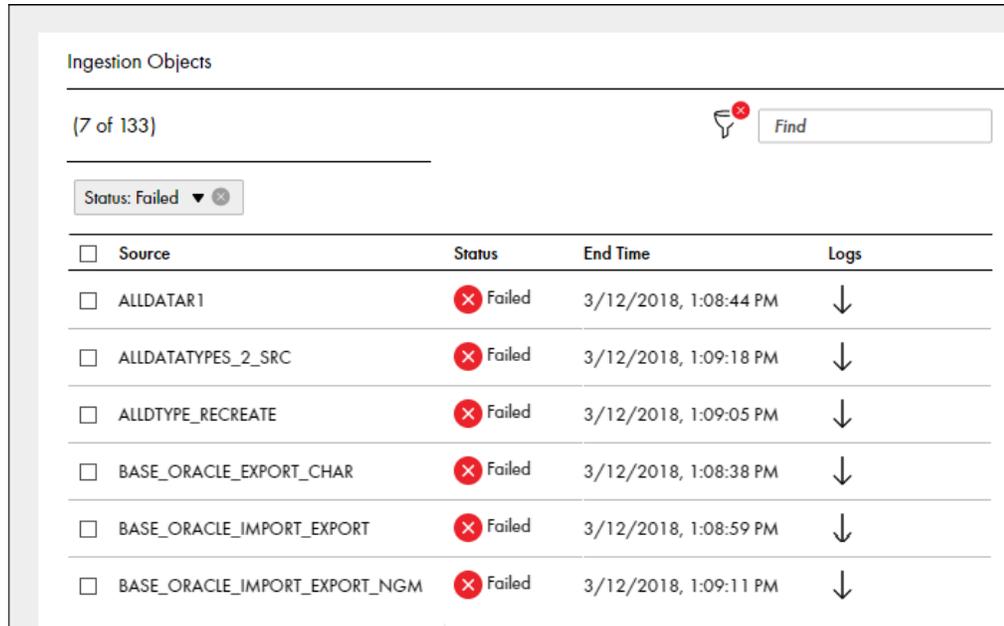
Status

- Completed
- Failed
- Running
- Aborted
- Canceled
- Queued
- Unknown
- Working

Source	Status	End Time	Logs
<input type="checkbox"/> ALLDATAR1	Failed	3/12/2018, 1:08:44 PM	↓
<input type="checkbox"/> ALLDATAR2	Unknown		
<input type="checkbox"/> ALLDATAR3	Unknown		
<input type="checkbox"/> ALLDATATYPES	Unknown		
<input type="checkbox"/> ALLDATATYPES_1_SRC	Unknown		
<input type="checkbox"/> ALLDATATYPES_1_TGT	Unknown		
<input type="checkbox"/> ALLDATATYPES_2_SRC	Failed	3/12/2018, 1:09:18 PM	↓

3. Seleccione el estado que desee utilizar para filtrar los objetos de ingesta.

Por ejemplo, puede seleccionar **Error**. La siguiente imagen muestra los objetos de ingesta filtrados por Error en la vista Objetos de ingesta:



Ingestion Objects

(7 of 133) 

Status: Failed ▼

<input type="checkbox"/>	Source	Status	End Time	Logs
<input type="checkbox"/>	ALLDATAR1	 Failed	3/12/2018, 1:08:44 PM	↓
<input type="checkbox"/>	ALLDATATYPES_2_SRC	 Failed	3/12/2018, 1:09:18 PM	↓
<input type="checkbox"/>	ALLDTYPE_RECREATE	 Failed	3/12/2018, 1:09:05 PM	↓
<input type="checkbox"/>	BASE_ORACLE_EXPORT_CHAR	 Failed	3/12/2018, 1:08:38 PM	↓
<input type="checkbox"/>	BASE_ORACLE_IMPORT_EXPORT	 Failed	3/12/2018, 1:08:59 PM	↓
<input type="checkbox"/>	BASE_ORACLE_IMPORT_EXPORT_NGM	 Failed	3/12/2018, 1:09:11 PM	↓

## Reiniciar trabajos de ingesta

Después de ejecutar la especificación de ingesta masiva, puede reiniciar los trabajos de ingesta de tablas de origen específicas.

Por ejemplo, puede que quiera reiniciar el trabajo de ingesta de una tabla de origen que ha generado errores de ingesta. Las tablas de origen pueden no ingerirse debido a una configuración de Hadoop que no es compatible con el trabajo de ingesta. También puede ocurrir que el destino de ingesta no tenga suficiente espacio para consumir las tablas ingeridas. Puede comprobar la configuración de Hadoop o reconfigurar el destino de ingesta. Luego, puede reiniciar los trabajos de ingesta de las tablas con errores de ingesta.

Para reiniciar los trabajos de ingesta, realice una de las siguientes tareas:

- Reiniciar los objetos seleccionados. Reinicie el trabajo de ingesta correspondiente únicamente a las tablas de origen que seleccione.
- Reiniciar los objetos con errores. Reinicie los trabajos de ingesta de todas las tablas de origen con errores de ingesta y todos los trabajos de ingesta que estén cancelados.

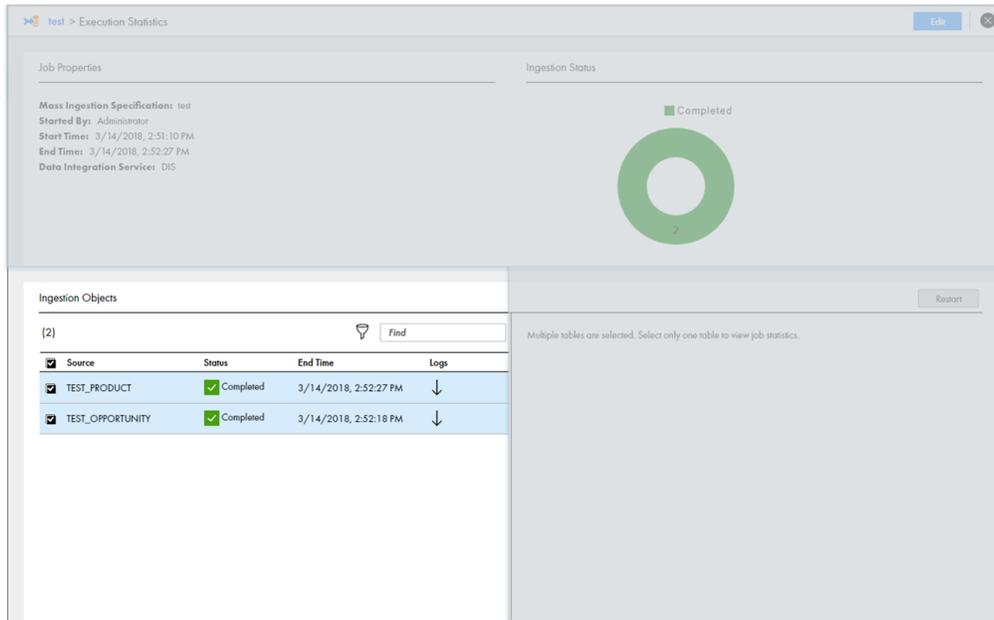
Cuando los trabajos de ingesta se reinician, no se vuelve a implementar o ejecutar la especificación de ingesta masiva completa. El estado de ingesta del trabajo se actualiza y se muestra como Realizando la operación. El estado seguirá mostrando Realizando la operación hasta que el servicio de ingesta masiva obtenga el estado de ingesta del servicio de integración de datos.

## Reiniciar objetos de ingesta seleccionados

Reinicie los objetos de ingesta seleccionados en la especificación de ingesta masiva para reiniciar el trabajo de ingesta de los objetos.

1. En la vista Objetos de ingesta de la página Estadísticas de ejecución, seleccione los objetos de ingesta que quiera reiniciar.

En la siguiente imagen se muestran varios objetos que están seleccionados:

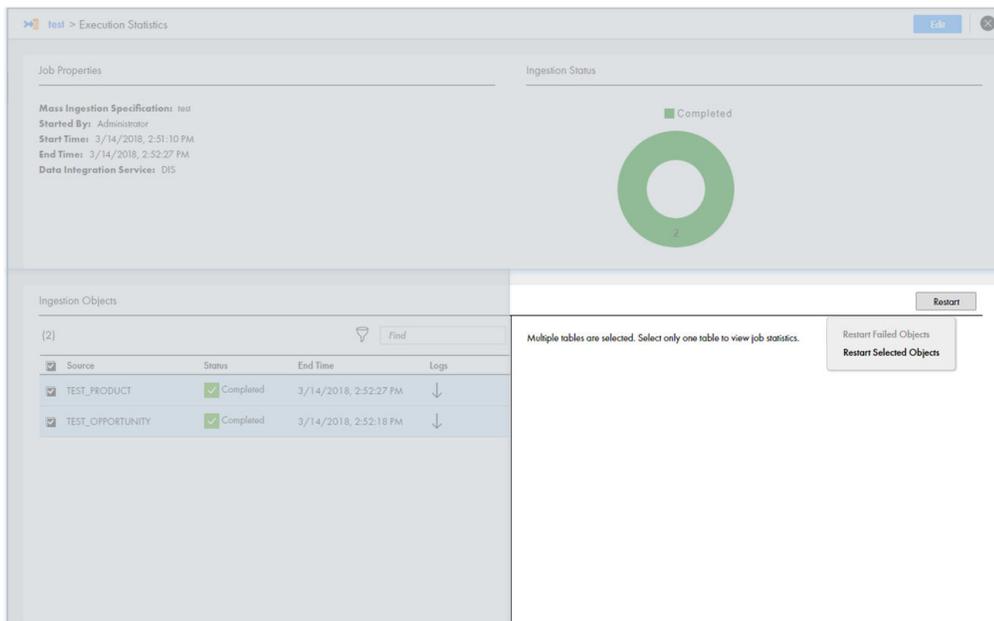


The screenshot shows the 'Execution Statistics' page for a job named 'test'. The 'Ingestion Status' section displays a green donut chart with the number '2' inside, indicating that 2 objects are completed. The 'Ingestion Objects' table has two rows selected, both with a green checkmark in the 'Status' column.

Source	Status	End Time	Logs
TEST_PRODUCT	Completed	3/14/2018, 2:52:27 PM	↓
TEST_OPPORTUNITY	Completed	3/14/2018, 2:52:18 PM	↓

2. En la vista Estadísticas de ingesta, seleccione **Reiniciar**.

En la siguiente imagen se muestra la opción **Reiniciar**:



The screenshot shows the same 'Execution Statistics' page. The 'Restart' button in the top right corner of the 'Ingestion Objects' section is highlighted. A tooltip is visible over the button, containing the text 'Restart Failed Objects' and 'Restart Selected Objects'.

3. Seleccione **Reiniciar objetos seleccionados**.

De este modo, se reinicia el trabajo de ingesta de los objetos seleccionados.

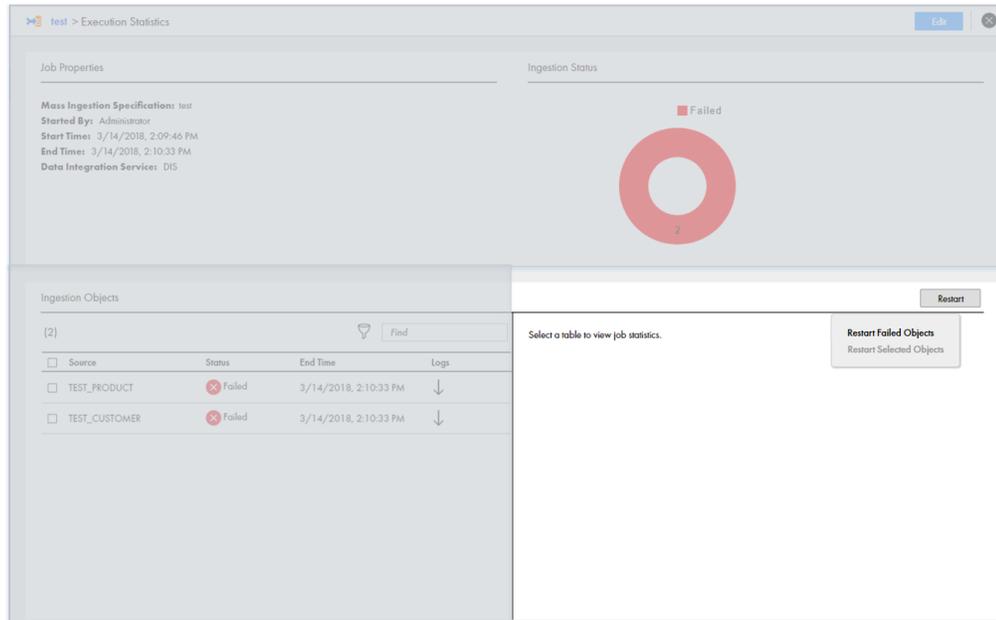
## Reiniciar objetos de ingesta con errores

Puede optar por reiniciar los objetos de ingesta para reiniciar el trabajo para ingerir solo las tablas de origen con errores o canceladas.

**Nota:** La opción para reiniciar objetos de ingesta con errores no reinicia los trabajos anulados.

1. En la vista Estadísticas de ingesta de la página Estadísticas de ejecución, seleccione **Reiniciar**.

En la siguiente imagen se muestra la opción **Reiniciar**:



2. Seleccione **Reiniciar objetos con errores**.

De este modo, se reinicia el trabajo de ingesta de todos los objetos con errores o cancelados.

## Supervisar la Herramienta del administrador

Puede supervisar una especificación de ingesta masiva en la Herramienta del administrador. Supervise la aplicación y los trabajos de asignación de ingesta que ingieren las tablas de origen configuradas en la especificación de ingesta masiva.

## Supervisar la aplicación y los trabajos de asignación de ingesta

Supervise la aplicación y los trabajos de asignación de ingesta en la ficha Supervisar. Puede consultar las estadísticas en la ficha Estadísticas de resumen o Estadísticas de ejecución.

### Estadísticas de resumen

En la ficha Estadísticas de resumen, puede ver resúmenes de diferentes aplicaciones y trabajos de asignación de ingesta. Los resúmenes proporcionan detalles sobre la distribución de objetos, el estado de

los objetos y el uso de recursos del servicio de integración de datos durante un intervalo de tiempo específico.

Puede elegir una de las siguientes opciones para ver un resumen:

- Distribución de gráficos. Muestra datos visuales sobre la distribución de los trabajos de asignación y aplicación.
- Distribución de tablas. Muestra el número total de trabajos completados, en ejecución, cancelados, anulados y con errores.
- Detalles. Muestra una lista de los trabajos de asignación o aplicaciones, solicitudes o conexiones que intervienen.
- Exportar datos. Exporte los datos de detalles de un trabajo de asignación o aplicación específico a un archivo .csv.

Para obtener más información, consulte la *Guía de Informática Administrator*.

## Estadísticas de ejecución

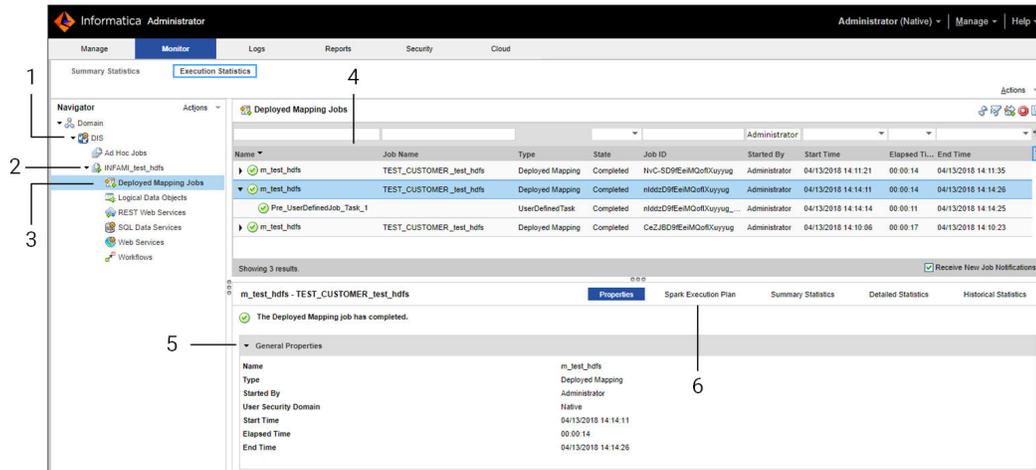
En la ficha Estadísticas de ejecución puede supervisar la aplicación y los trabajos de asignación de ingesta que se implementan en un servicio de integración de datos.

Puede seleccionar una aplicación en el navegador. Los paneles de contenido muestran las siguientes vistas:

- Vista Propiedades. Muestra las propiedades generales y las estadísticas en tiempo de ejecución. Utilice la vista Propiedades para ver más información sobre los trabajos de asignación y aplicación en la aplicación.
- Vista Informes. Muestra los informes de supervisión. Utilice la vista Informes para seguir de cerca los informes sobre los trabajos de asignación y aplicación.

Puede ir a los trabajos de asignación implementada de la aplicación para ver información sobre los trabajos de asignación de ingesta que ingieren las tablas en una especificación de ingesta masiva.

En la siguiente imagen se muestra la ficha Estadísticas de ejecución de un trabajo de asignación implementada:



1. Servicio de integración de datos. Servicio de integración de datos en el que se implementa la especificación de ingesta masiva.
2. Aplicación. Aplicación que contiene los trabajos de asignación en una especificación de ingesta masiva implementada.
3. Trabajos de asignación implementada. Asignaciones que llevan a cabo los trabajos de ingesta configurados en la especificación de ingesta masiva.
4. Panel de contenido. Propiedades que describen los trabajos de asignación.
5. Propiedades generales. Propiedades que resumen un trabajo de asignación.
6. Plan de ejecución. Plan de ejecución de Spark que resume las estadísticas de trabajo en Hadoop.

En la siguiente tabla se describe la información que se puede ver en el panel de contenido en relación un trabajo de asignación implementada:

Propiedad	Descripción
Nombre	Nombre de la especificación de ingesta masiva asociada al trabajo de asignación. El nombre se muestra con el siguiente formato: <code>m_&lt;mass ingestion specification name&gt;</code>
Nombre de trabajo	Nombre de la tabla de origen que el trabajo de asignación ingiere. El nombre se muestra con el siguiente formato: <code>&lt;source table name&gt;_&lt;mass ingestion specification name&gt;</code> Ordene los trabajos de asignación por nombre de trabajo para encontrar rápidamente la asignación que ingiere una tabla de origen.
Tipo	Tipo de trabajo que define una tarea en la asignación.
Estado	Estado del trabajo de asignación.
ID de trabajo	ID de trabajo que define el trabajo de asignación.
Iniciada por	Usuario que ha iniciado la instancia de ejecución de la especificación de ingesta masiva.
Hora de inicio	Hora a la que se ha iniciado el trabajo de asignación.

Propiedad	Descripción
Tiempo transcurrido	Tiempo que el trabajo de asignación requiere para completarse.
Hora de finalización	Hora a la que ha finalizado el trabajo de asignación.

Para obtener más información sobre las estadísticas de ejecución, vea la *Guía de Informática Administrator*.

Para obtener más información sobre los planes de ejecución de Spark, vea la guía *Informática Guía de Data Engineering Administrator*.

## Cancelar trabajos de ingesta

Puede cancelar un trabajo de asignación de ingesta en ejecución. Es posible que desee cancelar un trabajo que deja de responder o que tarda demasiado tiempo en completarse.

Al cancelar un trabajo de asignación de ingesta, se cancela la asignación que realiza el trabajo de ingesta en una tabla de origen de una instancia de ejecución de una especificación de ingesta masiva. Si supervisa la especificación de ingesta masiva desde la herramienta de ingesta masiva, el estado de la tabla de origen muestra **Cancelado**. Tras cancelar un trabajo, ninguno de los datos de la tabla de origen correspondiente se carga en el origen.

Tras cancelar un trabajo de asignación de ingesta, la asignación se puede volver a emitir para reiniciar el trabajo de asignación de ingesta de la tabla de origen. También puede reiniciar el trabajo de asignación de ingesta desde la herramienta de ingesta masiva.

1. Haga clic en la vista **Estadísticas de ejecución**.
2. En el navegador del dominio, expanda un servicio de integración de datos.
3. Expanda la aplicación que contiene la información de la especificación de ingesta masiva implementada y seleccione **Trabajos de asignación implementada**.  
Aparecerá en el panel de contenido una lista de trabajos.
4. En el panel de contenido, seleccione el trabajo de asignación de ingesta.
5. En el menú **Acciones**, seleccione **Cancelar objeto seleccionado**.
6. Para reiniciar el trabajo de asignación de ingesta, seleccione **Volver a emitir objeto seleccionado**.

## Anular trabajos de ingesta

Si una cantidad grande de trabajos en la especificación de ingesta masiva deja de responder, puede anular todos los trabajos desde la línea de comandos.

Después de anular los trabajos de una especificación de ingesta masiva, el estado de los trabajos que se estaban ejecutando cambia a **Anulado**. Ninguno de los datos en los trabajos anulados se carga en el destino.

Para obtener más información, vea [“abortRun” en la página 77](#) en el apéndice "Referencia de comandos infacmd mi".

# Solución de problemas de trabajos de ingesta masiva

Tenga en cuenta los siguientes consejos de solución de problemas cuando supervise trabajos de ingesta masiva:

## ¿Cómo puedo ver la traza de pila completa de una tabla concreta con un error de ingesta?

Para ver la traza de pila completa, realice las siguientes tareas:

1. En la Herramienta del administrador, vaya a la aplicación implementada en el servicio de integración de datos.
2. Detenga la aplicación implementada.
3. Dentro de la aplicación implementada, vaya al trabajo de ingesta correspondiente a la tabla.
4. Edite las propiedades de asignación para establecer el nivel de seguimiento en **Datos detallados**.
5. Inicie la aplicación implementada.
6. Ejecute de nuevo el trabajo de ingesta.  
Por ejemplo, puede utilizar el comando `infacmd ms runMapping`.
7. Vea los registros de implementación.

# APÉNDICE A

## Referencia de comando infacmd mi

Este apéndice incluye los siguientes temas:

- [abortRun, 77](#)
- [createService, 78](#)
- [deploySpec, 82](#)
- [exportSpec, 83](#)
- [extendedRunStats, 84](#)
- [getSpecRunStats, 86](#)
- [listSpecRuns, 87](#)
- [listSpecs, 88](#)
- [restartMapping, 89](#)
- [runSpec, 90](#)

### abortRun

Anula los trabajos de asignación de ingesta en una instancia de ejecución de una especificación de ingesta masiva. Cuando se anulan los trabajos de asignación de ingesta, el comando anula las asignaciones que realizan los trabajos de ingesta para todas las tablas de origen que se ejecutan o que están en cola. El comando no anula asignaciones de los trabajos de ingesta que se han completado.

Para anular los trabajos de asignación de ingesta, deberá especificar un RunID. Para buscar el RunID de una instancia de ejecución, enumere las instancias de ejecución de especificación mediante el comando infacmd mi listSpecRuns.

El comando infacmd mi abortRun utiliza la siguiente sintaxis:

```
abortRun
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
```

```
<-ServiceName|-sn> service_name
```

```
<-runID|-rid> run_id
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi abortRun:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-runID -rid	run_id	Obligatorio. Número de identificador de ejecución, o identificador de ejecución, de la instancia de ejecución de especificación de ingesta masiva.  Para buscar el RunID de una instancia de ejecución, enumere las instancias de ejecución de especificación mediante el comando infacmd mi listSpecRuns.

## createService

Crea un servicio de ingesta masiva. Al crear el servicio de ingesta masiva, se debe especificar un servicio de repositorio de modelos. El servicio de ingesta masiva está deshabilitado de forma predeterminada. Para habilitar el servicio de ingesta masiva, utilice el comando infacmd isp enableService.

El comando infacmd mi createService emplea la siguiente sintaxis:

```
createService  
  
<-DomainName|-dn> domain_name  
  
<-UserName|-un> user_name  
  
<-Password|-pd> password  
  
[<-SecurityDomain|-sdn> security_domain]
```

```

[<-Gateway|-hp> gateway_host1:port gateway_host2:port...]
[<-ResilienceTimeout|-re> timeout_period_in_seconds]
<-ServiceName|-sn> service_name
<-HttpPort|-http> http_port
[<-HttpsPort|-https> https_port]
[<-KeystoreFile|-kf> keystore_file_location]
[<-KeystorePassword|-kp> keystore_password]
<-LicenseName|-ln> license_name
[<-FolderPath|-fp> full_folder_path]
<-NodeName|-nn> node_name
<-RepositoryService|-rs> repository_service_name
[<-RepositoryUser|-ru> repository_user]
[<-RepositoryPassword|-rp> repository_password]
[<-RepositoryUserSecurityDomain|-rsdn> repository_user_security_domain]

```

La tabla siguiente describe las opciones y los argumentos de infacmd mi createService:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informática. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.

Opción	Argumento	Descripción
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas.  El valor predeterminado es Nativo.
-Gateway -hp	gateway_host1:port gateway_host2:port	Obligatorio si la información de conectividad de la puerta de enlace en el archivo domains.infa no está actualizada. Los nombres de host y los números de puertos de los nodos de puertas de enlace del dominio.
-ResilienceTimeout -re	timeout_period_in_seconds	Opcional. Tiempo, en segundos, durante el que infacmd intenta establecer o restablecer una conexión con el dominio. El tiempo de espera de fiabilidad se puede configurar con la opción -re o con la variable de entorno INFA_CLIENT_RESILIENCE_TIMEOUT. Si establece el tiempo de espera de fiabilidad con ambos métodos, la opción -re tendrá preferencia.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva.  No se aplica la distinción entre mayúsculas y minúsculas al nombre, el cual debe ser único en el dominio. Los caracteres deben ser compatibles con la página de códigos del repositorio asociado. El nombre no puede contener más de 230 caracteres y tampoco puede contener espacios iniciales ni finales, ni retornos de carro, tabulaciones o los siguientes caracteres:  / * ? < > "
-HttpPort -http	http_port	Obligatoria si no se especifica un puerto HTTPS. Número de puerto HTTP único que se utiliza por cada proceso del servicio de ingesta masiva. Después de crear el servicio, puede definir diferentes números de puerto para cada proceso del servicio de ingesta masiva.  El valor predeterminado es 9050. <b>Nota:</b> No se puede especificar un puerto HTTP y un puerto HTTPS al mismo tiempo.

Opción	Argumento	Descripción
-HttpsPort -https	https_port	Obligatoria si no se especifica un puerto HTTP. Número de puerto HTTPS único que se utiliza por cada proceso del servicio de ingesta masiva. Después de crear el servicio, puede definir diferentes números de puerto para cada proceso del servicio de ingesta masiva. <b>Nota:</b> No se puede especificar un puerto HTTP y un puerto HTTPS al mismo tiempo.
-KeystoreFile -kf	keystore_file_location	Obligatoria si se especifica un puerto HTTPS. Ruta de acceso y nombre de archivo del almacén de claves que contiene las claves y los certificados necesarios cuando se utiliza el protocolo HTTPS en el servicio de ingesta masiva. Para crear un archivo de almacén de claves, se usa una utilidad keytool. La utilidad keytool genera y almacena pares de claves privadas o públicas y los certificados asociados en un archivo de almacén de claves. Puede utilizar el certificado autofirmado o un certificado firmado por una autoridad de certificación.
-KeystorePassword -kp	keystore_password	Obligatoria si se especifica un puerto HTTPS. Contraseña del archivo de almacén de claves.
-LicenseName -ln	license_name	Obligatorio. Nombre de la licencia que desea asignar al servicio de ingesta masiva. Para aplicar cambios, reinicie el servicio de ingesta masiva.
-FolderPath -fp	full_folder_path	Opcional. La ruta de acceso completa, excluyendo el nombre de dominio, a la carpeta en la que desea crear el servicio de ingesta masiva. Debe tener el siguiente formato:  <i>/parent_folder/child_folder</i>  El valor predeterminado es el dominio.  /
-NodeName -nn	node_name	Obligatorio. Nodo en el que se ejecuta el servicio de ingesta masiva.
-RepositoryService -rs	repository_service_name	Obligatorio. Servicio de repositorio de modelos que almacena los metadatos para las especificaciones de ingesta masiva.
-RepositoryUser -ru	repository_user	Opcional. Nombre de usuario para acceder al servicio de repositorio de modelos.

Opción	Argumento	Descripción
-RepositoryPassword -rp	repository_password	Es obligatorio si especifica el nombre de usuario. Contraseña de usuario para acceder al servicio de repositorio de modelos.
-RepositoryUserSecurityDomain -rsdn	repository_user_security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del repositorio de modelos.

## deploySpec

Implementa una especificación de ingesta masiva. Al implementar la especificación, se deben especificar el servicio de integración de datos y la conexión de Hadoop. Debe implementar la especificación de ingesta masiva antes de ejecutarla. Después de implementar la especificación, ejecútela mediante el comando `infacmd mi runSpec`.

El comando `infacmd mi deploySpec` utiliza la siguiente sintaxis:

```

deploySpec
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-DISServiceName|-dsn> dis_service_name
<-MISpecName|-spec> mi_spec_name
<-HadoopConnection|-hc> hadoop_connection

```

En la tabla siguiente se describen las opciones y argumentos de `infacmd mi deploySpec`:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción <code>-dn</code> o con la variable de entorno <code>INFA_DEFAULT_DOMAIN</code> . Si se establece un nombre de dominio con ambos métodos, la opción <code>-dn</code> tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción <code>-un</code> o con la variable de entorno <code>INFA_DEFAULT_DOMAIN_USER</code> . Si se establece un nombre de usuario con ambos métodos, la opción <code>-un</code> tendrá preferencia.

Opción	Argumento	Descripción
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-DISServiceName -dis	data_integration_service	Obligatorio. Nombre del servicio de integración de datos donde desea implementar la especificación de ingesta masiva.
-MISpecName -spec	mi_spec_name	Obligatorio. Nombre de la especificación de ingesta masiva que desea implementar en el servicio de integración de datos.
-HadoopConnection -hc	hadoop_connection	Obligatorio. Conexión de Hadoop que el servicio de integración de datos usa para insertar la especificación de ingesta masiva en el entorno de Hadoop.

## exportSpec

Exporta la especificación de ingesta masiva a un archivo de archivado de la aplicación. Al exportar la especificación, se debe especificar el directorio en el que se desea guardar el archivo. Puede implementar el archivo de archivado de la aplicación en un servicio de integración de datos mediante el comando `infacmd dis DeployApplication`.

El comando `infacmd mi exportSpec` utiliza la siguiente sintaxis:

```
exportSpec
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-MISpecName|-spec> mi_spec_name
<-Directory|-dir> dir_path
<-HadoopConnection|-hc> hadoop_connection
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi exportSpec:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-MISpecName -spec	mi_spec_name	Obligatorio. Nombre de la especificación de ingesta masiva que desea exportar.
-Directory -dir	dir_path	Obligatorio. El directorio donde desea escribir el archivo de archivado de la aplicación.
-HadoopConnection -hc	hadoop_connection	Obligatorio. La conexión de Hadoop que el servicio de integración de datos usará para ejecutar el trabajo de ingesta masiva al importar el archivo de archivado de la aplicación y ejecutar la aplicación.  Debe especificar la conexión de Hadoop porque una conexión de Hadoop no persiste para la especificación de ingesta masiva mientras que la especificación se almacena en el repositorio de modelos.

## extendedRunStats

Obtiene las estadísticas de procesamiento extendidas para una tabla de origen específica en una especificación de ingesta masiva implementada. Para obtener las estadísticas extendidas, debe especificar el RunID de la especificación de ingesta masiva, el nombre de la tabla de origen y el tipo de asignación.

Las estadísticas extendidas informan de las estadísticas de ingesta de las filas de tabla ingeridas desde el origen y las estadísticas de ingesta de las filas de tabla ingeridas en el destino. Las estadísticas enumeran el número de filas que se han ingerido correctamente y el número de filas que contienen errores.

Si la instancia de ejecución utiliza una carga incremental, las estadísticas extendidas también indicarán la clave incremental y el valor de inicio. La clave incremental es el nombre de la columna que el motor de Spark ha usado para obtener datos incrementales en la tabla de origen. El valor de inicio es el valor que el motor de Spark ha usado para empezar a ingerir datos incrementales.

El comando `infacmd mi extendedRunStats` utiliza la siguiente sintaxis:

```
extendedRunStats
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-RunID|-rid> run_id
<-SourceName|-srcName> source_name
<-MappingTp|-mtp> mapping_type
```

En la tabla siguiente se describen las opciones y argumentos de `infacmd mi extendedRunStats`:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva asociada al trabajo de asignación de ingesta.
-RunID -rid	run_id	Obligatorio. Número de identificador de ejecución (o identificador de ejecución) de la instancia de ejecución de especificación de ingesta masiva. Para hallar el RunID de una instancia de ejecución, obtenga una lista de las instancias de ejecución de especificación mediante el comando <code>infacmd mi listSpecRuns</code> .

Opción	Argumento	Descripción
-SourceName -srcName	source_name	Obligatorio. Nombre de la tabla de origen en la instancia de ejecución de la especificación de ingesta masiva. Para hallar el nombre de la tabla de origen, obtenga las estadísticas de ejecución de la ingesta mediante el comando infacmd mi getSpecRunStats.
-MappingTp -Mtp	Mapping_type	Obligatorio. El tipo de asignación corresponde al motor de tiempo de ejecución que ejecuta el trabajo de asignación de ingesta relativo a la tabla de origen. Para hallar el tipo de asignación, obtenga las estadísticas de ejecución de la ingesta mediante el comando infacmd mi getSpecRunStats.

## getSpecRunStats

Obtiene las estadísticas de ejecución detalladas de una especificación de ingesta masiva implementada. Para obtener las estadísticas, se debe especificar un RunID. Para buscar el RunID de una instancia de ejecución, enumere las instancias de ejecución de especificación mediante el comando infacmd mi listSpecRuns.

Las estadísticas de ejecución detallada informan sobre el JobID para cada trabajo de asignación de ingesta en la especificación de ingesta masiva implementada, el nombre de la tabla de origen que ingiere cada trabajo de asignación, la hora de inicio de ejecución, la hora de finalización, el motor de tiempo de ejecución que ejecuta el trabajo de asignación y el estado del trabajo. El JobID es el identificador del trabajo de asignación de ingesta que ingiere la tabla de origen. El estado puede aparecer como completado, error, cancelado, en ejecución, interrumpido, en cola o desconocido.

El comando infacmd mi getSpecRunStats utiliza la siguiente sintaxis:

```
getSpecRunStats
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-runID|-rid> run_id
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi getSpecRunStats:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.

Opción	Argumento	Descripción
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-runID -rid	run_id	Obligatorio. Número de identificador de ejecución, o identificador de ejecución, de la instancia de ejecución de especificación de ingesta masiva.  Para buscar el RunID de una instancia de ejecución, enumere las instancias de ejecución de especificación mediante el comando infacmd mi listSpecRuns.

## listSpecRuns

Enumera las instancias de ejecución de una especificación de ingesta masiva implementada. Cada instancia de ejecución está definida por un RunID. Cuando se enumeran las instancias de ejecución, se debe especificar el servicio de ingesta masiva.

Las estadísticas de ejecución detalladas informan del RunID de cada instancia de ejecución de la especificación; del tipo de carga; de la hora de inicio de la instancia de ejecución; del servicio de integración de datos donde se implementa la especificación de ingesta masiva; del usuario que inició la ejecución, y del estado del trabajo de cada instancia de ejecución. El estado puede aparecer como completado, error, cancelado, en ejecución, en cola o desconocido.

El comando infacmd mi listSpecRuns utiliza la siguiente sintaxis:

```
listSpecRuns
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-MISpecName|-spec> mi_spec_name
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi listSpecRuns:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informática. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-MISpecName -spec	mi_spec_name	Obligatorio. Nombre de la especificación de ingesta masiva.

## listSpecs

Enumera las especificaciones de ingesta masiva. Cuando se enumeran las especificaciones, se debe especificar el servicio de ingesta masiva.

El comando infacmd mi listSpecs utiliza la siguiente sintaxis:

```
listSpecs
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi listSpecs:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informática. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra las especificaciones de ingesta masiva.

## restartMapping

Reinicia los trabajos de asignación de ingestas en una especificación de ingesta masiva. Especifique la lista de tablas de origen que desee reiniciar. Debe especificar el servicio de ingesta masiva y el RunID para la instancia de ejecución de la especificación de ingesta masiva. También puede especificar si desea reiniciar solo las tablas de origen que han fallado.

El comando infacmd mi restartMapping utiliza la siguiente sintaxis:

```
restartMapping
<-DomainName|-dn> domain_name
<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-RunID|-rid> run_id
<-SourceList|-srcList> comma_separated_source_list
[<-OnlyFailed|-failed> true|false]
```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi restartMapping:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas.  El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la ingesta de las tablas de origen.
-runID -rid	run_id	Obligatorio. Número identificativo de ejecución (ID de ejecución) de la instancia de ejecución de la especificación de ingesta masiva.
-SourceList -srcList	comma_separated_source_list	Obligatorio. La lista de tablas de origen que se van a reiniciar. Separe cada tabla de origen con una coma.
-OnlyFailed -failed	true false	Opcional. Escriba true para reiniciar solo las tablas de origen que no se han ingerido. Escriba false para reiniciar todas las tablas de origen.

## runSpec

Ejecuta una asignación de ingesta masiva que se implementa en un servicio de integración de datos. Antes de ejecutar una especificación, esta se debe implementar mediante infacmd mi deploySpec.

El comando infacmd mi runSpec utiliza la siguiente sintaxis:

```
runSpec
<-DomainName|-dn> domain_name
```

```

<-UserName|-un> user_name
<-Password|-pd> password
[<-SecurityDomain|-sdn> security_domain]
<-ServiceName|-sn> service_name
<-MISpecName|-spec> mi_spec_name
[<-LoadType|-lt> load_type]
<-DISServiceName|-dsn> dis_service_name
[<-OperatingSystemProfile|-osp> operating_system_profile_name]

```

En la tabla siguiente se describen las opciones y argumentos de infacmd mi runSpec:

Opción	Argumento	Descripción
-DomainName -dn	domain_name	Obligatorio. Nombre del dominio de Informatica. El nombre de dominio se puede establecer con la opción -dn o con la variable de entorno INFA_DEFAULT_DOMAIN. Si se establece un nombre de dominio con ambos métodos, la opción -dn tendrá preferencia.
-UserName -un	user_name	Obligatorio. Nombre de usuario para conectar con el dominio. El nombre de usuario se puede establecer con la opción -un o con la variable de entorno INFA_DEFAULT_DOMAIN_USER. Si se establece un nombre de usuario con ambos métodos, la opción -un tendrá preferencia.
-Password -pd	password	Obligatorio. Contraseña del nombre de usuario. La contraseña distingue entre mayúsculas y minúsculas. Las contraseñas se pueden establecer con la opción -pd o con la variable de entorno INFA_DEFAULT_DOMAIN_PASSWORD. Si se establece una contraseña con ambos métodos, la contraseña establecida con la opción -pd tendrá preferencia.
-SecurityDomain -sdn	security_domain	Opcional. Nombre del dominio de seguridad al que pertenece el usuario del dominio. Los dominios de seguridad se pueden establecer con la opción -sdn o con la variable de entorno INFA_DEFAULT_SECURITY_DOMAIN. Si establece un nombre de dominio de seguridad con ambos métodos, la opción -sdn tendrá preferencia. El dominio de seguridad distingue mayúsculas de minúsculas. El valor predeterminado es Nativo.
-ServiceName -sn	service_name	Obligatorio. Nombre del servicio de ingesta masiva que administra la especificación de ingesta masiva.
-MISpecName -spec	mi_spec_name	Obligatorio. Nombre de la especificación de ingesta masiva que se implementa en el servicio de integración de datos.

Opción	Argumento	Descripción
-LoadType -lt	load_type	Opcional. Tipo de carga de la ingesta de datos en la especificación de ingesta masiva. Use <code>full</code> o <code>incremental</code> . El valor predeterminado es <code>full</code> .  Si la carga incremental no está habilitada en la especificación de ingesta masiva, no puede usar una carga incremental para ingerir los datos.
-DISServiceName -dis	data_integration_service	Obligatorio. Nombre del servicio de integración de datos en el que se implementa la especificación de ingesta masiva.
-OperatingSystemProfile -osp	operating_system_profile_name	Opcional. Nombre del perfil del sistema operativo configurado para el servicio de integración de datos.

# INDICE

## A

abortRun (infacmd mi) [77](#)

## C

carga completa  
  opciones de carga incremental [56](#)  
carga incremental  
  clave [29](#)  
  clave de ID [30](#)  
  clave de marca de tiempo [29](#)  
  clave incremental [29](#)  
  ID [30](#)  
  marca de tiempo [29](#)  
  modo [36](#)  
  modo de anexión [36](#)  
  modo de sobrescritura [36](#)  
  opciones [36](#)  
CreateService (infacmd mi) [78](#)

## D

deploySpec (infacmd mi) [82](#)  
desfase de esquema [37](#)

## E

especificación  
  implementación en un archivo de archivado [83](#)  
especificación de ingesta masiva  
  anulando [77](#)  
  código de compresión [36](#)  
  consulta ddl [33](#)  
  create [47](#), [48](#)  
  ejecutar [58](#)  
  filtrar [69](#)  
  implementar [51](#)  
  supervisar  
    estadísticas de ejecución [67](#)  
especificaciones implementadas en un servicio de integración de  
  datos  
  Ejecución [90](#)  
estado de especificación  
  acceder con infacmd mi [86](#)  
exportSpec  
  infacmd mi [83](#)  
expresión regular  
  en mi spec [42](#)  
Expresión regular  
  en mi spec [42](#)

## F

filtrar objetos de ingesta [69](#)

## G

getSpecRunStats  
  infacmd mi [86](#)

## H

herramienta de ingesta masiva  
  interfaz [16](#)  
  interfaz de usuario [16](#)

## I

implementar mi spec  
  archivo de almacenamiento de la aplicación [53](#)  
infacmd mi  
  anular una especificación de ingesta masiva [77](#)  
  creación del servicio de ingesta masiva [78](#)  
  ejecutando mi spec [90](#)  
  extendedRunStats [84](#)  
  implementar especificación [83](#)  
  implementar la especificación de ingesta masiva [82](#)  
  listado de especificaciones de ingesta masiva [88](#)  
  listSpecRuns [87](#)  
  obtener las estadísticas de especificaciones [86](#)  
  reiniciar trabajos [89](#)  
ingesta  
  supervisión [61](#)  
ingesta masiva  
  abortRun [75](#)  
  antes de empezar [19](#)  
  anular trabajo [75](#)  
  arquitectura [12](#)  
  asignar función de administrador [24](#)  
  asignar funciones [24](#)  
  cadena entera [42](#)  
  cancelar trabajo [75](#)  
  carga incremental [26](#)  
  caso de uso [9](#)  
  cláusula de filtro [41](#)  
  clientes y herramientas [12](#)  
  configurar el destino [48](#)  
  configurar el destino de Hive [48](#)  
  configurar el origen [47](#)  
  configurar supervisión [22](#)  
  configurar transformaciones masivas [48](#), [49](#)  
  crear [25](#)  
  criterios de reemplazo [42](#)  
  datos incrementales [26](#)

- ingesta masiva (*continuado*)
  - definición [26, 47](#)
  - destino [31](#)
  - destino de Hive [48](#)
  - diagrama [14](#)
    - ejecutar
      - estado [57](#)
  - especificación [27, 31, 38, 44, 72, 73](#)
  - estadísticas de ejecución [64, 72, 73, 84](#)
  - estadísticas de ingesta [66](#)
  - estadísticas de resumen [72, 73](#)
  - estado de ingesta [66](#)
  - expresión regular [42](#)
  - Expresión regular [42](#)
  - filtrar por [41](#)
  - herramienta [15](#)
  - historial de ejecución [62](#)
    - implementar
      - vista [61](#)
  - inicio de sesión [16](#)
  - instancias de ejecución [57](#)
  - migrar [53](#)
  - objetos de ingesta [65](#)
  - origen [27](#)
  - parámetros de tabla [44](#)
  - parámetros masivos [38](#)
  - patrón [42](#)
  - preparar [19](#)
  - propiedades del trabajo [64](#)
  - reemplazar columnas [42](#)
  - reiniciar [70](#)
  - reiniciar con errores [72](#)
  - reiniciar ingesta [70](#)
  - reiniciar seleccionados [71](#)
  - reiniciar trabajos [71, 72](#)
  - reiniciar trabajos de ingesta [70](#)
  - repositorios [13](#)
  - resumen [8, 61](#)
  - servicios de aplicación [13](#)
  - source [47](#)
  - supervisar [60, 72, 73](#)
  - supervisión [64](#)
  - target [31, 34, 48](#)
  - transformaciones [48, 49](#)
  - transformaciones masivas [48, 49](#)
  - vistas [17](#)

## L

listSpecs (infacmd mi) [88](#)

## M

mi

- caso de uso [9](#)
- clientes y herramientas [12](#)

mi (*continuado*)

- migrar [53](#)
- repositorios [13](#)
- servicios de aplicación [13](#)

mi run

- estado [57](#)

mi spec

- cancelación [75](#)
- cómo crear [25](#)
- destino de HDFS [34](#)
- destino de Hive [31, 45](#)
- en ejecución [55](#)
- estadísticas de ejecución [64](#)
- implementar [50, 51](#)
  - parámetros de tabla
    - destino de Hive [45](#)
- resumen [61](#)
- supervisión [60](#)

## P

preferencias

- supervisión [23](#)

propagar cambios de esquema

- en el destino [37](#)
- en el origen [37](#)

## R

restartMapping (infacmd mi) [89](#)  
resumen de ejecuciones

- especificación de ingesta masiva implementada [87](#)

runSpec

- infacmd mi [90](#)

## S

Servicio de ingesta masiva

- asignar a nodo [19](#)
- creación [78](#)
- crear [19](#)

supervisión

- preferencias, configurar [23](#)

## T

tipo de carga [56](#)

## V

ver mi spec

- ejecuciones [57](#)