Informatica® Data Engineering Integration
10.5

Mass Ingestion Guide

Informatica Data Engineering Integration Mass Ingestion Guide
10.5
March 2021

Publication Date: 2021-04-01

Table of Contents

Preface

Use the *Informatica® Data Engineering Integration Mass Ingestion Guide* to understand and navigate the user interface in the Mass Ingestion tool. Learn how you can create, deploy, run, and monitor mass ingestion jobs.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit https://network.informatica.com.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit https://search.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit https://docs.informatica.com.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at http://velocity.informatica.com. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
https://www.informatica.com/services-and-training/customer-success-services/contact-us.html.

To find online support resources on the Informatica Network, visit https://network.informatica.com and select the eSupport option.

CHAPTER 1

Introduction to Mass Ingestion

This chapter includes the following topics:

Overview

Use Informatica Mass Ingestion (the Mass Ingestion tool) to ingest large amounts of data from a relational database to a Hive or HDFS target.

The Mass Ingestion tool simplifies the process of ingesting data by providing a wizard that you can use to create a mass ingestion specification. A mass ingestion specification is a configuration that you can design to specify the data that you want to ingest and how you want to ingest it.

The wizard walks you through the steps that you can use to configure each part of the specification, including the relational source and the Hive or HDFS target, and any parameters that you want to configure for the source, such as a parameter to filter certain columns or to mask the data to protect private information.

When you run the mass ingestion specification, the Mass Ingestion tool uses Data Engineering Integration to run the ingestion job on a Hadoop cluster. The specification replaces the need to manually create and run mappings, and it can ingest all of your data in one run. As the schemas in the relational database evolve, the specification can accommodate and ingest only the incremental data.

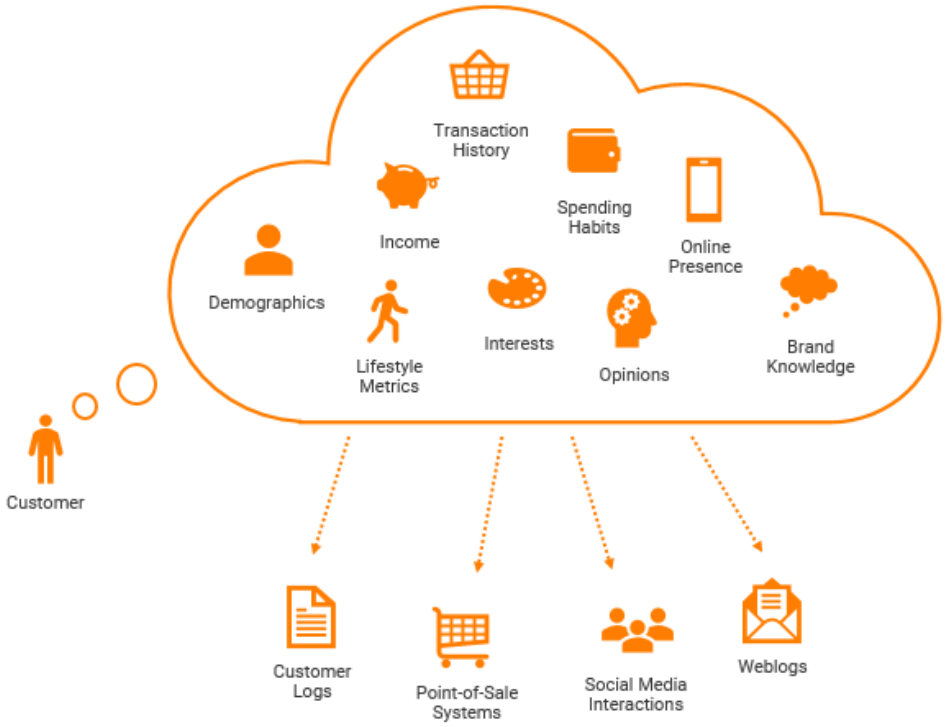
Use Case

You work in an IT group for a commercial bank. Your team is taking on a new project to personalize the rewards program that you offer to customers who open checking and savings accounts at your bank.

You plan to collect and analyze data on your customers to understand the types of rewards that customers are interested in. For example, one customer might be interested in saving money on groceries while another customer might be interested in travel deals.

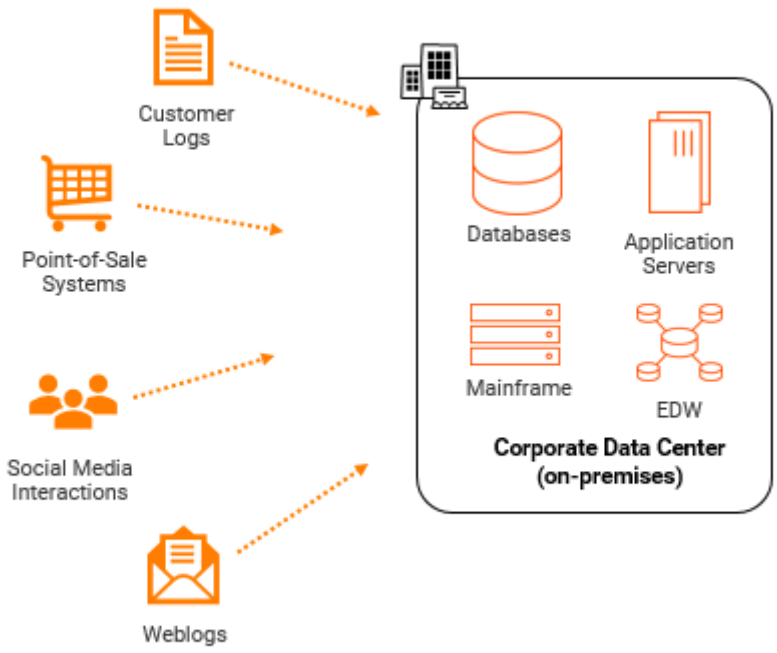
You collect data on customer demographics, lifestyle metrics, income, transaction history, spending habits, online presence, interests, opinions, and brand knowledge. The data is collected through different media, including customer logs on file with the bank, point-of-sale systems at companies that the bank partners with, social media interactions, and customer weblogs.

The following image shows the types of data that you collect and the media that you use to collect the different types of data:



When the data is collected, the data is stored in the bank's corporate data center which includes various relational databases.

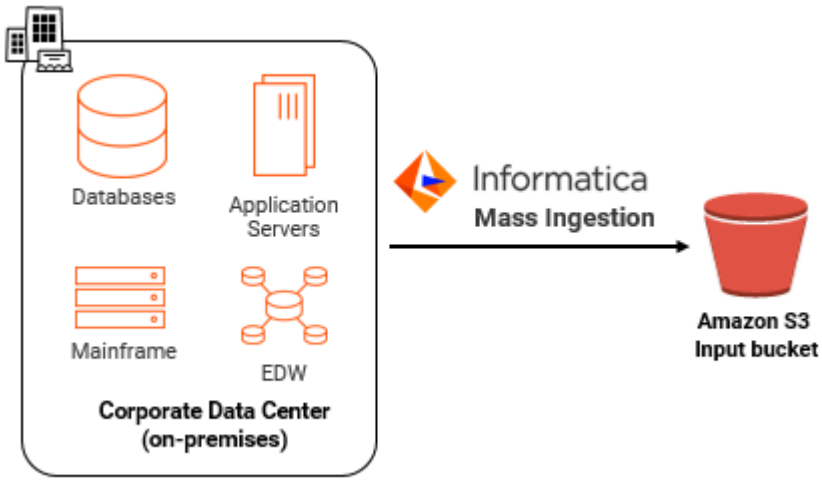
The following image shows how the data might be stored::



Before your data analysts can begin working with the data, you need to ingest the data from the relational databases into Amazon S3 buckets. But you cannot spend the time and resources required to ingest the large amounts of data. You will have to develop numerous mappings and parameter sets to ingest the data to make sure that the data is ingested properly. You also have to make sure that you do not ingest sensitive customer information such as customer credit card numbers. You then have to maintain the mappings when relational schemas change.

Instead of manually creating and running mappings, you can use the Mass Ingestion tool to create one mass ingestion specification that ingests all of your data at once. You have to specify only the source, the target, and any parameters to apply across source tables. When you deploy and run the specification, the Spark engine ingests all of the data to Amazon S3.

The following image shows how mass ingestion can branch the link between the data that the bank stores in its relational databases and the Amazon S3 buckets:

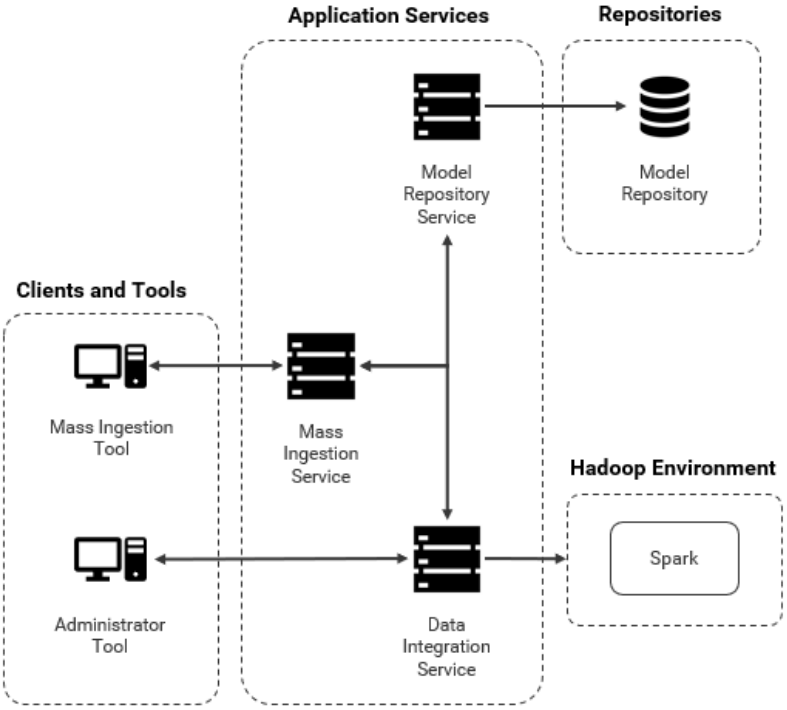


Mass ingestion saved you a lot of time and resources and your data analysts have more time to analyze the data and develop a new system for the bank's rewards program.

Architecture

The mass ingestion architecture includes the components that create, deploy, run, and monitor a mass ingestion specification.

The following image shows the mass ingestion components:



Clients and Tools

The mass ingestion components include the following clients and tools:

Informatica Administrator

Administrator tool. The client where you can create and configure a Mass Ingestion Service. You can also use the client to monitor the specific ingestion mapping jobs that the Spark engine runs to ingest the data in a mass ingestion specification.

Informatica Mass Ingestion

Mass Ingestion tool. The web application where you design and implement mass ingestion specifications. Use the web application to configure, deploy, run, and monitor a mass ingestion specification.

Application Services

The mass ingestion components include the following application services:

Mass Ingestion Service

Manages and validates mass ingestion specifications that you create in the Mass Ingestion tool. The Mass Ingestion Service deploys specifications and schedules the specifications to run. When a specification runs, the Mass Ingestion Service generates ingestion job statistics.

Model Repository Service

Manages the Model repository. The Model Repository Service stores the mass ingestion specifications that you create in the Model repository.

Note: Mass ingestion does not support a Model Repository Service that is integrated with a version control system.

Data Integration Service

Manages deployed mass ingestion specifications. The Data Integration Service connects to the Hadoop environment to allow the Spark engine to ingest the data that is configured in a mass ingestion specification.

Repositories

The mass ingestion components include the following repositories:

Model repository

The repository that stores the metadata for the mass ingestion specifications that you create in the Mass Ingestion tool.

The Model repository stores mass ingestion specifications in the project `INFA_MI_PROJECT`. If a specification is stored in a different project, the Mass Ingestion Service ignores the specification. To allow the Mass Ingestion Service to read all specifications, do not move the specification to a different project.

Note: You cannot view the project in the Developer tool, and you cannot view the specifications stored in the project from the Developer tool. To view the list of specifications that are stored in the project, use the command line.

Connections

When you run mass ingestion jobs, the mass ingestion components use the following connections:

JDBC

A JDBC connection accesses the tables in a relational database in a mass ingestion job.

The source connection that you use for a mass ingestion job must be a JDBC connection. For example, to access an Oracle schema, you must configure a JDBC connection that uses an Oracle driver to connect to an Oracle database. You cannot use an Oracle connection.

Sqoop

When you configure a JDBC connection with Sqoop arguments, tasks are divided between JDBC and Sqoop. JDBC is used to import metadata from the relational database, while Sqoop reads the data.

If you use an incremental load to ingest data using a Sqoop connection, the Mass Ingestion Service leverages Sqoop's incremental import mode. When the Mass Ingestion Service configures the filter for incremental data, the filter is pushed down to the Sqoop source.

If you use a Sqoop connection, consider the following limitations:

- A source table cannot be ingested using a Sqoop connection if the table contains special characters in the table metadata.
- Blob data types cannot be ingested using a Sqoop connection.

Hadoop

A Hadoop connection allows the Data Integration Service to push mass ingestion jobs to the Hadoop environment where the jobs run on the Spark engine.

Hive

A Hive connection accesses Hive data and allows a mass ingestion job to write Hive data to a Hive target.

HDFS

An HDFS connection accesses data on the Hadoop cluster to allow a mass ingestion job to write flat-file data to the cluster.

For information about connection properties, see the "Connections" appendix in the *Informatica Data Engineering Integration User Guide*.

Process

The mass ingestion process incorporates the components within the mass ingestion architecture that create, deploy, run, and monitor a mass ingestion specification.

The mass ingestion process includes the following tasks:

Create

You create a mass ingestion specification in the Mass Ingestion tool. The Mass Ingestion Service validates and connects to the Model Repository Service to store the specification in a Model repository.

After you create the specification, you can migrate the specification between Model repositories.

Deploy

You deploy the mass ingestion specification to a Data Integration Service and specify a Hadoop connection. The Mass Ingestion Service processes and deploys the specification to the Data Integration Service.

You can also deploy the mass ingestion specification to an application archive file to save the information about the specification as an application. If you deploy the specification to an application archive file, you can import the application to the Model repository and deploy the application to a Data Integration Service.

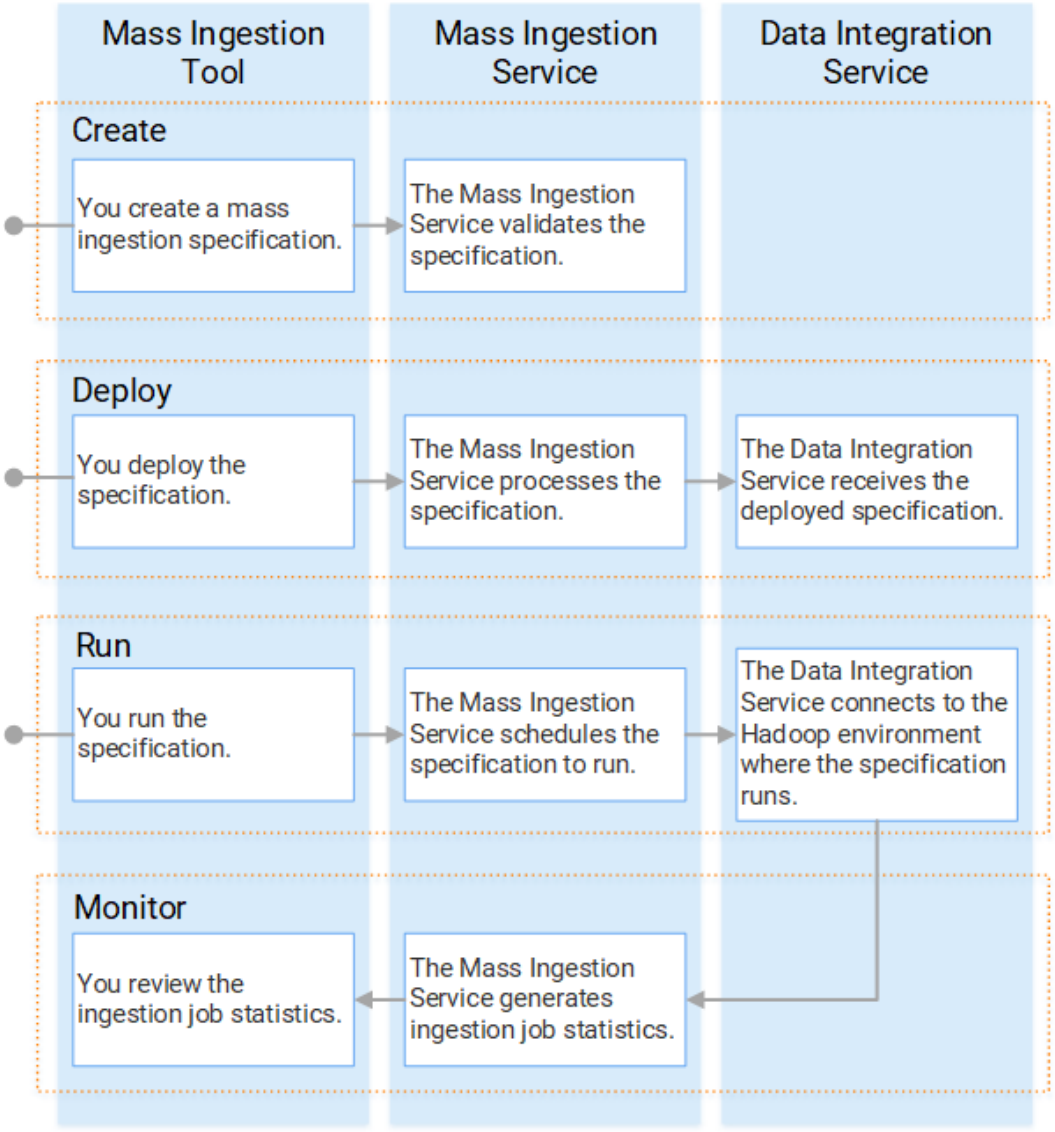
Run

You run the mass ingestion specification to ingest data to a Hive or HDFS target. The Mass Ingestion Service schedules the specification to run, and the Data Integration Service pushes the specification to the Spark engine in the Hadoop environment.

Monitor

The Mass Ingestion Service generates ingestion job statistics. You can monitor the statistics in the Mass Ingestion tool or in the Administrator tool.

The following diagram illustrates the detailed mass ingestion process when you create, deploy, run, and monitor a mass ingestion specification:



Mass Ingestion Tool

The Mass Ingestion tool is a web application that provides the interface to design mass ingestion specifications and run mass ingestion jobs.

When you design a mass ingestion specification in the Mass Ingestion tool, you specify how you want to ingest your source data. You can use the Mass Ingestion tool to deploy and run the mass ingestion

specification to perform the ingestion job, and you can monitor the ingestion job to review ingestion statistics.

Logging In to the Mass Ingestion Tool

After the system administrator configures the Mass Ingestion Service, you can log in to the Mass Ingestion tool. To get the URL for the Mass Ingestion tool, contact your Informatica system administrator.

1. Open a web browser.
2. In the address field, enter the following URL provided by your Informatica system administrator:

   `http://<host>:<port>/mi/login`
3. If you configured LDAP security in the Administrator tool, the login page for the Mass Ingestion tool prompts you to select a security domain. Select either the native security domain or a configured LDAP security domain.

   If you configured SAML authentication on the domain, you are redirected to the third-party security provider to sign in.
4. If the domain is not configured for SAML authentication, enter your user name and password on the login page. Click **Log In**.

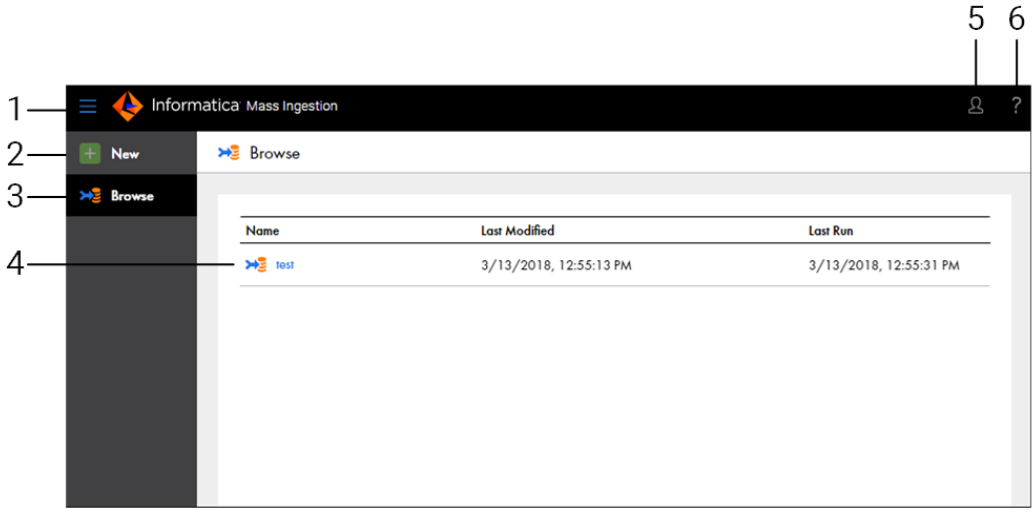
The landing page for the Mass Ingestion tool appears.

Mass Ingestion Tool User Interface

Use the Mass Ingestion tool to design and implement mass ingestion jobs.

You can create multiple mass ingestion specifications in the Mass Ingestion tool, but you can work only on one mass ingestion specification at a time. To work in the Mass Ingestion tool, you access the Mass Ingestion tool workbench.

The following figure shows the Mass Ingestion tool workbench:



1. Menu. View navigation options to create a mass ingestion specification or browse mass ingestion specifications.
2. New. Create a mass ingestion specification.
3. Browse. Browse created mass ingestion specifications.
4. Mass ingestion specification. Review the name of the mass ingestion specification, the time the specification was last modified, and the time the specification was last run.
5. User. View the user who is currently signed in, or sign out of the Mass Ingestion tool.
6. Help. Access help for the Mass Ingestion tool.

Navigation

You can navigate to additional pages from the Mass Ingestion tool workbench.

You can navigate to the following pages:

Browse

Displays the mass ingestion specifications that you create. To navigate to the Browse page, select Browse from the menu.

Create

Displays the wizard where you can create a mass ingestion specification. To navigate to the creation wizard, select New from the menu.

Overview

Displays details on a mass ingestion specification. To navigate to the Overview page, select a mass ingestion specification on the Browse page.

On the Overview page, you can access the following views:

- Summary view. Review general mass ingestion specification properties.
- Deploy view. Deploy, redeploy, and run a mass ingestion specification.
- Execution History view. Browse the run instances for a mass ingestion specification.

Execution Statistics

Displays detailed information on the ingestion job for a mass ingestion specification. To navigate to the Execution Statistics page, navigate to the Overview page. On the Overview page, select a mass ingestion specification run instance.

On the Execution Statistics page, you can access the following views:

- Job Properties view. Review general mass ingestion specification properties for the run instance.
- Ingestion Objects view. Browse the source tables that are ingested in the mass ingestion specification.
- Ingestion Status view. View a summary of the mass ingestion job configured in the mass ingestion specification.
- Ingestion Statistics view. Review ingestion job statistics for a specific source table.

Help

Displays the help for the Mass Ingestion tool.

CHAPTER 2

Prepare

This chapter includes the following topics:

Prepare Overview

Before you can use the Mass Ingestion tool, an administrator must configure the Mass Ingestion Service in the Administrator tool.

The administrator must perform the following configuration tasks:

- Create the Mass Ingestion Service in the Informatica domain.
- Configure monitoring on the Mass Ingestion service.
- Assign the Administrator role to a user or group for the Mass Ingestion Service.

Creating a Mass Ingestion Service

When you create a Mass Ingestion Service, you must associate a Model Repository Service with the Mass Ingestion Service. A Model Repository Service cannot be associated with more than one Mass Ingestion Service.

Note: You must create the Mass Ingestion Service in a domain that uses native authentication. If you create the Mass Ingestion Service in a domain that uses LDAP or Kerberos authentication, you cannot log in to the Mass Ingestion tool.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. In the Domain Navigator, select the domain.
4. Click **Actions** > **New** > **Mass Ingestion Service**.

   The **New Mass Ingestion Service** wizard appears.

5. On the **New Mass Ingestion Service - Step 1 of 3** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Node	Node on which the service runs.

6. Click **Next**.
7. On the **New Mass Ingestion Service - Step 2 of 3** page, enter the following properties:

Property	Description
Model Repository Service	Model Repository Service to associate with the service.
User Name	User name that the service uses to access the Model Repository Service. Enter the Model repository user that you created.
Password	Password for the Model repository user.

8. Click **Next**.

   The **New Mass Ingestion Service - Step 3 of 3** page appears.
9. On the **New Mass Ingestion Service - Step 3 of 3** page, enter the following properties:

Property	Description
HTTP Port	Unique HTTP port number for the Mass Ingestion Service process when the service uses the HTTP protocol. Default is 9050.
Enable Transport Layer Security (TLS)	Enables the Transport Layer Security protocol to encrypt connections between the Mass Ingestion Service and external components. If you enable the TLS protocol, you must specify an HTTPS port and a keystore file. You do not specify an HTTP port.
HTTPS Port	Unique HTTPS port number for the Mass Ingestion Service process when the service uses the HTTPS protocol. When you set an HTTPS port number, you must also define the keystore file that contains the required keys and certificates.

Property	Description
Keystore File	Path and file name of the keystore file that contains the keys and certificates required if you use HTTPS connections for the Mass Ingestion Service. You can create a keystore file with a keytool. keytool is a utility that generates and stores private or public key pairs and associated certificates in a keystore file. You can use the self-signed certificate or use a certificate signed by a certificate authority.
Keystore Password	Password for the keystore file.

10. To enable the Mass Ingestion Service, select **Enable Service**.
11. Click **Finish**.

    The domain creates the Mass Ingestion Service. If you selected **Enable Service**, the domain enables the Mass Ingestion Service.
12. In the **Domain Navigator**, select the Mass Ingestion Service.
13. Click the URL to access the Mass Ingestion tool.

Configuring Monitoring

To monitor a mass ingestion specification, you must configure the domain to display statistics and reports about objects in the domain.

When you configure monitoring in the domain, the Data Integration Services that manage deployed mass ingestion specifications store statistics and reports in a Model repository. The statistics include historical information about the deployed specifications.

If you do not configure monitoring, you cannot monitor the ingestion status in the Mass Ingestion tool. In the Administrator tool, some views on the Manage and Monitor tabs do not have content and you cannot monitor the ingestion mapping jobs. The workflow graph is empty, and notifications disappear when you refresh the page.

To set up monitoring statistics and reports, perform the following tasks:

1. Configure monitoring settings. Configure a Model repository to store run-time statistics for objects that the Data Integration Services run.
2. Configure reports and statistics views. Choose which statistics appear in the **Statistics** and **Reports** views.

Step 1. Configure Monitoring Settings

Configure monitoring settings for the domain. When you configure monitoring settings, you specify a Model repository to store run-time statistics about deployed mass ingestion specification on the Data Integration Service.

Create the Model repository contents before you configure the monitoring settings. If you create the contents after you configure the monitoring settings, then you must recycle the Model Repository Service after the contents are created.

1. In the Administrator tool, click the **Manage** tab > **Services and Nodes** view.

2. Click the **Monitoring Configuration** view.

   The current monitoring configuration displays.

3. Click **Edit** to change the monitoring configuration.
4. Edit the following options:

Option	Description
Model Repository Service	Name of the Model repository that stores the historical information. The Model repository must not be integrated with a version control system.
Username	User name to access the Model Repository Service.
Password	Password of the user name to access the Model Repository Service.
Modify Password	Modify the Model Repository Service password.
Security Domain	Name of the security domain that the Model repository user belongs to.
Preserve Summary Historical Data	Number of days that the Model repository saves averaged data. If purging is disabled, then the Model repository saves the data indefinitely. Default is 180. Minimum is 0. Maximum is 366.
Preserve Detailed Historical Data	Number of days that the Model repository saves per-minute data. If purging is disabled, then the Model repository saves the data indefinitely. Default is 14. Minimum is 1. Maximum is 14.
Purge Statistics Every	Interval, in days, at which the Model Repository Service purges data that is older than the values configured in the **Preserve Historical Data** option. Default is 1 day.
Days At	Time of day when the Model Repository Service purges statistics. Default is 1:00 a.m.
Maximum Number of Sortable Records	Maximum number of records that can be sorted in the **Monitor** tab. If the number of records on the **Monitor** tab is greater than this value, then you can only sort by **Start Time** and **End Time**. Default is 3,000.
Maximum Delay for Update Notifications	Maximum time, in seconds, that the Data Integration Service buffers statistics before it stores them in the Model repository and displays them in the **Monitor** tab. If the Data Integration Service shuts down unexpectedly before it stores the statistics in the Model repository, then the statistics are lost. Default is 10.
Show Milliseconds in Date Time Field	Include milliseconds for date and time fields in the **Monitor** tab.

5. Click **OK**.
6. Click **Save**.

To apply the settings, you must restart all of the Data Integration Services.

Step 2. Configure Reports and Statistics

By default, the ingestion job statistics for a mass ingestion specification are empty in the Mass Ingestion tool. To view statistics and reports, you must configure the Report and Statistic settings in the domain. These settings apply to all Data Integration Services in the domain.

Before you configure statistics and reports, you must specify a Model Repository Service in the Monitoring Configuration and enable the Model Repository Service.

1. In the Administrator tool, click the **Monitor** tab.
2. From the **Actions** menu, select **Report and Statistic Settings**.
3. On the **Statistics** tab, configure the time ranges that you want to use for statistics, and then select the frequency at which the statistics assigned to each time range should be updated.
4. Select a default time range to appear for all statistics.
5. Click the **Reports** tab.
6. Enable the time ranges that you want to use for reports, and then select the frequency at which the reports assigned to each time range should be updated.
7. Select a default time range to appear for all reports, and then click **OK**.
8. Click **Select Reports**.
9. Add the reports that you want to run to the **Selected Reports** box.
10. Organize the reports in the order in which you want to view them on the **Monitor** tab.
11. Click **OK** to close the **Select Reports** window.
12. Click **OK** to the settings and close the **Report and Statistic Settings** window.

Assigning the Administrator Role

Assign the Administrator role to a user or group for the Mass Ingestion Service. The Administrator role assigned to the user or group also determines tasks that the user or group can complete in the Mass Ingestion tool.

1. In the Administrator tool, click the **Security** tab.
2. In the Navigator, select a user or group.
3. Click the **Privileges** tab.
4. Click Edit.

   The **Edit Roles and Privileges** dialog box appears.
5. To assign the Administrator role, expand the Mass Ingestion Service on the **Roles** tab.
6. To grant the role, select the Administrator role to assign to the user or group for the domain or application service.
7. To revoke the role, clear the roles assigned to the user or group.
8. Click **OK**.

CHAPTER 3

Create

This chapter includes the following topics:

Create Overview

Create a mass ingestion specification in the Mass Ingestion tool.

When you create a mass ingestion specification, you configure the job that ingests data from a relational database to a Hive or HDFS target.

To create the mass ingestion specification, you use a wizard that helps you define the ingestion job. In the wizard, you configure different properties on each of the following pages:

1. Definition. The type of specification that you want to create.
2. Source. The source data that you want to ingest.
3. Target. The Hive or HDFS target where you want to ingest data.
4. Mass Transformations. Parameters that apply across all source tables that you ingest.
5. Transformation Override. Parameters that apply to specific source tables that you ingest. These parameters override the parameters that you configure as mass transformations.

As you configure the mass ingestion specification, the Mass Ingestion Service validates the inputs. When you save the specification, the specification is stored in the Model repository.

Incremental Data

You can configure a mass ingestion specification to ingest incremental data, which is data that has been modified in between runs of the specification.

To ingest incremental data, enable incremental load in the specification and configure the following incremental load options:

- Incremental key. The key that the Spark engine uses to fetch incremental data.
- Incremental mode. The mode that the Spark engine uses to ingest incremental data.

If you ingest data to a Hive target, you can also propagate schema changes to accommodate for schema drift in the relational database.

If you do not enable incremental load, you can use only full loads to ingest the data. A full load ingests all of the data to the target each time that the specification runs.

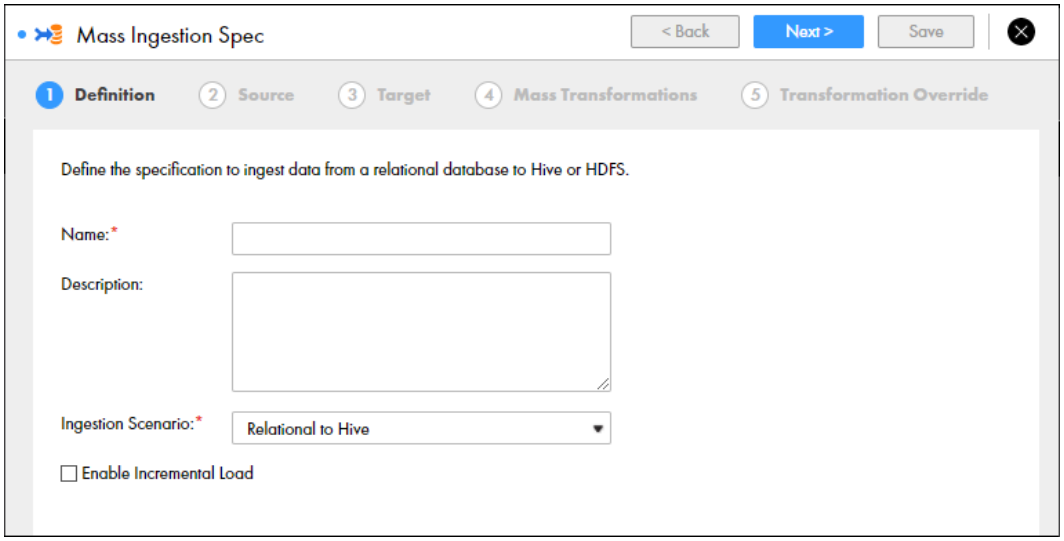
Definition

The definition describes the type of mass ingestion specification that you want to create.

In the definition, you specify a name and a description and select the scenario for the specification. The scenario determines whether the specification ingests data from a relational database to a Hive target or from a relational database to an HDFS target. The scenario that you specify defines the properties that you can configure in the rest of the specification.

You can also enable incremental load. If you enable incremental load, you can ingest incremental data when you run the specification. After you enable incremental load, you must configure incremental load options when you configure the relational source and the Hive or HDFS target.

The following image shows the Definition page of the mass ingestion specification:



The following table describes the properties that you can configure in the definition:

Property	Description
Name	Required. Name of the mass ingestion specification. The name is not case sensitive and must be unique. You cannot edit the name after you create the mass ingestion specification.
Description	Optional. Description of the specification.
Ingestion Scenario	Required. The scenario where data ingestion takes place. You can ingest data from a relational database to a Hive target or from a relational database to an HDFS target. You cannot edit the ingestion scenario after you create the mass ingestion specification.
Enable Incremental Load	Optional. Allows you to ingest data using an incremental load when you run the specification.

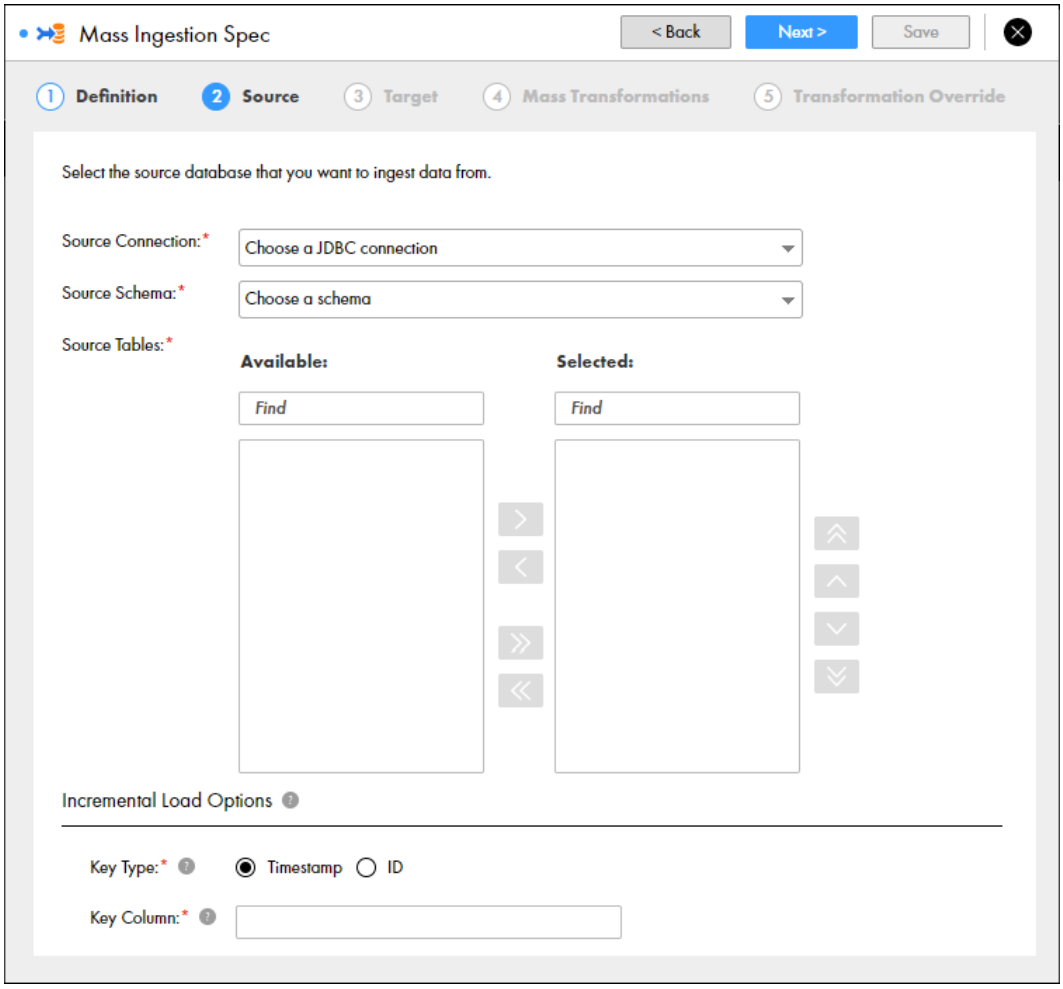
Source

The source describes the source data that you want to ingest from a relational database.

To define the source, select a JDBC connection and a schema to locate the database that you want to use. Then specify the tables that you want to ingest.

If you enable incremental load in the definition of the mass ingestion specification, you must configure the incremental load options for the source to define an incremental key.

The following image shows the Source page of the mass ingestion specification:



The following table describes the properties that you can configure to define the source:

Property	Description
Source Connection	Required. The JDBC connection used to find the source schema. If changes are made to the available JDBC connections, refresh the browser or log out and log back in to the Mass Ingestion tool.
Source Schema	Required. The schema that defines where the source tables are stored.
Source Tables	Required. The tables that will be ingested. Select the tables in the Available column and move the tables to the Selected column.
Key Type	Required if you enable incremental load. The type of incremental key. Select Timestamp or ID. Default is Timestamp.
Key Column	Required if you enable incremental load. The column name of the incremental key. If the key is a timestamp, the key column must store date/time data. If the key is an ID, the key column must store numeric data.

Incremental Keys

If you enable incremental load in the definition of the mass ingestion specification, you must configure an incremental key.

An incremental key is a column in a source table that the Spark engine uses to fetch incremental data from the source. The incremental key might be a timestamp or an ID depending on how your organization manages updates to records.

Timestamp Keys

If the incremental key is a timestamp, the incremental key column must store date/time data, and the timestamp must indicate the last time that the row of data was modified.

When you run the mass ingestion specification, the Spark engine fetches the rows in the source table with a timestamp that is more recent than the oldest timestamp that was previously ingested. If the timestamp for a row in the table is more recent than the oldest timestamp that was ingested, the Spark engine fetches the row associated with the timestamp as incremental data.

For example, you might have ingested the following source table in the previous run of the specification:

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

Note that the oldest timestamp is `01/27/2017 02:43:05`.

The following table shows the current data in the source table:

EmpLastName	LastModified
'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

Since the oldest timestamp that was ingested is `01/27/2017 02:43:05`, the Spark engine fetches the rows from the source table with a timestamp that is more recent than `01/27/2017 02:43:05`.

In the current source table, there are two timestamps that are more recent: `10/22/2018 04:20:57` and `09/13/2018 04:24:26`, so the rows that are associated with these timestamps are incremental data.

When you run the specification, the Spark engine ingests the following rows of data:

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

ID Keys

If the incremental key is an ID, the incremental key column must store numeric data, and the ID must indicate that a new row of data has been added to the source table.

When you run the mass ingestion specification, the Spark engine fetches the rows in the source table with an ID that is higher than the maximum ID value of the rows that have been previously ingested. If the ID value for a row in the table is greater than the maximum ID that has been ingested, the Spark engine fetches the row associated with the ID as incremental data.

For example, you might have ingested the following source table in the previous run of the specification:

EmpID	EmpLastName
481530	'Basquez'
481531	'Savage'
481532	'Greene'

Note that maximum ID value is 481532.

The following table shows the current data in the source table:

EmpID	EmpLastName
481530	'Basquez'
481531	'Savage'
481532	'Greene'
481533	'Caldwell'
481534	'Galloway'

Because the IDs `481533` and `481534` are greater than the maximum ID `481532` that was previously ingested, the rows that are associated with these IDs are incremental data.

When you run the specification, the Spark engine ingests the following rows of data:

EmpID	EmpLastName
481533	'Caldwell'
481534	'Galloway'

Target

The target describes the Hive or HDFS target where you want to ingest data in the source tables.

The target is defined by the scenario that you select on the Definition page. Use the Target page to configure the target properties.

If you enable incremental load in the mass ingestion specification, you must configure incremental load options for the target. Use the incremental load options to specify a mode to ingest data to the target and whether you want to propagate schema changes on the source.

Hive Target

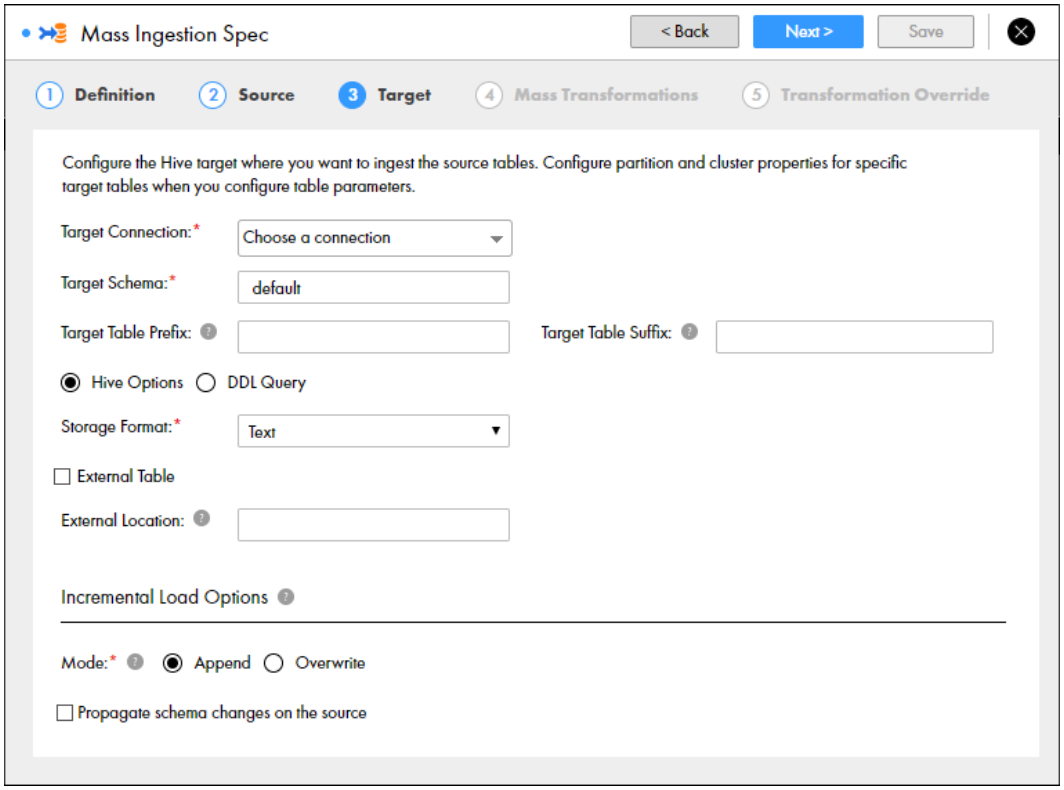
Configure a Hive target to ingest source data to Hive target tables.

When you configure the mass ingestion specification to ingest data to a Hive target, you configure a Hive connection and Hive properties to define the target.

You can ingest data to an internal or external Hive table. Internal Hive tables are managed by Hive. External Hive tables are unmanaged tables. You can specify an external location for an external Hive table such as Amazon S3, Microsoft Azure Data Lake Store, or HBase.

If you enable incremental load in the definition of the mass ingestion specification, you must configure incremental load options for the Hive target to select a mode to ingest the data. You can also choose to propagate schema changes on the source.

The following image shows the Target page for a Hive target:



The following table describes the properties that you can configure to define the Hive target:

Property	Description
Target Connection	Required. The Hive connection used to find the Hive storage target. If changes are made to the available Hive connections, refresh the browser or log out and log back in to the Mass Ingestion tool.
Target Schema	Required. The schema that defines the target tables.
Target Table Prefix	The prefix added to the names of the target tables. Enter a string. You can enter alphanumeric and underscore characters. The prefix is not case sensitive.
Target Table Suffix	The suffix added to the names of the target tables. Enter a string. You can enter alphanumeric and underscore characters. The prefix is not case sensitive.
Hive Options	Select this option to configure the Hive target location.
DDL Query	Select this option to configure a custom DDL query that defines how data from the source tables is loaded to the target tables.
Storage Format	Required. The storage format of the target tables. You can select Cluster default, Text, Avro, Parquet, or ORC. Default is Cluster default. If you select Cluster default, the specification uses the default storage format on the Hadoop cluster.
External Table	Select this option if the table is external.
External Location	The external location of the Hive target. By default, tables are written to the default Hive warehouse directory. A sub-directory is created under the specified external location for each source that is ingested. For example, you can enter `/temp`. A source table named `PRODUCT` is ingested to the external location `/temp/PRODUCT/`.
Mode	Required if you enable incremental load. Select Append or Overwrite. Append mode appends the incremental data to the target. Overwrite mode overwrites the data in the target with the incremental data. Default is Append.
Propagate schema changes on the source	Optional. If new columns are added to the source tables or existing columns are changed, the changes are propagated to the target tables.

Configure partition and cluster properties for specific target tables when you configure the transformation override.

When you ingest to a Hive target, consider the following guidelines:

- If a source table is ingested to a Hive target and the name of the source table contains a reserved keyword on Hive, the data in the source table is ingested to a target table that has a randomly-generated name.
- A source table cannot be ingested into Hive if the table metadata uses UTF-8 characters. To resolve the issue, configure the Hive metastore for UTF-8 data processing.

- A source table cannot be ingested to an Avro file in a Hive target if the source table contains a column with a timestamp data type or the incremental load is configured with a timestamp key. To ingest timestamp data to an Avro file, the third-party Hive JDBC driver must have a Hive version higher than 1.1.
- When you run a full load to ingest data to a Hive target in an external location, all rows in the source table are added to the target Hive table. For example, if the source table contains 500 rows and you run a full load twice, the Hive table contains 1000 rows. To reset the table, you must clear the data in the external location.

DDL Query

When you configure a mass ingestion specification to ingest data to a Hive target, you can configure a custom DDL query to define how data from the source tables is loaded to the target tables.

You can define the DDL query to customize the target table or specify additional parameters. The target table contains the columns that you define in the DDL query.

To define a DDL query, use SQL statements and placeholders. Use the placeholders to fetch the table name, column list, and column names. The Data Integration Service substitutes the placeholders with actual values at run time according to the tables that you ingest. You must enclose the placeholders within curly brackets. For example, {INFA_TABLE_NAME}.

You can use the following placeholders:

INFA_TABLE_NAME

Fetches the target table name at run time.

INFA_COLUMN_LIST

Fetches a list of columns in the target table at run time.

For example, you might ingest a table `CUSTOMER`. To define how you want to ingest the table in the target, you can enter the following DDL query:

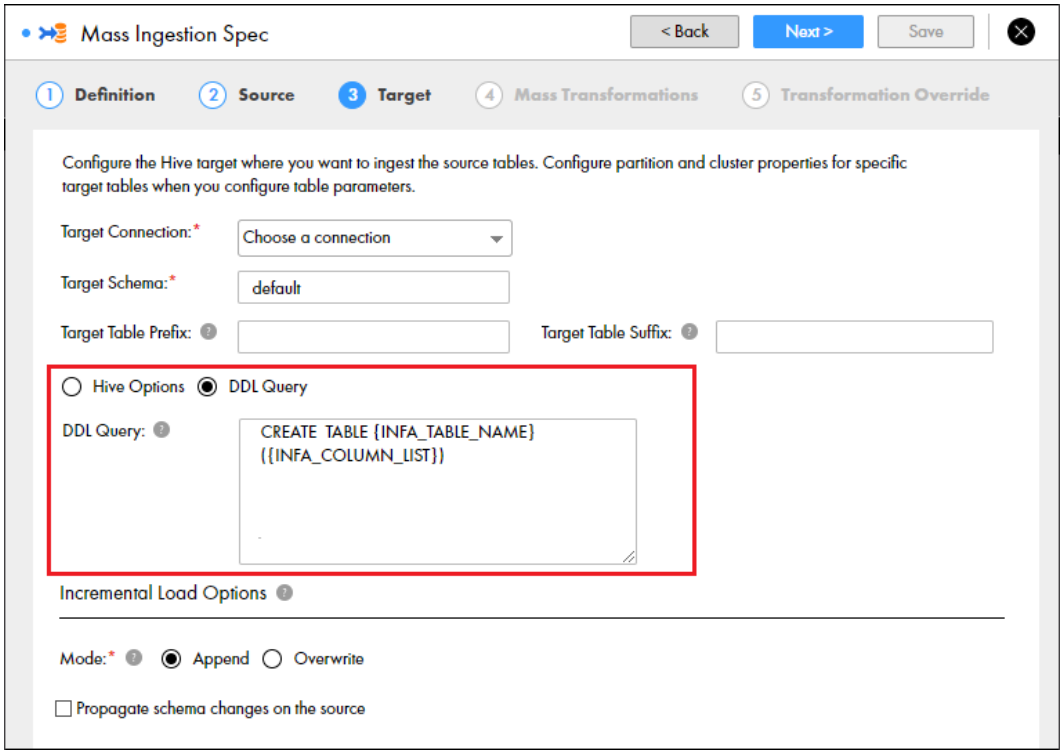
```
CREATE TABLE {INFA_TABLE_NAME} ({INFA_COLUMN_LIST}) CLUSTERED BY (LAST_NAME) INTO 10
BUCKETS STORED AS TEXT
```

At run time, the Data Integration Service substitutes `{INFA_TABLE_NAME}` with `CUSTOMER`, and it substitutes `{INFA_COLUMN_LIST}` with the list of columns that appear in the table `CUSTOMER`. The Data Integration Service might expand the DDL query to the following query:

```
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
```

Note: You cannot use a placeholder to specify the partition columns and clustered by columns. When you specify the partition columns and clustered by columns, enter the column name in the DDL query.

The following image shows the option to configure a DDL query for a Hive target:



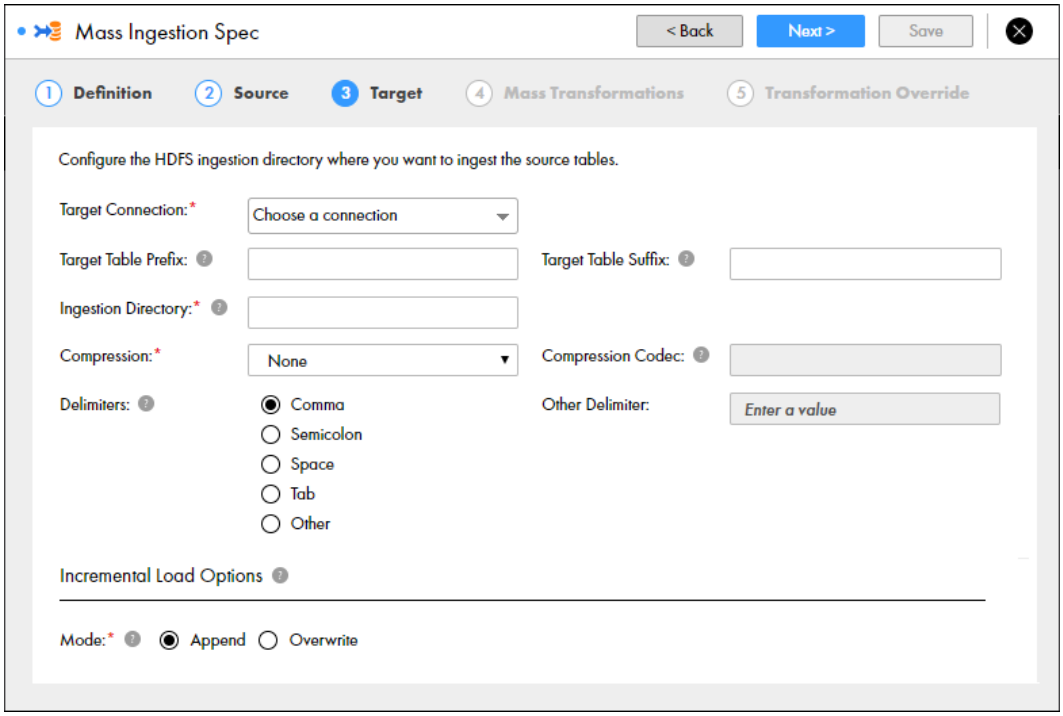
HDFS Target

Configure an HDFS target to ingest source data to a flat file on HDFS.

When you configure the mass ingestion specification to ingest data to an HDFS target, you configure an HDFS connection and an ingestion directory to define the target.

If you enable incremental load in the definition of the mass ingestion specification, you must configure incremental load options for the HDFS target to select a mode to ingest the data.

The following image shows the Target page for an HDFS target:



The following table describes the properties that you can configure to define the HDFS target:

Property	Description
Target Connection	Required. The HDFS connection used to find the HDFS storage target. If changes are made to the available HDFS connections, refresh the browser or log out and log back in to the Mass Ingestion tool.
Target Table Prefix	The prefix added to the names of the target files. Enter a string. You can enter alphanumeric and underscore characters. The prefix is not case sensitive.
Target Table Suffix	The suffix added to the names of the target files. Enter a string. You can enter alphanumeric and underscore characters. The prefix is not case sensitive.
Ingestion Directory	Required. The target directory on HDFS. A sub-directory is created under the ingestion directory for each source that is ingested. If the specified directory already exists, the directory is replaced. For example, you can enter `/temp`. A source table named `PRODUCT` is ingested to the directory `/temp/PRODUCT/`.
Compression	Required. The compressed file format that stores the target files. You can select None, Gzip, Bzip2, LZO, Snappy, or Custom. If you select Custom, enter the compression codec. Default is None.
Compression Codec	If you select custom compression, enter the fully qualified class name implementing the Hadoop CompressionCodec interface.

Property	Description
Delimiters	The delimiters used to separate data in the target files. You can select comma, semicolon, space, tab, or other. If you select Other, you can define a custom delimiter.
Other Delimiter	Required if you choose `Other` for the delimiter. Enter a custom delimiter.
Mode	Required if you enable incremental load. Select Append or Overwrite. Append mode appends the incremental data to the target. Overwrite mode overwrites the data in the target with the incremental data. Default is Append.

Note: When the Data Integration Service stores temporary files that you ingest to an HDFS target, it appends a unique ID to the original file name. The resulting file name can have a maximum length of 255 characters.

Compression Codec

When you configure a mass ingestion specification to ingest data to an HDFS target directory, you can configure a compression codec to write the ingested data to a compressed file.

You can select one of the following compression options:

- Gzip
- Bzip2
- LZO
- Snappy
- Custom

If you specify a custom compression codec, you must specify the fully qualified class name implementing the Hadoop `CompressionCodec` interface.

Incremental Load Options

If you enable incremental load in the definition of the mass ingestion specification, you must configure incremental load options for the target.

The incremental load options for the target include an incremental mode. The Spark engine uses the incremental mode to determine how to load incremental data to the target.

If you ingest data to a Hive target, you can also configure the incremental load options to propagate schema changes on the source. Propagating schema changes accommodates for schema drift in the relational database, such as new columns that have been added to the source tables.

Incremental Modes

If you enable incremental load in a mass ingestion specification, you must select a mode to ingest the data.

You can select append mode or overwrite mode. Append mode appends the incremental data to the data in the target. Overwrite mode overwrites the data in the target with the incremental data. You might use overwrite mode if the target is a staging area and not the final operational data store.

For example, the target might contain the following data from a previous run of the specification:

EmpID	EmpLastName
481530	'Basquez'
481531	'Savage'
481532	'Greene'

The following table shows the incremental data that the Spark engine ingests in the current run of the specification:

EmpID	EmpLastName
481533	'Caldwell'
481534	'Galloway'

If you use append mode, the incremental data is appended to the target. When the ingestion job is complete, the target contains the following data:

EmpID	EmpLastName
481530	'Basquez'
481531	'Savage'
481532	'Greene'
481533	'Caldwell'
481534	'Galloway'

If you use overwrite mode, the incremental data overwrites the target. When the ingestion job is complete, the target contains the following data:

EmpID	EmpLastName
481533	'Caldwell'
481534	'Galloway'

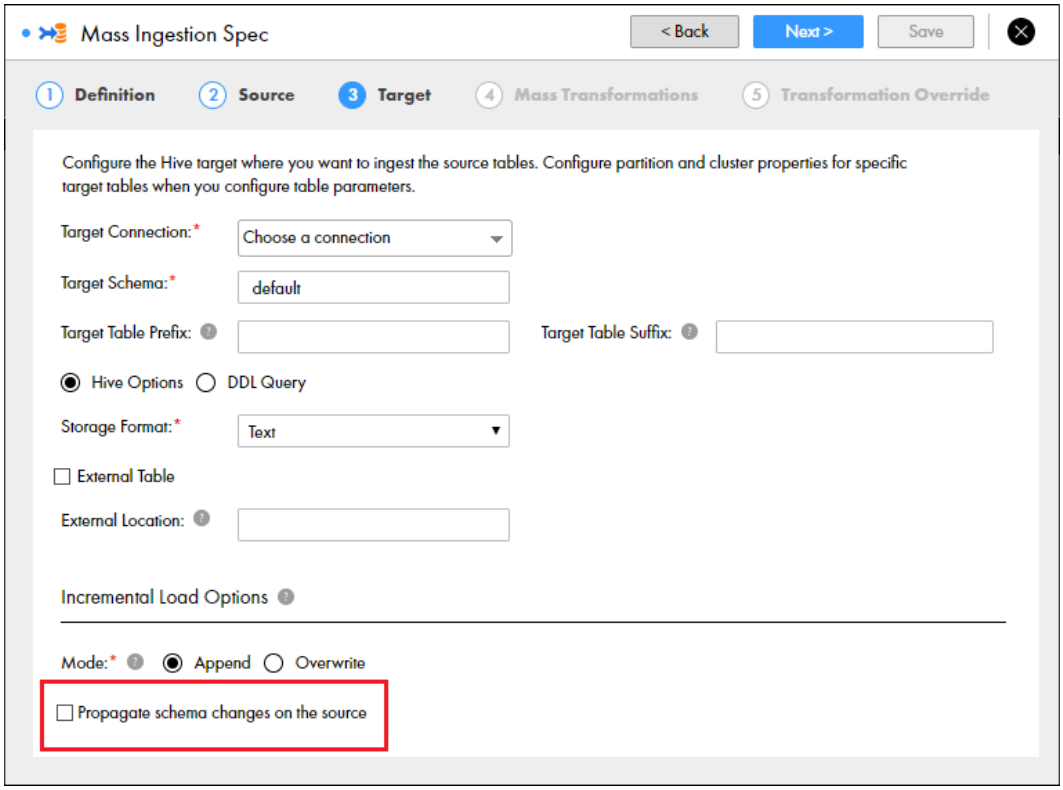
Propagate Schema Changes

When you ingest data to a Hive target, you can propagate schema changes on the source to accommodate for schema drift in the relational database.

Schema drift occurs when you change the schema of a source table. For example, you might add new columns or change existing columns in the table. It can be a time-intensive task to revise the schemas of the target tables to match the schemas of the source tables before you perform each load of the data.

To accommodate for potential schema drift in the source tables, you can configure incremental loads to propagate schema changes on the source to the target. If the source tables have new columns or changes to existing columns, this option propagates the changes to the target tables.

The following image shows the property that you can use to propagate schema changes:



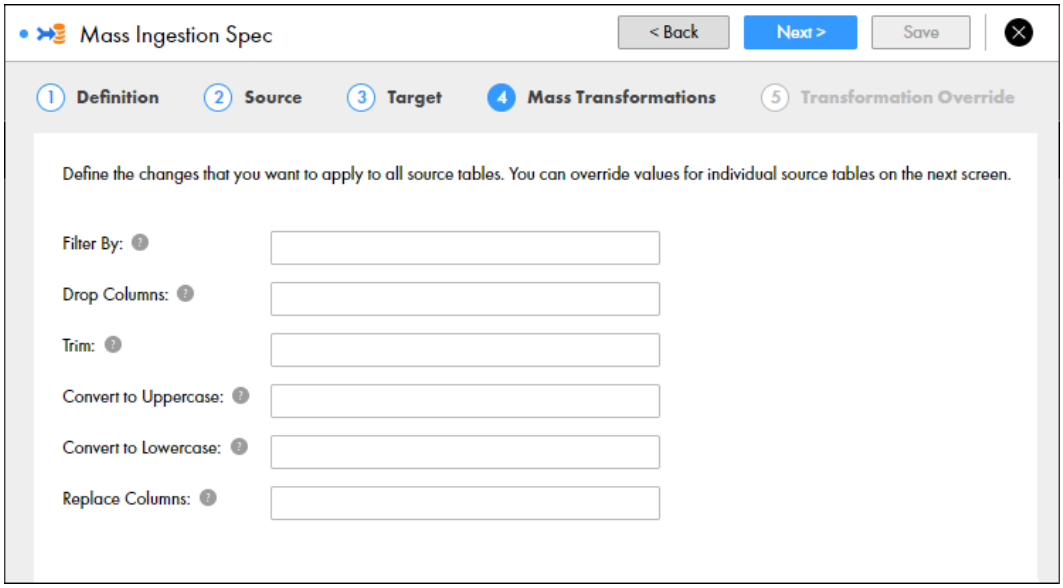
For more information about propagating schema changes, see the *Informatica PowerExchange for Hive User Guide*. Mass ingestion leverages the APPLYNEWSCHEMA target schema strategy.

Mass Transformations

Configure mass transformations to define any parameters that you want to apply across all source tables that you ingest from a relational database.

To set parameters for specific tables, configure the transformation override.

The following image shows the Mass Transformations page of the mass ingestion specification:



The following table describes the parameters that you can configure:

Property	Description
Filter By	Filters rows in the target table based on criteria for a column in the table. Enter a filter clause to determine the criteria, such as `STATE='California'`. You can use any transformation language functions in the filter clause. The filter clause must evaluate to TRUE or FALSE. To set criteria for multiple columns, use the operators AND and OR. For example, use the operator AND to specify a filter clause such as `STATE='California' AND STATUS='Single'`.
Drop Columns	Drops columns in the target table. Enter each column as a string and separate column names with a comma, or specify a regular expression. For example, if you enter `COL1`, the column COL1 will be dropped in the target table. If you enter a regular expression such as `.*SSN.*`, columns that have SSN in the column name will be dropped.
Trim	Trims column values in the target table to remove spaces before and after the values. Enter each column as a string and separate column names with a comma, or specify a regular expression. For example, if you enter `COL1`, values in the column COL1 will be trimmed in the target table. If you enter a regular expression such as `.*SSN.*`, values in columns that have SSN in the column name will be trimmed.
Convert to Uppercase	Converts column values in the target table to uppercase. Enter each column as a string and separate column names with a comma, or specify a regular expression. For example, if you enter `COL1`, values in the column COL1 will be converted to upper case in the target table. If you enter a regular expression such as `.*SSN.*`, values in columns that have SSN in the column name will be converted to upper case.

Property	Description
Convert to Lowercase	Converts column values in the target table to lowercase. Enter each column as a string and separate column names with a comma, or specify a regular expression. For example, if you enter `COL1`, values in the column COL1 will be converted to lower case in the target table. If you enter a regular expression such as `.*SSN.*`, values in columns that have SSN in the column name will be converted to lower case.
Replace Columns	Replaces column values in the target table. Enter each column as a string and separate column names with a comma, or specify a regular expression. For example, if you enter `COL1`, values in the column COL1 will be replaced in the target table. If you enter a regular expression such as `.*SSN.*`, values in columns that have SSN in the column name will be replaced. If you specify columns to replace, you must specify the replace criteria.
Replace Criteria	Required if you specify columns to replace. Determines how to replace column values in the target table. You can select `Pattern` or `Entire String`. If you select `Entire String`, all values in the columns that you specify are replaced by the value that you configure. If you select `Pattern`, enter the pattern to be replaced. Then enter the value to replace the pattern.

Pattern	Required if you configure the replace criteria to be a pattern. Determines the pattern to replace. The pattern must be a regular expression. For example, if you want to replace the values in columns that contain Social Security numbers, you can enter the pattern:
	`^\d(3)-?\d(2)-?\d(4)$`
	The pattern replaces all values that correspond to the pattern. In the column that contains Social Security numbers, the pattern replaces the entire Social Security number.
	To mask only the first five digits of the Social Security number, you can enter the pattern:
	`^\d(3)-?\d(2)$`

Value	Required if you configure columns to replace. Replaces the pattern or the entire string according to the configured criteria.
	If you select `Entire String` for the replace criteria, the value that you enter replaces the values in all of the columns that you specify to replace. For example, if you replace COL1 and COL2 and enter the value `XXX`, all values in columns COL1 and COL2 are replaced with the value XXX.
	If you select `Pattern` for the replace criteria, the value that you enter replaces all values that correspond to the pattern.
	For example, you want to mask the entire Social Security number according to the pattern:
	`^\d(3)-?\d(2)-?\d(4)$`
	Enter the value `XXX-XX-XXXX`. All values in the SSN column will appear as XXX-XX-XXXX.
	You might also want to mask only the first five digits of the Social Security number according to the pattern
	`^\d(3)-?\d(2)$`
	Enter the value `XXX-XX`. The first five digits of every Social Security number will appear as XXX-XX. For example, if the original Social Security number is 123-45-6789, the replaced value is XXX-XX-6789.

If you configure parameters for a table column that does not exist in all of the tables, the tables where the column does not exist will fail to be ingested. You must reconfigure the parameters for each table where the column does not exist when you configure the transformation override.

Filter Clauses

The filter clause determines how rows are filtered in the target table based on criteria for a column in the table.

To design a filter clause, you can specify the column names in the target table and any transformation language functions that use the column names as arguments. To set multiple criteria, use the operators AND and OR. The filter clause must evaluate to TRUE or FALSE.

For example, you might ingest the following source table:

EMPLOYEEID	PHONENUMBER
607014	(630)4468851
620368	(904)3854084
698107	(549)5694371
621861	(904)9062721

To filter the rows that you ingest to the target, you might use the following filter clause to filter the rows by area code `(904)`:

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

The rows in the table are filtered by area code `(904)`. The following table is ingested to the target:

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

For more information about transformation language functions, see the "Functions" chapter in the *Informatica Transformation Language Reference*.

Regular Expressions

A regular expression describes a range or pattern of values.

You can use a regular expression specify the columns that you want to parameterize in a mass ingestion specification. Use a regular expression when the columns in different source tables have varying names but contain the same information. If you choose to replace columns, you also use a regular expression to specify the pattern in the replace criteria.

For example, you might want to drop the columns that contain Social Security numbers. All of the column names contain `SSN`, but the column names have different prefixes depending on the source table where a column appears. To specify all variations in the column names, you can use a regular expression such as `.*SSN`.

Replace Criteria

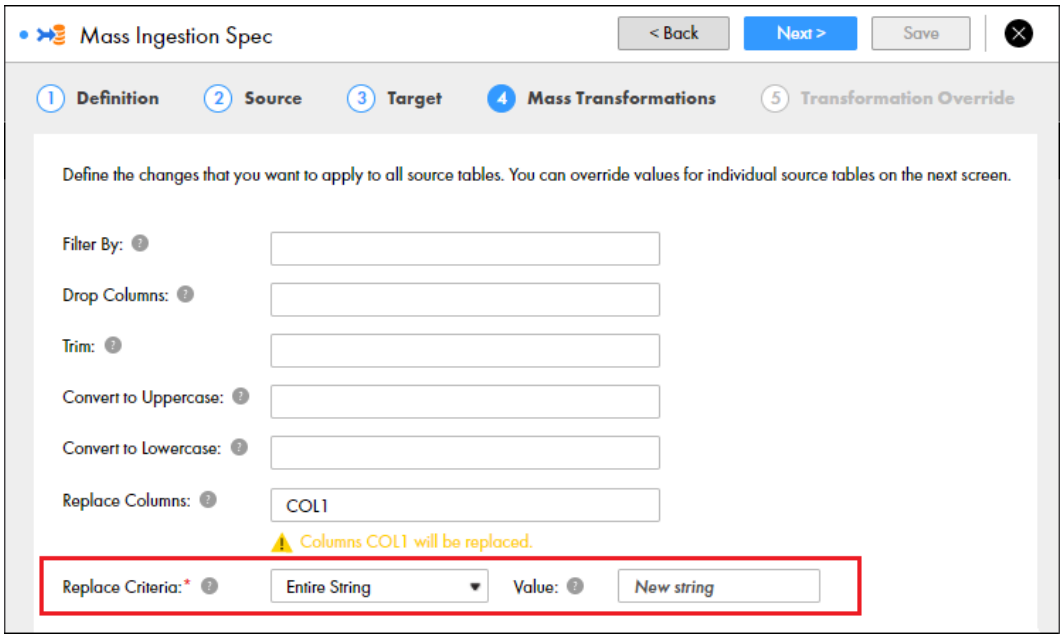
The replace criteria determines how to replace column values in the target table.

Configure the replace criteria to replace values in columns according to the entire string or a pattern.

Replacing an Entire String

If you configure the replace criteria to replace an entire string, all values in the columns that you specify are replaced according to the new value that you configure. For example, you configure the criteria to replace values in columns `COL1` and `COL2` and configure the new value to be `XXX`. All values in columns `COL1` and `COL2` are replaced with the value `XXX`.

The following image shows the options to configure the replace criteria based on an entire string:



Replacing a Pattern

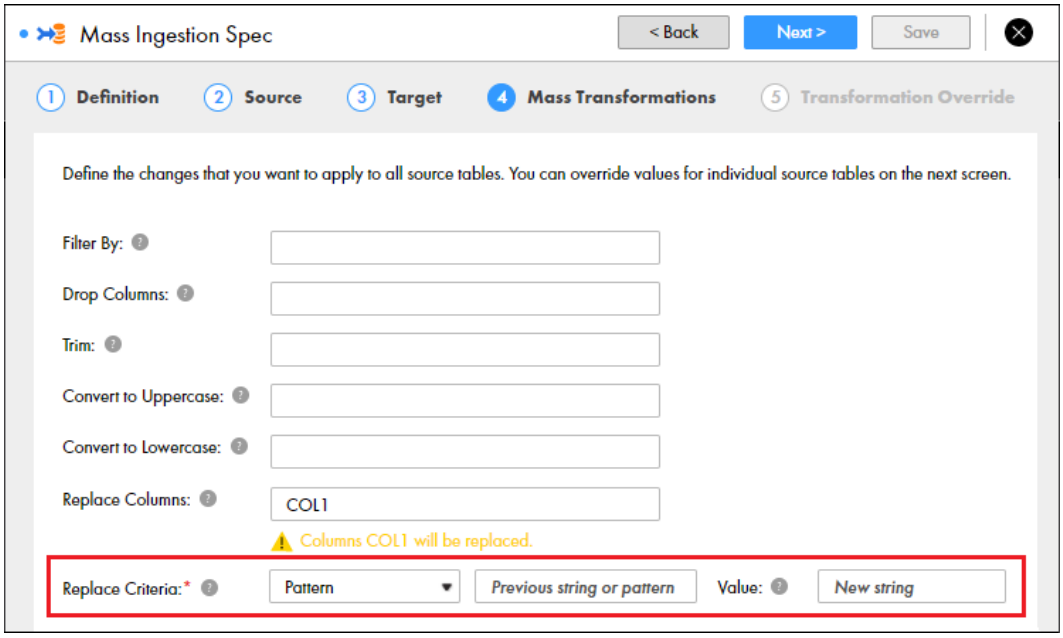
If you configure the replace criteria to replace a pattern, you enter a pattern and a new value to replace values that match the pattern. The pattern that you enter must be a regular expression.

For example, you want to mask all values in a column that contains IP addresses. To match an IP address, you can enter the following pattern:

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

All values that match the pattern are replaced according to the new value that you configure. If you configure the new value `XXX`, all values that match the pattern are replaced with `XXX`.

The following image shows the options to configure the replace criteria based on a pattern:



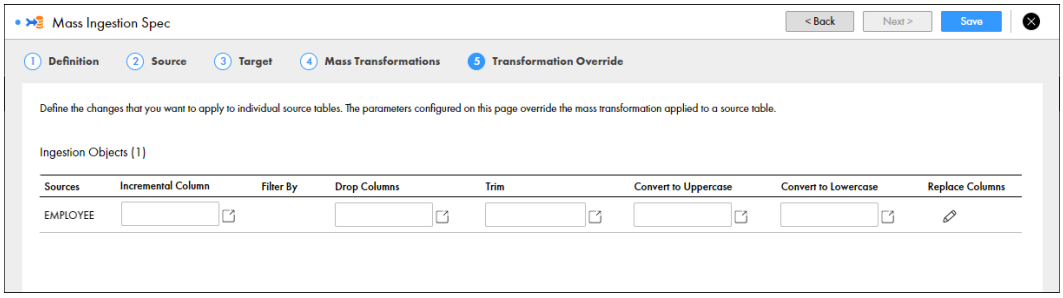
Transformation Override

Configure the transformation override to override the mass transformations that are applied to all source tables in a mass ingestion specification. When you configure the override, you can configure parameters to apply to specific tables that you ingest from the relational database.

You can configure the transformation override in one of the following ways:

- Before you configure the transformation override, configure mass transformations to apply parameters to all source tables. Then, navigate to each table and edit the applied parameter.
- Clear the mass transformations, and configure specific parameters for each table.

The following image shows the Transformation Override page for an HDFS target:



You can edit the following parameters for each source table:

- Incremental Column. If incremental load is enabled, edit the column that determines incremental data in the source tables.
- Filter By. Filter the columns in the source tables.
- Drop Columns. Drop columns from the source tables.

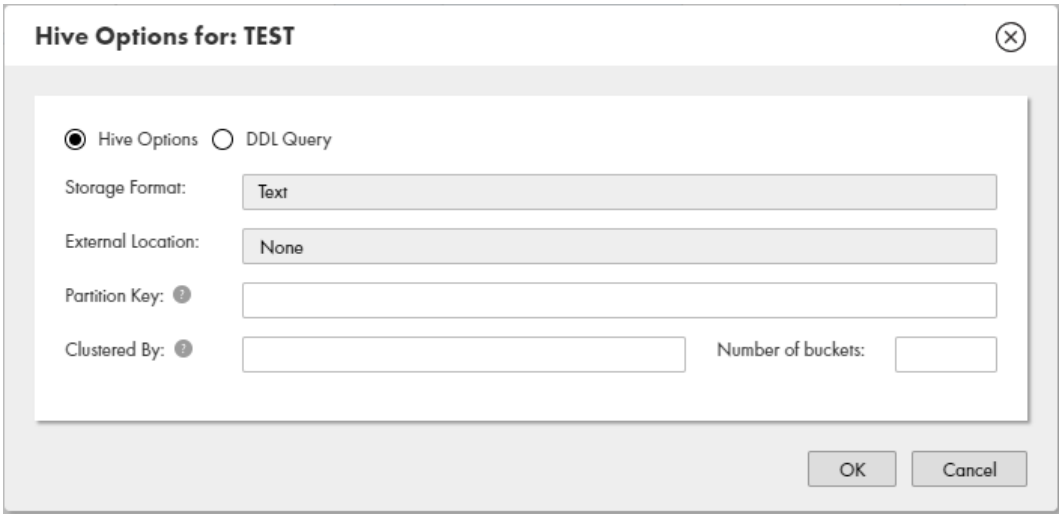
- Replace Columns. Replace values in the source table columns.
- Trim. Trim spaces from the beginning and end of values in source table columns.
- Convert to Uppercase. Convert values in source table columns to uppercase.
- Convert to Lowercase. Convert values in source table columns to lowercase.
- Hive. If the specification ingests data to a Hive target, specify Hive properties for the target tables.

When you configure parameters for a source table, you cannot configure different parameters for different columns in the table. If you configure parameters for a column that does not exist in the table, the table fails to be ingested.

Hive Options

When you configure a transformation override for a Hive target, you can configure additional parameters to specify how the data is loaded to the Hive target tables.

The following image shows the dialog box that appears when you edit the properties for the Hive target table:

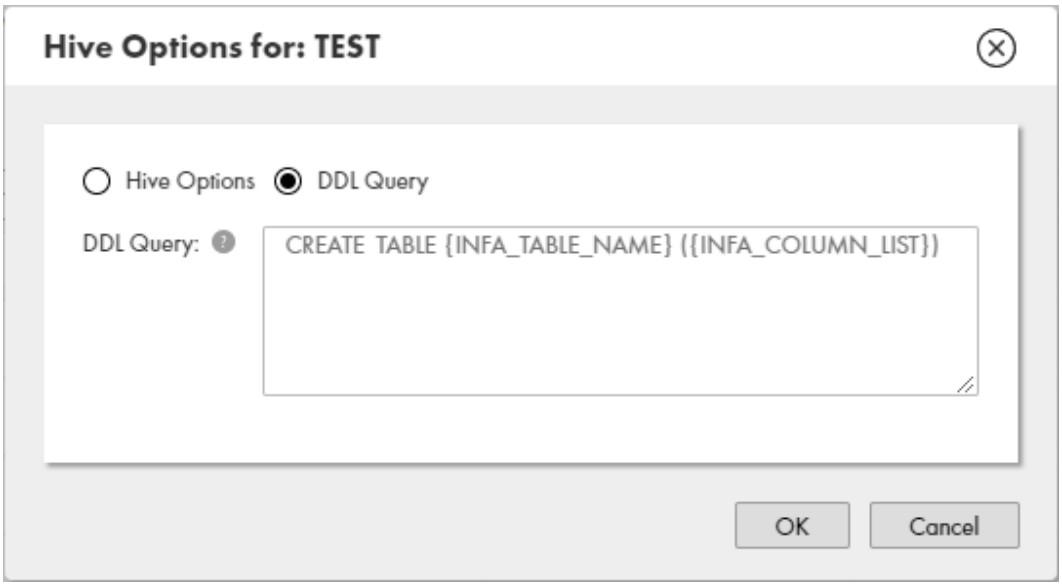


The following table describes the Hive target table properties:

Property	Description
Hive Options	Select this option to configure the Hive target location.
DDL Query	Select this option to configure a custom DDL query that defines how data from the source tables is loaded to the target tables.
Storage Format	Required. The storage format of the target tables. You can select Text, Avro, Parquet, or ORC. Default is Text.
External Location	The external location of the Hive target. Enter an external location to specify a location other than the default Hive warehouse directory. A sub-directory is created under the specified external location for each source that is ingested. For example, you can enter `/temp`. A source table named `PRODUCT` is ingested to the external location `/temp/PRODUCT/`.
Partition Key	The partition key for the target Hive table.

Property	Description
Clustered By	The cluster key for the target Hive table.
Number of buckets	Required if you specify a cluster key. The number of buckets to divide the target Hive table.

The following image shows the dialog box that appears when you edit the DDL query in the properties for the Hive target table:



To edit the DDL query, use SQL statements and the following placeholders:

INFA_TABLE_NAME

Fetches the target table name at run time.

INFA_COLUMN_LIST

Fetches a list of columns in the target table at run time.

Creating a Mass Ingestion Specification

Create a mass ingestion specification in the Mass Ingestion tool.

1. In the Mass Ingestion tool, click **New** to create a new mass ingestion specification.

   The **Definition** page for a new mass ingestion specification appears.
2. Design the mass ingestion specification by configuring the properties on each page in the wizard.
3. Click **Save**.

   The **Overview** page for the mass ingestion specification appears.

Configuring the Definition

Configure the definition to determine the type of mass ingestion specification that you want to create.

1. On the **Definition** page, enter the name and description for the mass ingestion specification.
2. Choose one of the following ingestion scenarios:
   - Relational to Hive. Select this scenario to ingest data from a relational database to a Hive target.
   - Relational to HDFS. Select this scenario to ingest data from a relational database to an HDFS target.
3. Optionally, select **Enable Incremental Load** if you expect to run incremental loads on the data.

   If you do not enable incremental load, you can use only full loads to ingest the data.

Configuring the Source

Configure the source to specify the relational database and the relational source tables that you want to ingest.

1. On the **Source** page, select the source connection and the source schema.

   The source tables that are available in the schema appear under the **Selected** column.
2. To add source tables to the mass ingestion specification, drag the source tables from the **Available** column to the **Selected** column.
3. If you enabled incremental load in the definition of the mass ingestion specification, configure the incremental key using the following properties:
   a. Key type. Select **Timestamp** or **ID**. Default is Timestamp.
   b. Key column. Enter the name of the column that stores the incremental key. If the key is a timestamp, the column must store date/time data. If the key is an ID, the column must store numeric data.

Configuring a Hive Target

If the mass ingestion specification ingests data to a Hive target, configure the Hive target to specify how you want to ingest data to the target tables.

1. Select the target connection.
2. Enter the target schema.
3. Optionally, enter the prefix and suffix to add to the target tables.
4. Choose one of the following options:
   - Select **Hive Options**.
     1. Select the storage format. Default is Text.
     2. If the target is an external Hive target, select **External Table** and enter the external location.
   - Select **DDL Query** and enter a custom DDL query.
5. If you enabled incremental load in the definition of the mass ingestion specification, configure the following incremental load options:
   - Mode. Select **Append** or **Overwrite**. Default is Append.
   - Propagate schema changes on the source. Select this option if you expect to update the columns in the source tables or add new columns to the source tables.

Configuring an HDFS Target

If the mass ingestion specification ingests data to an HDFS target, configure the HDFS target to specify how you want to ingest data to flat files on HDFS.

1. Select the target connection.
2. Optionally, enter the prefix and suffix to add to the target files.
3. Enter the ingestion directory.
4. Select the compression format. If you select to use custom compression, enter the compression codec.
5. Select the delimiter. If you select **Other**, enter a custom delimiter.
6. If you enabled incremental load in the definition of the mass ingestion specification, configure the following incremental load option:
   - Mode. Select **Append** or **Overwrite**. Default is Append.

Configuring Mass Transformations

Optionally, configure mass transformation to apply parameters across all of the source tables that you ingest.

1. For **Filter By**, enter a filter clause.
2. For **Drop Columns**, list the columns to drop or enter a regular expression.
3. For **Trim**, list the columns to trim or enter a regular expression.
4. For **Convert to Uppercase**, list the columns to convert to uppercase or enter a regular expression.
5. For **Convert to Lowercase**, list the columns to convert to lowercase or enter a regular expression.
6. For **Replace Columns**, list the columns to replace or enter a regular expression.
   a. For **Replace Criteria**, select **Entire String** or **Pattern**. Default is Entire String.
   b. If you select Entire String, enter the new string.
   c. If you select Pattern, enter the previous string or pattern. Then, enter the new string.

Configuring the Transformation Override

Optionally, configure the transformation override to override the parameters that are applied to specific source tables.

Note: For some parameters, you can click the arrow next to the text box to select columns from a list of columns that exist in the table.

1. If you enabled incremental load in the definition of the mass ingestion specification, edit the incremental key column for a source table in the **Incremental Column** column.
2. In the **Filter By** column, edit the filter clause.
3. In the **Drop Columns** column, edit the list of columns or the regular expression.
4. In the **Trim** column, edit the list of columns or the regular expression.
5. In the **Convert to Uppercase** column, edit the list of columns or the regular expression.
6. In the **Convert to Lowercase** column, edit the list of columns or the regular expression.

7. In the **Replace Columns** column, click the pencil.
    a. For **Replace**, edit the list of columns or the regular expression.
    b. For **Replace Criteria**, select **Entire String** or **Pattern**.
    c. If you select Entire String, edit the new string.
    d. If you select Pattern, edit the previous string or pattern, or edit the new string.
8. If the target is a Hive target, click the pencil in the **Hive** column and choose one of the following options:
    - Select **Hive Options** and enter the partition key, the cluster key, and the number of buckets.
    - Select **DDL Query** and edit the custom DDL query.

CHAPTER 4

Deploy

This chapter includes the following topics:

Deploy Overview

Deploy a mass ingestion specification to a Data Integration Service to prepare to run the specification and ingest data to the target.

Before you deploy a specification, you must specify a Data Integration Service and a Hadoop connection. The Data Integration Service uses the Hadoop connection to connect to the Hadoop environment when you run the deployed specification. After you deploy the mass ingestion specification, you can redeploy the specification using a different Hadoop connection or you can redeploy the specification to a different Data Integration Service.

When you deploy the mass ingestion specification, you isolate the specification from changes in the Mass Ingestion tool. For example, you deploy the mass ingestion specification to a Data Integration Service. After testing the ingestion output, you edit the mass ingestion specification. The changes that you make to the mass ingestion specification do not affect the deployed specification.

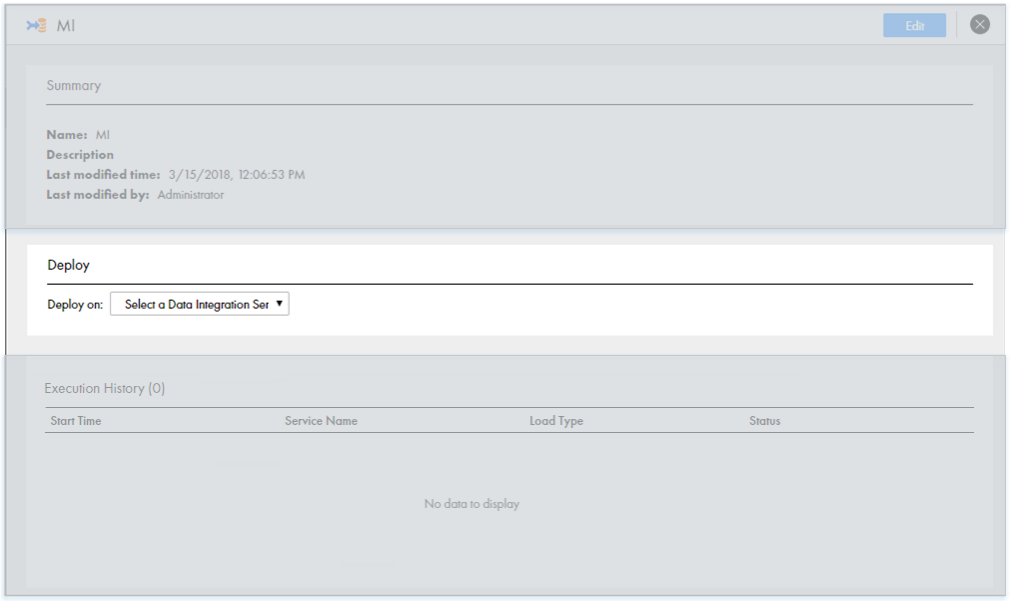
Deploying a Mass Ingestion Specification

Deploy a mass ingestion specification in the Mass Ingestion tool.

Note: If you make changes to the available Data Integration Services or Hadoop connections, refresh your browser or log out and log back in to the Mass Ingestion tool.

1. Browse to a mass ingestion specification in the Mass Ingestion tool.

   The **Overview** page for the mass ingestion specification appears.
   The following image shows the **Deploy** view:

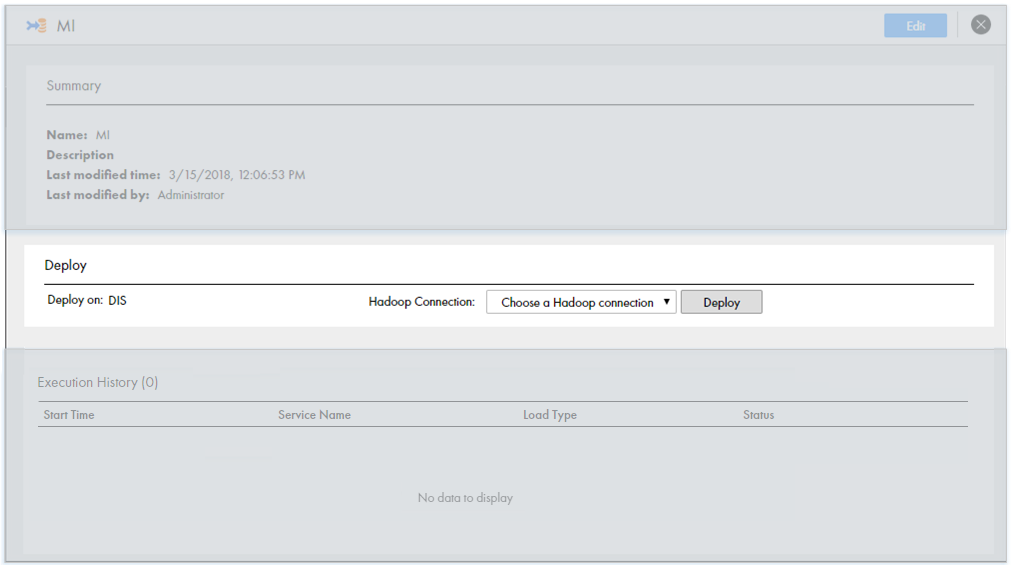


2. In the **Deploy** view, select a Data Integration Service.

   The option to select a Hadoop connection appears.

3. Select a Hadoop connection.

   The following image shows the option to select a Hadoop connection in the **Deploy** view:



4. Click **Deploy** to deploy the mass ingestion specification.

Redeploying a Mass Ingestion Specification

Redeploy a mass ingestion specification to a different Data Integration Service or a different Hadoop connection.

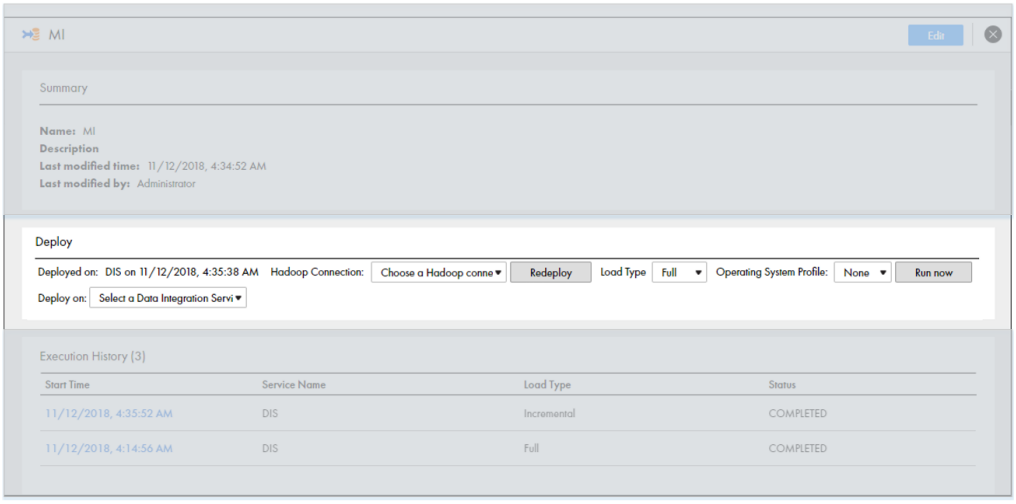
Note: If you make changes to the available Data Integration Services or Hadoop connections, refresh your browser or log out and log back in to the Mass Ingestion tool.

1. Browse to a mass ingestion specification in the Mass Ingestion tool.

   The **Overview** page for the mass ingestion specification appears.

2. To redeploy the mass ingestion specification to a different Hadoop connection, select a Data Integration Service in the **Deploy** view.

   The following image shows the **Deploy** view on the **Overview** page where you can redeploy the specification:



   Note: If you do not select a Data Integration Service, the mass ingestion specification is redeployed to the same Data Integration Service on a different Hadoop connection.

3. Select a Hadoop connection.
4. Click **Redeploy** to redeploy the mass ingestion specification.

Deploy to an Application Archive File

An application archive file contains the objects and metadata of an application in XML format. When you deploy a mass ingestion specification to an application archive file, you save all the information about the specification as an application in the XML file.

The file has an .iar extension. You can deploy a mass ingestion specification to an application archive file only through the command line.

You might want to create an application archive file for any of the following reasons:

- Deploy the application. If your organization restricts the ability to deploy applications to a Data Integration Service to administrators, an administrator can deploy an application from an archive file to a Data Integration Service. The administrator can use the Administrator tool or infacmd dis deployApplication.

- Import the application to a Model repository. You can use the Developer tool to import the application to a Model repository, or an administrator can import the application to a Model repository using infacmd tools importObjects.
- Archive the application archive file in another system. For example, if the Model repository is not integrated with a version control system, an administrator can check the archive file into a version control system.
- Back up the specification. You might want to back up the mass ingestion specification to prevent data loss due to hardware or software problems. If you need to recover the specification, you can import the application archive file to a Model repository.

For more information about infacmd, see the *Informatica Command Reference*.

For more information about application archive files, see the *Informatica Developer Tool Guide*.

Migrate a Mass Ingestion Specification

Migrate a mass ingestion specification to move the mass ingestion specification to a Model repository in a different environment. For example, you might want to migrate a mass ingestion specification between development, test, and production environments.

When you migrate a mass ingestion specification, the source and target connection names and IDs must be the same across the environments.

You use a different method to migrate a mass ingestion specification depending on the state of the source and target repositories. The source repository is the Model repository that contains the mass ingestion specifications that you want to migrate. The target repository is the Model repository where you want to migrate mass ingestion specifications.

Choose a method to migrate a mass ingestion specification depending on one of the following repository states:

- The target repository is empty and you want to migrate all Model repository contents. To migrate the contents, back up the contents in the source repository and restore the contents in the target repository.
- The target repository is empty and you want to migrate only the mass ingestion specifications. Before you migrate the mass ingestion specifications, create a Mass Ingestion Service and associate the Mass Ingestion Service with the target repository. The Mass Ingestion Service creates a project `INFA_MI_PROJECT` in the target repository where you can import the mass ingestion specifications. To migrate a mass ingestion specification, export the mass ingestion specification from the project `INFA_MI_PROJECT` in the source repository using infacmd tools exportObjects. Import the mass ingestion specification to the project `INFA_MI_PROJECT` in the target repository using infacmd tools importObjects.
- The target repository contains mass ingestion contents and you want to update the mass ingestion specifications in the target repository. To migrate a mass ingestion specification, export the mass ingestion specification from the project `INFA_MI_PROJECT` in the source repository using infacmd tools exportObjects. Import the mass ingestion specification to the project `INFA_MI_PROJECT` in the target repository using infacmd tools importObjects with the option `[<-ConflictResolution|-cr>]` to replace or rename the specifications in the target repository.

After you migrate a mass ingestion specification to a different environment, deploy the mass ingestion specification to a Data Integration Service.

For more information on infacmd tools, see the "infacmd tools" chapter in the *Informatica Command Reference*.

For more information on exporting and importing Model repository contents, see the "Object Import and Export" chapter in the *Informatica Developer Tool Guide*.

CHAPTER 5

Run

This chapter includes the following topics:

Run Overview

After you deploy a mass ingestion specification, you can run the specification to ingest the data.

When you run the specification, you can select one of the following load types:

- Full load. A full load ingests all of the data to the target. When you use a full load, the Spark engine deletes the existing data in the Hive or HDFS target and replaces the data with the data that is configured in the specification.
- Incremental load. An incremental load ingests only the incremental data to the target. The Spark engine appends the incremental data to the target or overwrites the target with the incremental data depending on the mode that you configure in the specification.

Each time that you run the specification, the Mass Ingestion Service generates a new run instance for the specification. The Data Integration Service processes the instance and connects to the Hadoop environment. In the Hadoop environment, the Spark engine runs the individual ingestion mappings jobs that ingest the data in the source tables to the target.

If the ingestion job stops responding or takes an excessive amount of time to complete, you can cancel the job in the Monitor tool.

For information on canceling ingestion jobs, see “Canceling Ingestion Jobs” on page 71 in the "Monitor" chapter.

Load Types

Run a mass ingestion specification using a full load or an incremental load. You can decide whether you want to use a full load or an incremental load depending on the state of the data in the relational database and in the target.

Full Load

Use a full load to ingest all of the data in the mass ingestion specification to the target. When you use a full load, the existing data in the Hive or HDFS target is deleted and replaced with the data in the source tables.

You might want to run a full load for any of the following reasons:

As a prerequisite for running incremental loads.

When you create a mass ingestion specification, run an initial full load before you begin running incremental loads on the data. The initial full load allows the Spark engine to create a basis to fetch incremental data in subsequent runs.

An initial full load can also help administrators maintain self-documented records. For example, it is possible to run an incremental load using overwrite mode as the first run of the specification, but the Spark engine does not have a basis to fetch incremental data. As a result, the Spark engine ingests all of the data from the source and effectively runs a full load. The records would indicate that a user ran an incremental load, but it might be unclear whether all data or only incremental data was ingested to the target.

If you run a initial full load followed by subsequent incremental loads, the administrator can distinguish whether the Spark engine ingested all data or only incremental data for each run of the specification.

To update the basis for incremental loads.

Run a full load to update the target based on UPSERT and DELETE statements that have been run against the relational database.

If you run an incremental load, the Spark engine fetches the rows that have been added to a relational table using INSERT statements. The Spark engine cannot fetch the rows that have been changed by UPSERT and DELETE statements, so an incremental load from the relational database might not provide an accurate representation of the source data.

Incremental Load

Use an incremental load to ingest only incremental data to the target. Before you can use an incremental load, you must enable incremental load and configure incremental load options in the mass ingestion specification.

You might want to use an incremental load since a full load can be time- and resource-intensive, especially if the data is largely unchanged between runs of a specification. For a more cost-efficient solution, you can use an incremental load to ingest only the incremental data each time that the specification runs.

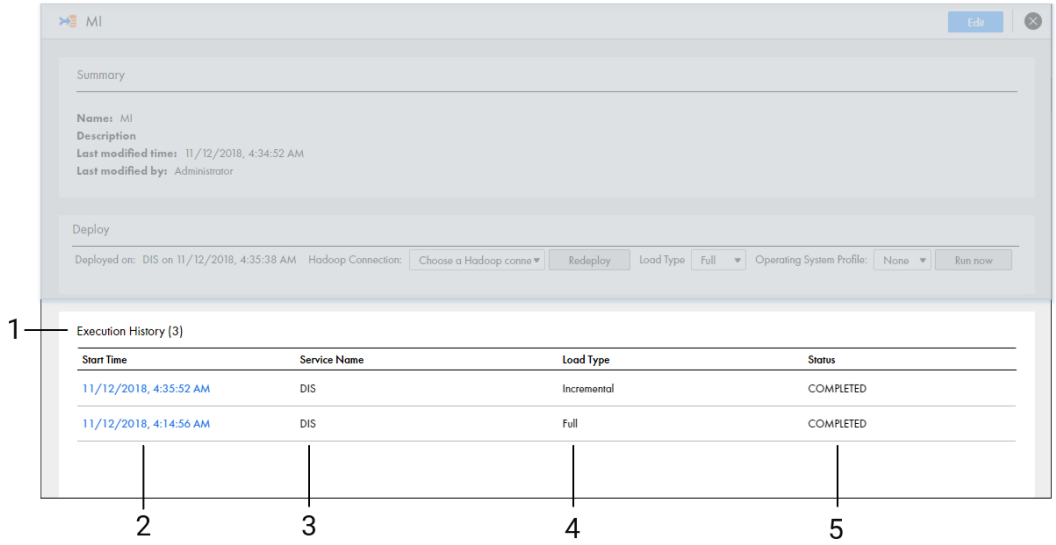
Before you begin running incremental loads on a specification, run an initial full load. An initial full load allows the Spark engine to internally persist the original source data and use the persisted values as a basis to fetch incremental data in an incremental load. After the Spark engine persists the original source data, you can ingest incremental data based on the original source data to any target.

Note: If you run an incremental load using append mode, the target must contain existing tables or files so that the Spark engine can append incremental data to the existing data. To create the tables or files, run a full load.

Viewing Run Instances

Each time that you run a mass ingestion specification, the Mass Ingestion Service generates a new run instance of the specification. You can view the run instances in the Execution History view on the Overview page in the Mass Ingestion tool.

The following image shows run instances in the Execution History view:



1. Execution History view. Displays the run instances of the mass ingestion specification.
2. Start Time. Displays the start time for the run instance.
3. Service Name. Displays the Data Integration Service where the mass ingestion specification is deployed.
4. Load Type. Displays the type of load that the Spark engine used to ingest the data.
5. Status. Displays the status of the run instance.

You can also list the run instances for a mass ingestion specification from the command line.

For more information, see "listSpecRuns" on page 82 in the "infacmd mi Command Reference" appendix.

Run Instance Status

The run instance status might display Completed, Failed, Running, Aborted, Canceled, Queued, or Unknown.

The following table describes the status that might be displayed for each run instance:

Status	Description
Completed	The run instance of the mass ingestion specification is completed. All data is ingested to the target successfully.
Failed	The run instance of the mass ingestion specification failed. Some data failed to be ingested to the target.
Running	The run instance of the mass ingestion specification is running. Some data is being ingested to the target.
Aborted	The run instance of the mass ingestion specification is aborted. The run instance was aborted from the command line.

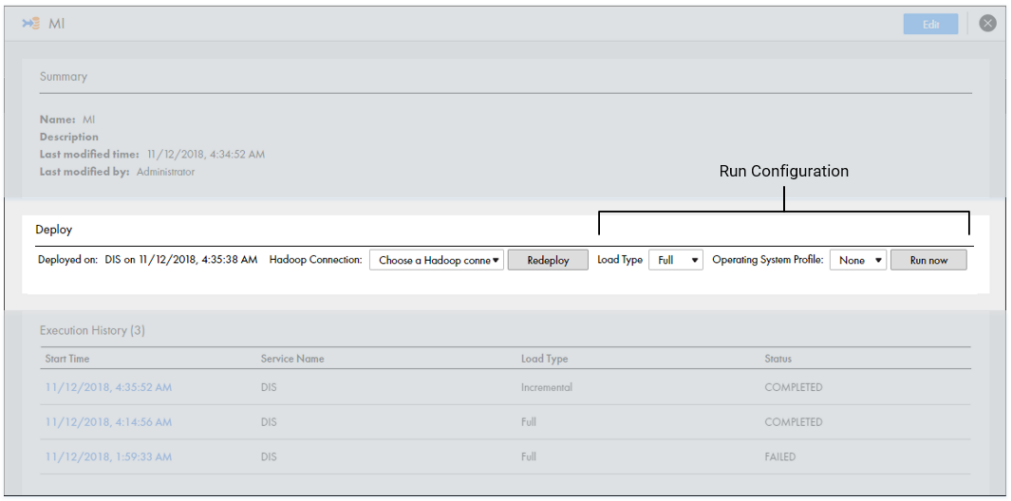
Status	Description
Canceled	The run instance of the mass ingestion specification is canceled. The jobs to ingest some source tables in the mass ingestion specification are canceled on the Data Integration Service.
Queued	The run instance of the mass ingestion specification is queued. The jobs to ingest some source tables in the mass ingestion specification are queued on the Data Integration Service.
Unknown	The status of the mass ingestion specification run instance is unknown. The Mass Ingestion Service cannot fetch the status of the run instance. The status for the run instance might display `Unknown` in the following circumstances: - A Model repository is not configured in the monitoring settings for the Mass Ingestion Service. - The Data Integration Service is disabled.

Running a Mass Ingestion Specification

Run a mass ingestion specification to ingest the data from the source tables to the Hive or HDFS target.

1. In the Mass Ingestion tool, browse to a mass ingestion specification.

   The **Overview** page appears.
   The following image shows the run configuration in the **Deploy** view on the **Overview** page:



2. In the **Deploy** view, select the load type. You can select **Full** or **Incremental**. Default is **Full**.

   Note: If incremental load is not enabled in the specification, you cannot use an incremental load. To use an incremental load, edit the specification. In the specification, enable incremental load and configure the incremental load options.
3. Select an operating system profile. Default is None.
4. Click **Run** to run the mass ingestion specification.

   A run instance for the mass ingestion specification appears in the **Execution History** view.

CHAPTER 6

Monitor

This chapter includes the following topics:

Monitor Overview

When you run the mass ingestion specification, you can begin monitoring the job statistics. Monitor the job statistics in the Mass Ingestion tool or in the Administrator tool.

Monitoring in the Mass Ingestion Tool

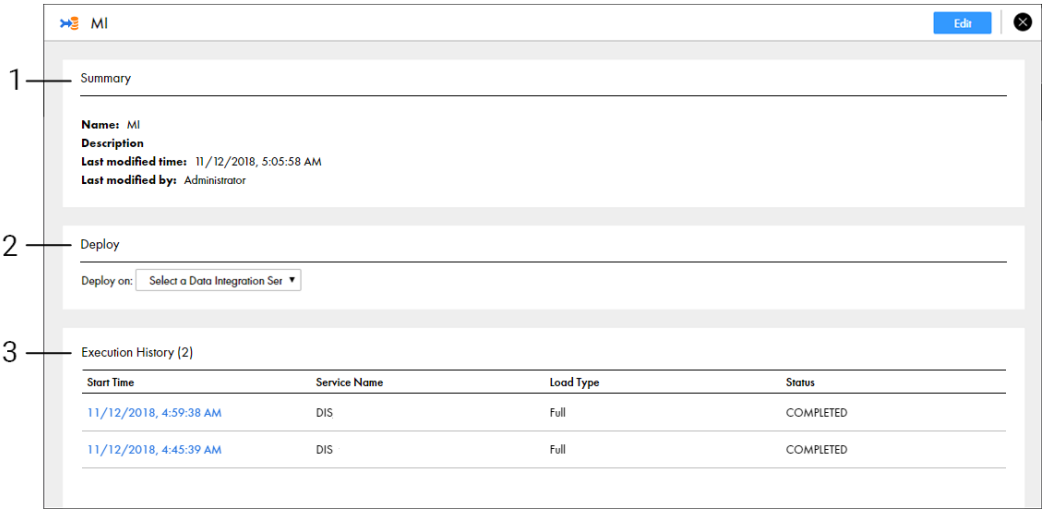
You can monitor a mass ingestion specification in the Mass Ingestion tool. Monitor the mass ingestion specification details and specification history on the Overview page, and monitor ingestion job statistics on the Execution Statistics page.

The Overview page provides general details on the specification, the deployment history, and the execution history. The Execution Statistics page provides detailed statistics on the ingestion job for a specification run instance.

Overview Page

On the Overview page, review details on a mass ingestion specification, the deployment and run options, and the execution history.

The following image shows the Overview page:



1. Summary. Displays general mass ingestion specification properties.
2. Deploy. Displays options to deploy the mass ingestion specification.
3. Execution History. Displays the run instances of the mass ingestion specification.

Summary View

The Summary view displays general mass ingestion specification properties.

The following table describes the specification properties that you can review:

Property	Description
Name	The name of the mass ingestion specification.
Description	The description of the mass ingestion specification.
Last modified time	The time that the mass ingestion specification was last modified.
Last modified by	The last user who modified the mass ingestion specification.

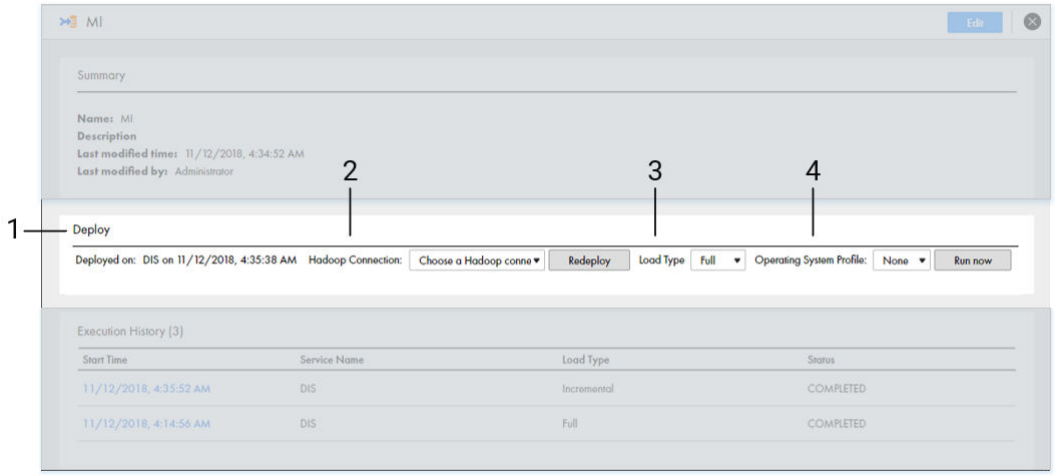
Deploy View

The Deploy view displays options to deploy and run the mass ingestion specification.

To deploy the specification, you can specify a Data Integration Service and a Hadoop connection. After the mass ingestion specification is deployed, you can redeploy the specification using a different Hadoop connection or to a different Data Integration Service.

The Data Integration Service uses the Hadoop connection to connect to the Hadoop environment. When the mass ingestion specification is deployed, you can run the specification from the Deploy view using options to select a load type and an operating system profile.

The following image shows the Deploy view after the mass ingestion specification has been deployed:



1. Deployed on. Displays the name of the Data Integration Service where the specification was deployed and the time that the specification was deployed.
2. Hadoop Connection. Option to select a Hadoop connection to redeploy the specification.
3. Load Type. Option to select a load type to define how the Spark engine ingests the source data when you run the specification. Default is Full.
4. Operating System Profile. Option to select an operating system profile when you run the specification. Default is None.

Execution History View

The Execution History view displays the run instances of the mass ingestion specification and information about each run instance.

The following table describes the run instance information that you can view in the Execution History view:

Property	Description
Start Time	The time that the run instance is initiated.
Service Name	The name of the Data Integration Service where the mass ingestion specification is deployed.
Load Type	The type of load that the Spark engine used to ingest the data in the mass ingestion specification. The load type might be Full or Incremental.
Status	The run instance status. The status might display Completed, Failed, Running, Aborted, Canceled, Queued, or Unknown.

The following table describes the status that might be displayed for each run instance:

Status	Description
Completed	The run instance of the mass ingestion specification is completed. All data is ingested to the target successfully.
Failed	The run instance of the mass ingestion specification failed. Some data failed to be ingested to the target.

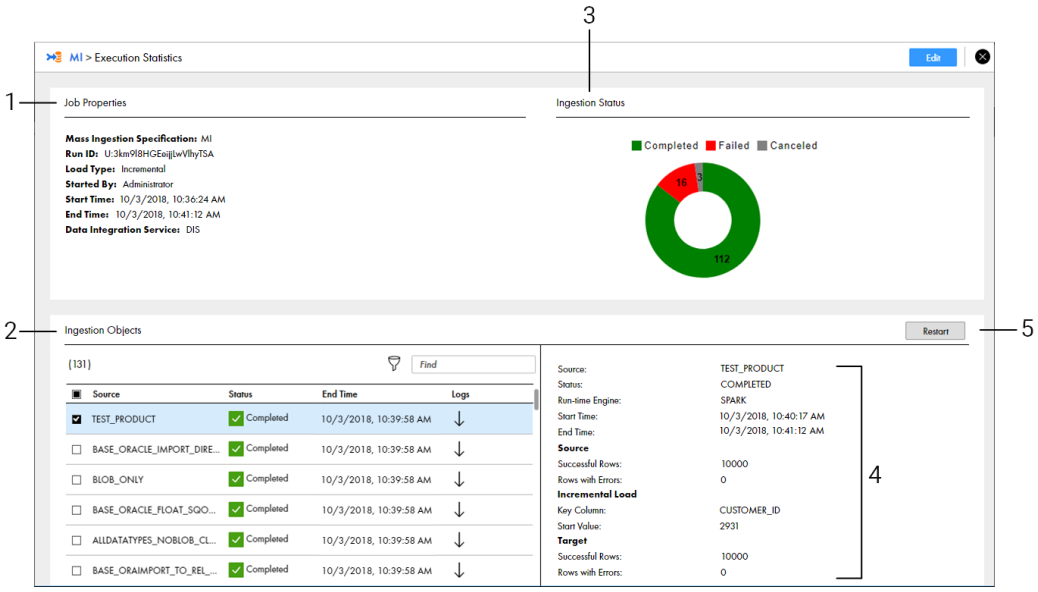
Status	Description
Running	The run instance of the mass ingestion specification is running. Some data is being ingested to the target.
Aborted	The run instance of the mass ingestion specification is aborted. The run instance was aborted from the command line.
Canceled	The run instance of the mass ingestion specification is canceled. The jobs to ingest some source tables in the mass ingestion specification are canceled on the Data Integration Service.
Queued	The run instance of the mass ingestion specification is queued. The jobs to ingest some source tables in the mass ingestion specification are queued on the Data Integration Service.
Unknown	The status of the mass ingestion specification run instance is unknown. The Mass Ingestion Service cannot fetch the status of the run instance. The status for the run instance might display `Unknown` in the following circumstances: - A Model repository is not configured in the monitoring settings for the Mass Ingestion Service. - The Data Integration Service is disabled.

When you click the start time for a run instance in the Execution History view, you can see the Execution Statistics page for the specification run. On the Execution Statistics page, you can monitor the ingestion job statistics.

Execution Statistics Page

On the Execution Statistics page, view detailed information on the ingestion job for a run instance of the mass ingestion specification.

The following image shows the Execution Statistics page:



1. Job Properties. Displays general properties for the ingestion job.
2. Ingestion Objects. Lists the source tables that the job ingests to the target.
3. Ingestion Status. Displays a graphical representation of the ingestion job.
4. Ingestion Statistics. Lists the ingestion job statistics for a specific source table.
5. Restart. Restarts the ingestion job for ingestion objects.

Job Properties View

The Job Properties view lists general properties to describe the run instance of the mass ingestion specification.

The following table describes the properties that you can view in the Job Properties view:

Property	Description
Mass Ingestion Specification	The name of the mass ingestion specification.
Run ID	The run identifier number, or the Run ID, of the mass ingestion specification run instance.
Load Type	The type of load that the Spark engine used to ingest the data. The load type might be Full or Incremental.
Started By	The user who initiated the mass ingestion specification run.
Start Time	The time that the ingestion job is initiated.

Property	Description
End Time	The time that the ingestion job is completed.
Data Integration Service	The name of the Data Integration Service where the mass ingestion specification is deployed.

Ingestion Objects View

The Ingestion Objects view lists the source tables that are part of the ingestion job.

Use the Ingestion Objects view to monitor the tables that are ingested successfully and the tables that fail. If certain source tables fail to be ingested, restart the ingestion process for the failed tables.

The following table describes the properties that you can view in the Ingestion Objects view:

Property	Description
Source	The name of the source table that is configured in the mass ingestion specification.
Status	The ingestion status of the source table. The status might display Completed, Failed, Running, Canceled, Queued, Unknown, or Working.
End Time	The time that the ingestion job for the source table is completed.
Logs	The option to download the ingestion log for the source table.

The following table describes the status that might be displayed for each source table:

Status	Description
Completed	The ingestion job for the source table is completed. All data in the source table was ingested to the target successfully.
Failed	The ingestion job for the source table failed. The data in the source table failed to be ingested to the target.
Running	The ingestion job for the source table is running. Some data is being ingested to the target.
Aborted	The ingestion job for the source table is aborted. The run instance of the mass ingestion specification that contains the ingestion job was aborted from the command line.
Canceled	The ingestion job for the source table is canceled. The deployed mapping that performs the ingestion job is canceled on the Data Integration Service.
Queued	The ingestion job for the source table is queued. The Mass Ingestion Service is waiting to schedule the deployed mapping that performs the ingestion job to run.

Status	Description
Unknown	The status of the ingestion job for the source table is unknown. The Mass Ingestion Service cannot fetch the status of the source table. The status for the run instance might display `Unknown` in the following circumstances: - A Model repository is not configured in the monitoring settings for the Mass Ingestion Service. - The Data Integration Service is disabled.
Working	The Mass Ingestion Service is fetching the status of the source table from the Data Integration Service.

Ingestion Status View

The Ingestion Status view displays a graphical representation of the mass ingestion job.

The graphic summarizes the ingestion status listed in the Ingestion Objects view. The status for each source table might be Completed, Failed, Running, Canceled, Queued, Unknown, or Working.

Based on the ingestion status, you might decide to modify the mass ingestion specification. If you modify the specification and run the specification again, review the statistics on a unique Execution Statistics page for the generated run instance.

Ingestion Statistics View

The Ingestion Statistics view lists the ingestion job statistics for a specific source table.

The following table describes the properties that you can view in the Ingestion Statistics view:

Property	Description
Source	The name of the source table that is ingested.
Status	The ingestion status of the source table. The status might display Completed, Failed, Running, Canceled, Queued, Unknown, or Working.
Run-time Engine	The run-time engine in the Hadoop environment that performs the ingestion job. The run-time engine is Spark.
Start Time	The time that the ingestion job for the source table is initiated.
End Time	The time that the ingestion job for the source table is completed.
Source	The statistics for the table rows that are from the source database. You can view the number of rows that were ingested successfully and the number of rows that contain errors.
Incremental Load	Details that describe the incremental data that was ingested. The details appear if the Spark engine used an incremental load to ingest the data. You can view the name of the column that the Spark engine used to fetch incremental data in the source table and the start value that determines the value that the Spark engine used to start ingesting the incremental data. For example, if the incremental key is an ID column and the start value is 2500, the Spark engine ingested the rows that have an ID of 2500 or higher.
Target	The statistics for the table rows that are ingested in the target. You can view the number of rows that were ingested successfully and the number of rows that contain errors.

Note: To view certain statistics, you must enable Spark monitoring. Spark monitoring might impact performance.

Monitoring Execution Statistics

Monitor the execution statistics for a mass ingestion specification run instance on the Execution Statistics page.

1. Browse to a mass ingestion specification in the Mass Ingestion tool.

   The **Overview** page appears.

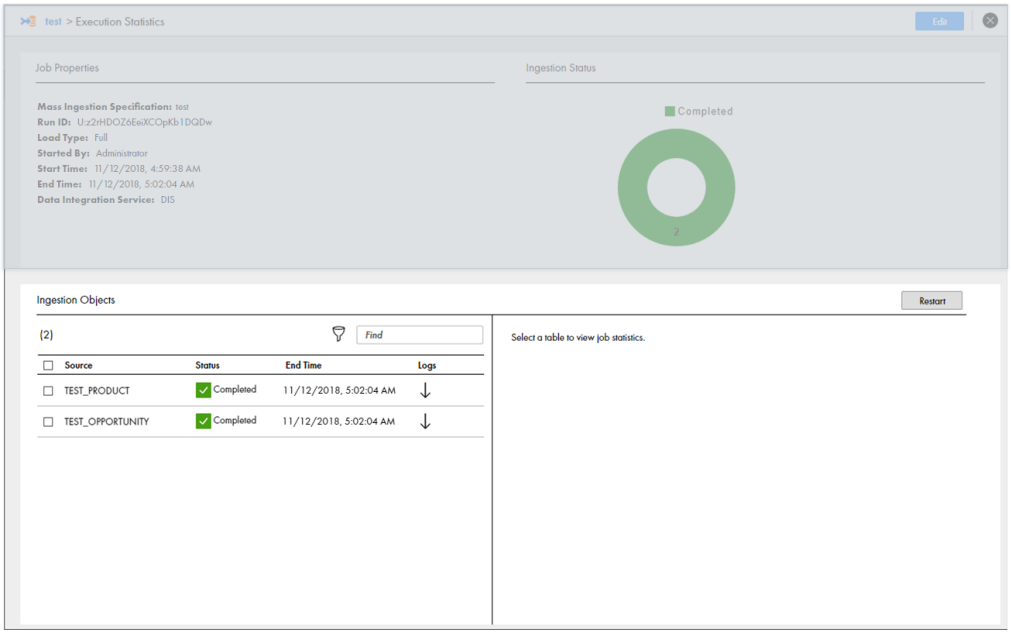
2. Select a run instance in the Execution History view to view the ingestion job statistics.

   The **Execution Statistics** page appears for the run instance that you select.

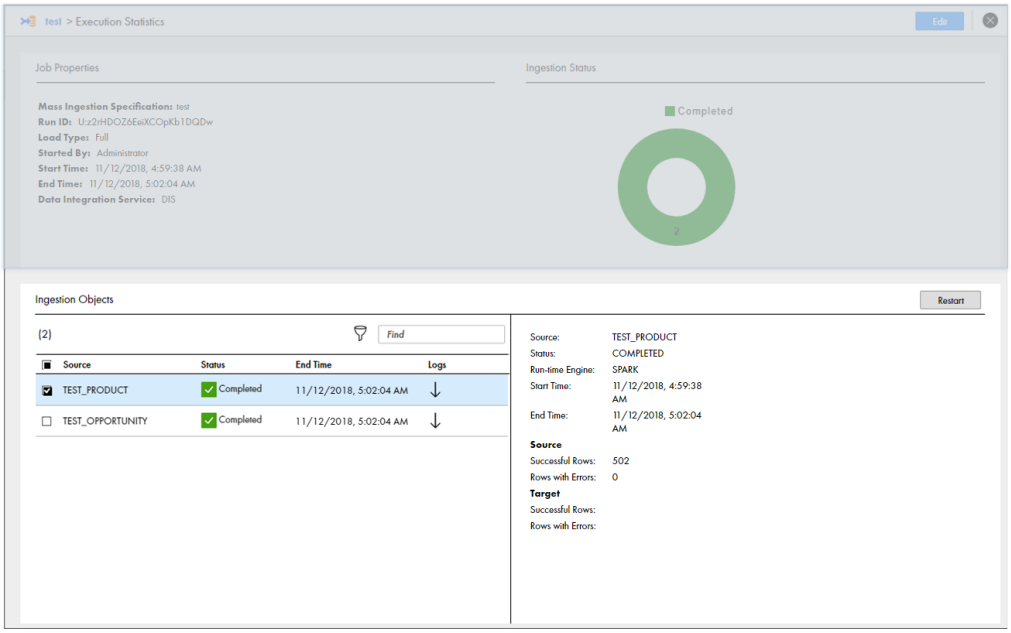
3. On the **Execution Statistics** page, select a source table in the Ingestion Objects view to view statistics for the job that ingests the source table.

   The statistics appear in the Ingestion Statistics view. You can view ingestion statistics for one table at a time.

   For example, the following image shows the different source tables that might appear in the Ingestion Objects view:



   The following image shows the statistics that might appear in the Ingestion Statistics view for the first source table in the Ingestion Objects view:

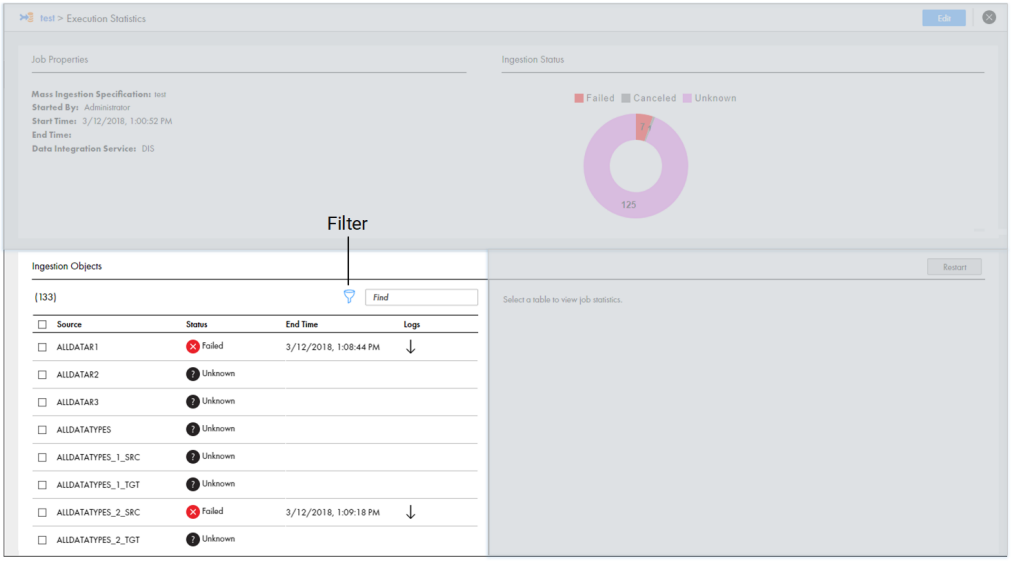


Filtering Ingestion Objects by Status

Filter ingestion objects on the Execution Statistics page to find the specific tables that you want to monitor. You can filter the ingestion objects by the ingestion status.

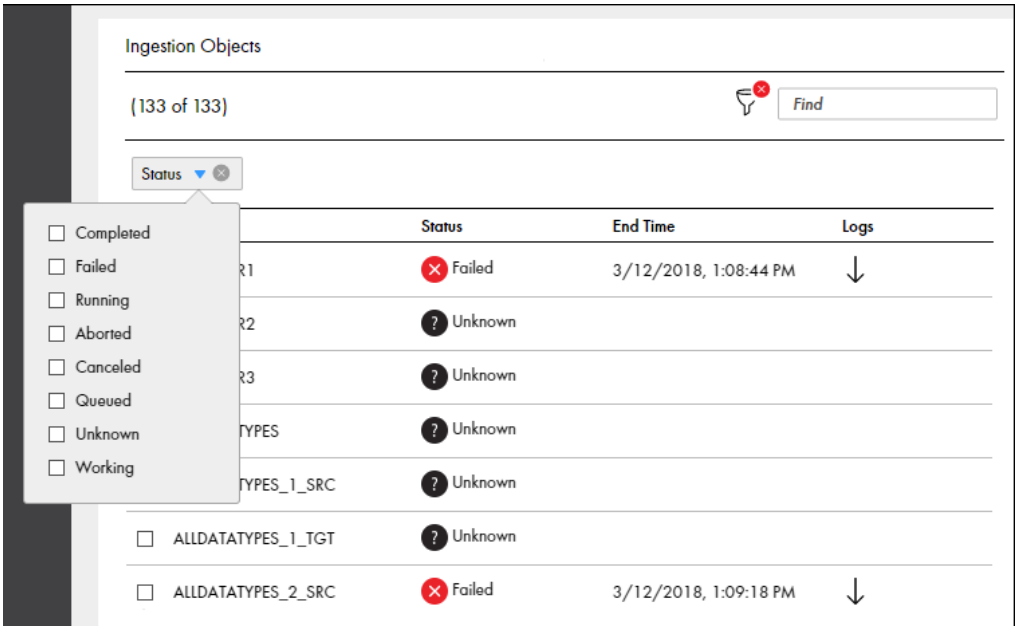
1. In the Ingestion Objects view on the Execution Statistics page, select the **Filter** icon.

   The following image shows the **Filter** icon in the Ingestion Objects view:



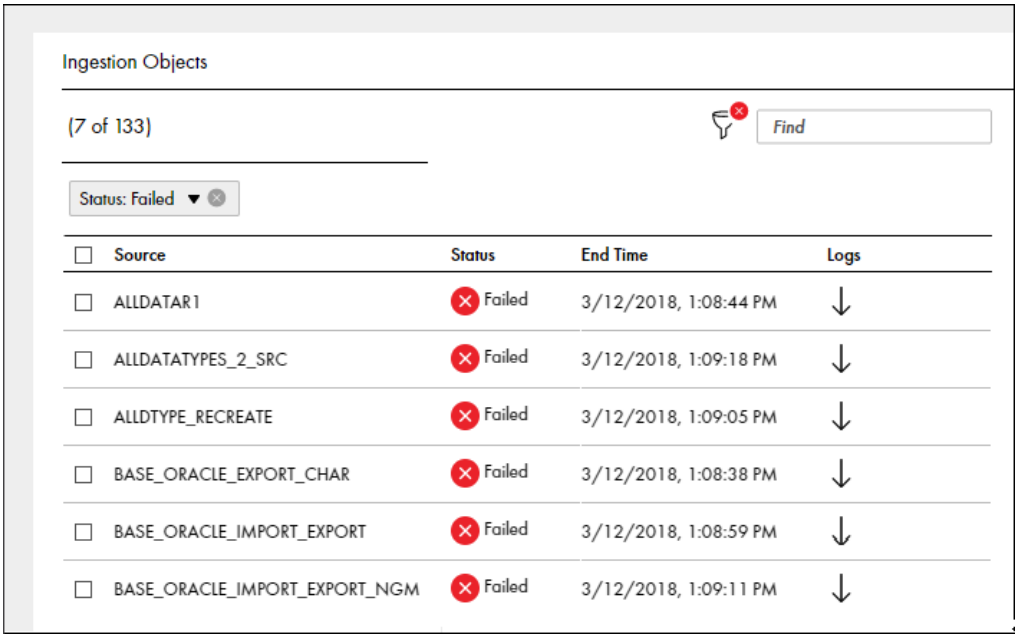
2. Select **Status** to filter the ingestion objects by the ingestion status.

   The following image shows the **Status** option in the Ingestion Objects view:



3. Select the status that you want to use to filter the ingestion objects.

   For example, you can select **Failed**. The following image shows the ingested objects filtered by Failed in the Ingestion Objects view:



Restart Ingestion Jobs

After the mass ingestion specification runs, you can restart the ingestion jobs for specific source tables.

For example, you might want to restart the ingestion job for a source table that failed to be ingested. Source tables might fail to be ingested due to a Hadoop configuration that is not compatible with the ingestion job, or they might fail if the ingestion target does not have enough space to consume the ingested tables. You might verify the Hadoop configuration or reconfigure the ingestion target. Then, you can restart the ingestion jobs for the tables that failed to be ingested.

To restart the ingestion jobs, perform one of the following tasks:

- Restart Selected Objects. Restart the ingestion job only for the source tables that you select.
- Restart Failed Objects. Restart the ingestion jobs for all source tables that failed to be ingested and all of the ingestion jobs that are canceled.

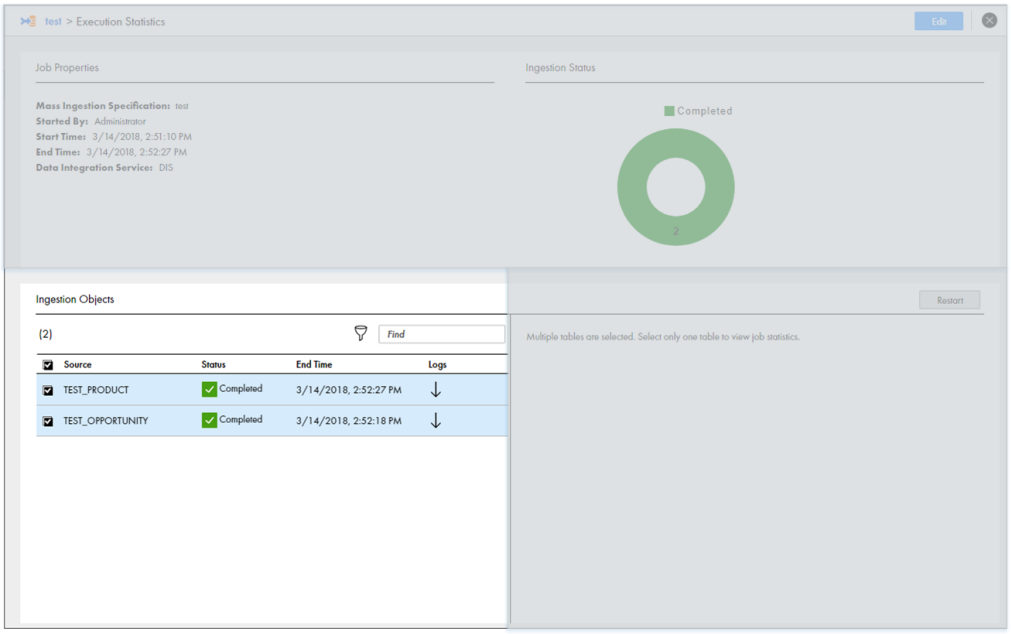
When you restart the ingestion jobs, you do not redeploy or rerun the entire mass ingestion specification. The ingestion status for the job refreshes and the status displays Working. The status continues to display Working until the Mass Ingestion Service fetches the ingestion status from the Data Integration Service.

Restarting Selected Ingestion Objects

Restart selected ingestion objects in the mass ingestion specification to restart the job to ingest the objects.

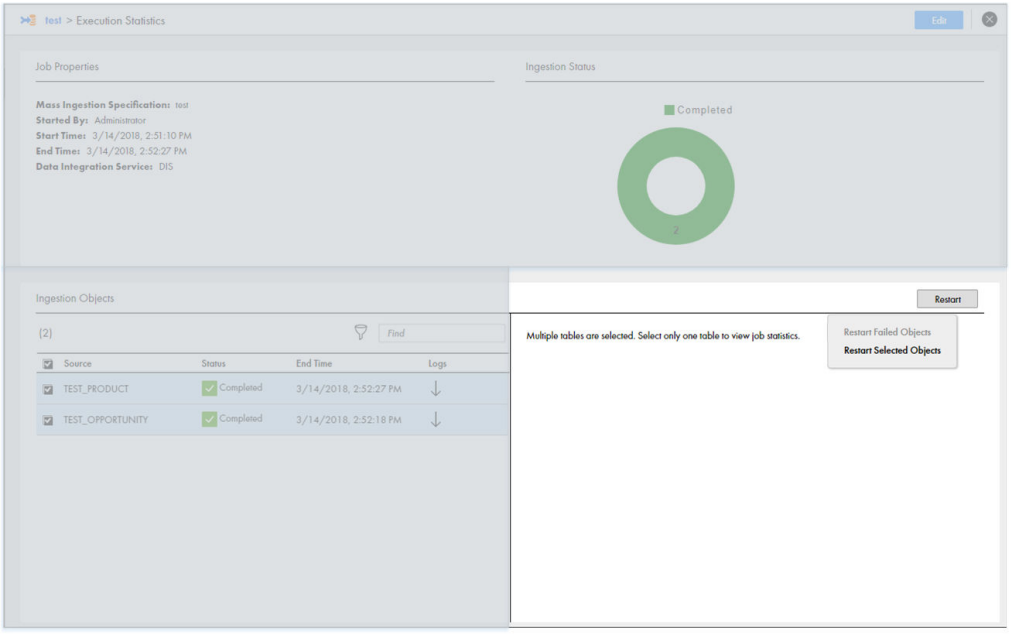
1. In the Ingestion Objects view on the Execution Statistics page, select the ingestion objects that you want to restart.

   The following image shows several objects that are selected:



2. In the Ingestion Statistics view, select **Restart**.

   The following image shows the **Restart** option:



3. Select **Restart Selected Objects**.

   The ingestion job restarts for the selected objects.

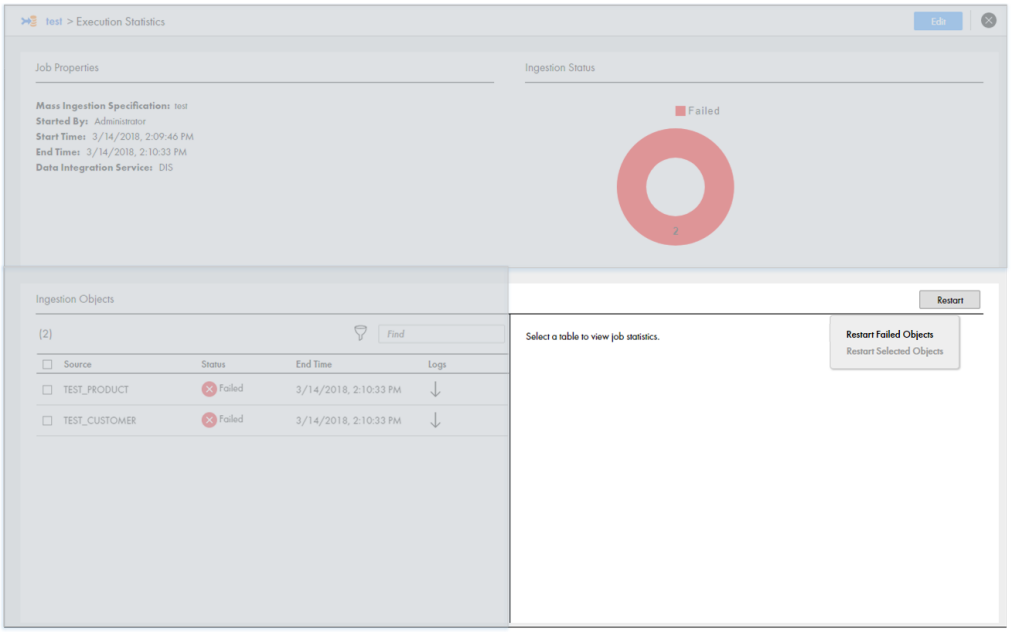
Restarting Failed Ingestion Objects

You can choose to restart the ingestion objects to restart the job to ingest only the source tables that failed or were canceled.

Note: The option to restart failed ingestion objects does not restart the jobs that are aborted.

1. In the Ingestion Statistics view on the Execution Statistics page, select **Restart**.

   The following image shows the **Restart** option:



2. Select **Restart Failed Objects**.

   The ingestion job restarts for all failed and canceled objects.

Monitoring in the Administrator Tool

You can monitor a mass ingestion specification in the Administrator tool. Monitor the application and the ingestion mapping jobs that ingest the source tables that you configure in a mass ingestion specification.

Monitor the Application and Ingestion Mapping Jobs

Monitor the application and ingestion mapping jobs on the Monitor tab. You can view statistics on the Summary Statistics tab or the Execution Statistics tab.

Summary Statistics

On the Summary Statistics tab, you can view summaries of multiple applications and ingestion mapping jobs. The summaries provide details on object distribution, object state, and Data Integration Service resource usage for a specified time range.

You can choose one of the following options to view a summary:

- Graphical distribution. Displays visual data on the application and mapping job distributions.
- Tabular distribution. Displays the total completed, running, canceled, aborted, and failed jobs.
- Details. Displays a list of the application or mapping jobs, requests, or connections that are involved
- Export data. Exports the detail data for a specific application or mapping job to a .csv file.

For more information, see the *Informatica Administrator Guide*.

Execution Statistics

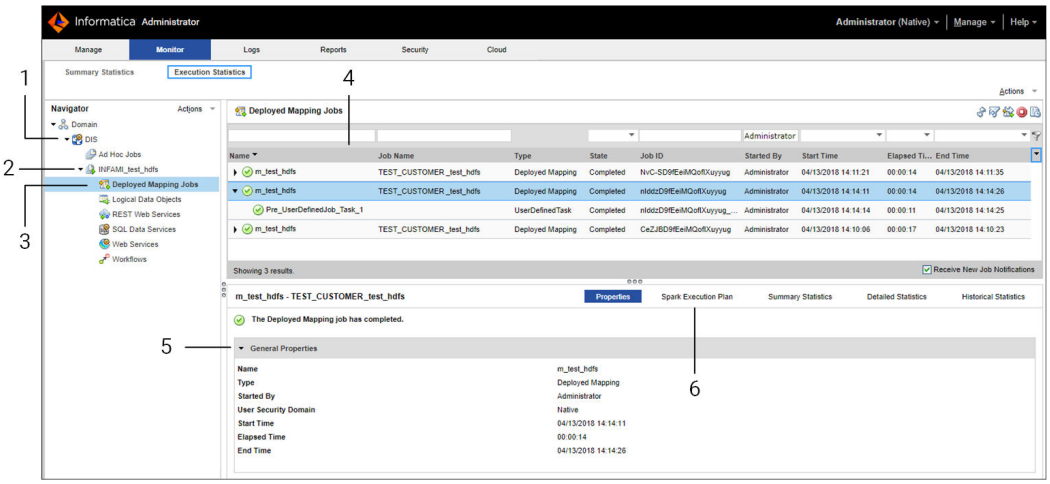
On the Execution Statistics tab, you can monitor the application and the ingestion mapping jobs that are deployed to a Data Integration Service.

You can select an application in the Navigator. The contents panel shows the following views:

- Properties view. Shows the general properties and run-time statistics. Use the properties view to view additional information about the application and mapping jobs in the application.
- Reports view. Shows monitoring reports. Use the Reports view to monitor reports about the application and mapping jobs.

You can browse to the deployed mapping jobs in the application to view information about the ingestion mapping jobs that ingest the source tables in a mass ingestion specification.

The following image shows the Execution Statistics tab for a deployed mapping job:



1. Data Integration Service. The Data Integration Service where the mass ingestion specification is deployed.
2. Application. The application that contains the mapping jobs in a deployed mass ingestion specification.
3. Deployed mapping jobs. The mappings that perform the ingestion jobs configured in the mass ingestion specification.
4. Contents panel. The properties that describe the mapping jobs.
5. General properties. The properties that summarize a mapping job.
6. Execution plan. The Spark execution plan that summarizes the job statistics on Hadoop.

The following table describes the information that you can view in the contents panel for a deployed mapping job:

Property	Description
Name	The name of the mass ingestion specification associated with the mapping job. The name appears in the following format: `m_<mass ingestion specification name>`
Job Name	The name of the source table that the mapping job ingests. The name appears in the following format: `<source table name>_<mass ingestion specification name>` Sort the mapping jobs by job name to quickly find the mapping that ingests a source table.
Type	The job type that defines a task in the mapping.
State	The status of the mapping job.
Job ID	The Job ID that identifies the mapping job.
Started By	The user who initiated the mass ingestion specification run instance.
Start Time	The time that the mapping job is initiated.
Elapsed Time	The time that the mapping job requires to complete.
End Time	The time that the mapping job is completed.

For more information about execution statistics, see the *Informatica Administrator Guide*.

For information about Spark execution plans, see the *Informatica Data Engineering Administrator Guide*.

Canceling Ingestion Jobs

You can cancel a running ingestion mapping job. You might want to cancel a job that stops responding or that is taking an excessive amount of time to complete.

When you cancel an ingestion mapping job, you cancel the mapping that performs the ingestion job for a source table in a run instance of a mass ingestion specification. If you monitor the mass ingestion specification from the Mass Ingestion tool, the source table status displays `Canceled`. After a job is canceled, none of the data for the corresponding source table is loaded to the target.

After you cancel an ingestion mapping job, you can reissue the mapping to restart the ingestion mapping job for the source table. You can also restart the ingestion mapping job from the Mass Ingestion tool.

1. Click the **Execution Statistics** view.
2. In the Domain Navigator, expand a Data Integration Service.
3. Expand the application that contains the information for a deployed mass ingestion specification and select **Deployed Mapping Jobs**.

   A list of jobs appears in the contents panel.
4. In the contents panel, select the ingestion mapping job.
5. From the **Actions** menu, select **Cancel Selected Object**.
6. To restart the ingestion mapping job, select **Reissue Selected Object**.

Aborting Ingestion Jobs

If a large number of jobs in the mass ingestion specification stop responding, you can abort all of the jobs from the command line.

After you abort the jobs in a mass ingestion specification, the status of the jobs that were running changes to `Aborted`. None of the data in the aborted jobs is loaded to the target.

For more information, see "abortRun" on page 73 in the "infacmd mi Command Reference" appendix.

Troubleshooting Mass Ingestion Jobs

Consider the following troubleshooting tips when you monitor mass ingestion jobs:

How do I view the full stack trace for a specific table that failed to be ingested?

To view the full stack trace, complete the following tasks:

1. In the Administrator tool, navigate to the deployed application on the Data Integration Service.
2. Stop the deployed application.
3. Within the deployed application, navigate to the ingestion job that corresponds to the table.
4. Edit the mapping properties to set the tracing level to **Verbose Data**.
5. Start the deployed application.
6. Run the ingestion job again.

   For example, you can use the `infacmd ms runMapping` command.
7. View the mapping logs.

APPENDIX A

infacmd mi Command Reference

This appendix includes the following topics:

abortRun

Aborts the ingestion mapping jobs in a run instance of a mass ingestion specification. When you abort the ingestion mapping jobs, the command aborts the mappings that perform the ingestion jobs for all source tables that are running or queued. The command does not abort mappings for the ingestion jobs that are completed.

To abort the ingestion mapping jobs, you must specify a RunID. To find the RunID for a run instance, list the specification run instances using infacmd mi listSpecRuns.

The infacmd mi abortRun command uses the following syntax:

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

The following table describes infacmd mi abortRun options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-runID -rid	run_id	Required. Run identifier number, or the Run ID, of the mass ingestion specification run instance. To find the RunID for a run instance, list the specification run instances using infacmd mi listSpecRuns.

createService

Creates a Mass Ingestion Service. When you create the Mass Ingestion Service, you must specify a Model Repository Service. The Mass Ingestion Service is disabled by default. To enable the Mass Ingestion Service, use infacmd isp enableService.

The infacmd mi createService command uses the following syntax:

```
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]
```

The following table describes infacmd mi createService options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-Gateway -hp	gateway_host1:port gateway_host2:port	Required if the gateway connectivity information in the domains.infa file is out of date. The host names and port numbers for the gateway nodes in the domain.

Option	Argument	Description
-ResilienceTimeout -re	timeout_period_in_seconds	Optional. Amount of time in seconds that infacmd attempts to establish or re-establish a connection to the domain. You can set the resilience timeout period with the -re option or the environment variable INFA_CLIENT_RESILIENCE_TIMEOUT. If you set the resilience timeout period with both methods, the -re option takes precedence.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service. The name is not case sensitive and must be unique within the domain. The characters must be compatible with the code page of the associated repository. The name cannot exceed 230 characters, have leading or trailing spaces, or contain carriage returns, tabs, or the following characters: / * ? < > " |
-HttpPort -http	http_port	Required if you do not specify an HTTPS port. Unique HTTP port number used for each Mass Ingestion Service process. After you create the service, you can define different port numbers for each Mass Ingestion Service process. Default is 9050. **Note:** You cannot specify both an HTTP port and an HTTPS port.
-HttpsPort -https	https_port	Required if you do not specify an HTTP port. Unique HTTPS port number used for each Mass Ingestion Service process. After you create the service, you can define different port numbers for each Mass Ingestion Service process. **Note:** You cannot specify both an HTTP port and an HTTPS port.
-KeystoreFile -kf	keystore_file_location	Required if you specify an HTTPS port. Path and file name of the keystore file that contains the keys and certificates required if you use the HTTPS protocol for the Mass Ingestion Service. You can create a keystore file with a keytool. keytool is a utility that generates and stores private or public key pairs and associated certificates in a keystore file. You can use the self-signed certificate or use a certificate signed by a certificate authority.
-KeystorePassword -kp	keystore_password	Required if you specify an HTTPS port. Password for the keystore file.
-LicenseName -ln	license_name	Required. Name of the license you want to assign to the Mass Ingestion Service. To apply changes, restart the Mass Ingestion Service.

Option	Argument	Description
-FolderPath -fp	full_folder_path	Optional. Full path, excluding the domain name, to the folder in which you want to create the Mass Ingestion Service. Must be in the following format: */parent_folder/child_folder* Default is the domain: */*
-NodeName -nn	node_name	Required. Node where the Mass Ingestion Service runs.
-RepositoryService -rs	repository_service_name	Required. Model Repository Service that stores the metadata for mass ingestion specifications.
-RepositoryUser -ru	repository_user	Optional. User name to access the Model Repository Service.
-RepositoryPassword -rp	repository_password	Required if you specify the user name. User password to access the Model Repository Service.
-RepositoryUserSecurityDomain -rsdn	repository_user_security_domain	Optional. Name of the security domain that the Model repository user belongs to.

deploySpec

Deploys a mass ingestion specification. When you deploy the specification, you must specify the Data Integration Service and the Hadoop connection. You must deploy a mass ingestion specification before you can run it. After you deploy the specification, run the specification using infacmd mi runSpec.

The infacmd mi deploySpec command uses the following syntax:

```
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
```

The following table describes infacmd mi deploySpec options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-DISServiceName -dis	data_integration_service	Required. Name of the Data Integration Service where you want to deploy the mass ingestion specification.
-MISpecName -spec	mi_spec_name	Required. Name of the mass ingestion specification that you want to deploy to the Data Integration Service.
-HadoopConnection -hc	hadoop_connection	Required. The Hadoop connection that the Data Integration Service uses to push the mass ingestion specification to the Hadoop environment.

exportSpec

Exports the mass ingestion specification to an application archive file. When you export the specification, you must specify the directory where you want to save the file. You can deploy the application archive file to a Data Integration Service using infacmd dis DeployApplication.

The infacmd mi exportSpec command uses the following syntax:

```
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
```

The following table describes infacmd mi exportSpec options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-MISpecName -spec	mi_spec_name	Required. Name of the mass ingestion specification that you want to export.
-Directory -dir	dir_path	Required. The directory where you want to write the application archive file.
-HadoopConnection -hc	hadoop_connection	Required. The Hadoop connection that the Data Integration Service will use to run the mass ingestion job when you import the application archive file and run the application. You must specify the Hadoop connection because a Hadoop connection does not persist for the mass ingestion specification while the specification is stored in the Model repository.

extendedRunStats

Gets the extended ingestion statistics for a specific source table in a deployed mass ingestion specification. To get the extended statistics, you must specify the RunID of the mass ingestion specification, the name of the source table, and the mapping type.

The extended statistics report the ingestion statistics for table rows ingested from the source and the ingestion statistics for table rows ingested in the target. The statistics list the number of rows that were ingested successfully and the number of rows that contain errors.

If the run instance uses an incremental load, the extended statistics also report the incremental key and the start value. The incremental key is the name of the column that the Spark engine used to fetch incremental data in the source table. The start value is the value that the Spark engine used to start ingesting incremental data.

The infacmd mi extendedRunStats command uses the following syntax:

```
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
```

The following table describes infacmd mi extendedRunStats options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification associated with the ingestion mapping job.

Option	Argument	Description
-RunID -rid	run_id	Required. Run identifier number, or the Run ID, of the mass ingestion specification run instance. To find the RunID for a run instance, list the specification run instances using infacmd mi listSpecRuns.
-SourceName -srcName	source_name	Required. Name of the source table in the run instance of the mass ingestion specification. To find the name of the source table, get the ingestion run statistics using infacmd mi getSpecRunStats.
-MappingTp -mtp	mapping_type	Required. The mapping type corresponds to the run-time engine that runs the ingestion mapping job for the source table. To find the mapping type, get the ingestion run statistics using infacmd mi getSpecRunStats.

getSpecRunStats

Gets the detailed run statistics for a deployed mass ingestion specification. To get the statistics, you must specify a RunID. To find the RunID for a run instance, list the specification run instances using infacmd mi listSpecRuns.

The detailed run statistics report the JobID for each ingestion mapping job in the deployed mass ingestion specification, the name of the source table that each mapping job ingests, the run start time, the end time, the run-time engine that runs the mapping job, and the job status. The JobID is the ID of the ingestion mapping job that ingests the source table. The status might display Completed, Failed, Canceled, Running, Aborted, Queued, or Unknown.

The infacmd mi getSpecRunStats command uses the following syntax:

```
getSpecRunStats

<-DomainName|-dn> domain_name

<-UserName|-un> user_name

<-Password|-pd> password

[<-SecurityDomain|-sdn> security_domain]

<-ServiceName|-sn> service_name

<-runID|-rid> run_id
```

The following table describes infacmd mi getSpecRunStats options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-runID -rid	run_id	Required. Run identifier number, or the Run ID, of the mass ingestion specification run instance. To find the RunID for a run instance, list the specification run instances using infacmd mi listSpecRuns.

listSpecRuns

Lists the run instances of a deployed mass ingestion specification. Each run instance is defined by a RunID. When you list the run instances, you must specify the Mass Ingestion Service.

The detailed run statistics report the RunID for each specification run instance, the load type, the run instance start time, the Data Integration Service where the mass ingestion specification is deployed, the user who started the run, and the job status for each run instance. The status might display Completed, Failed, Cancelled, Running, Queued, or Unknown.

The infacmd mi listSpecRuns command uses the following syntax:

```
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
```

The following table describes infacmd mi listSpecRuns options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-MISpecName -spec	mi_spec_name	Required. Name of the mass ingestion specification.

listSpecs

Lists the mass ingestion specifications. When you list specifications, you must specify the Mass Ingestion Service.

The infacmd mi listSpecs command uses the following syntax:

```
listSpecs

<-DomainName|-dn> domain_name

<-UserName|-un> user_name

<-Password|-pd> password

[<-SecurityDomain|-sdn> security_domain]

<-ServiceName|-sn> service_name
```

The following table describes infacmd mi listSpecs options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specifications.

restartMapping

Restarts the ingestion mapping jobs in a mass ingestion specification. Specify the list of source tables to restart. You must specify the Mass Ingestion Service and the RunID for the run instance of the mass ingestion specification. You can also specify whether you want to restart only the source tables that failed.

The infacmd mi restartMapping command uses the following syntax:

```
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]
```

The following table describes infacmd mi restartMapping options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the ingestion of the source tables.
-runID -rid	run_id	Required. Run identifier number (Run ID) of the mass ingestion specification run instance.
-SourceList -srcList	comma_separated_source_list	Required. The list of source tables to restart. Separate each source table with a comma.
-OnlyFailed -failed	true|false	Optional. Enter true to restart only the source tables that failed to be ingested. Enter false to restart all source tables.

runSpec

Runs a mass ingestion specification that is deployed to a Data Integration Service. Before you can run a specification, you must deploy the specification using infacmd mi deploySpec.

The infacmd mi runSpec command uses the following syntax:

```
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]
```

The following table describes infacmd mi runSpec options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	security_domain	Optional. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. Default is Native.
-ServiceName -sn	service_name	Required. Name of the Mass Ingestion Service that manages the mass ingestion specification.
-MISpecName -spec	mi_spec_name	Required. Name of the mass ingestion specification that is deployed to the Data Integration Service.
-LoadType -lt	load_type	Optional. Type of load to ingest the data in the mass ingestion specification. Use `full` or `incremental`. Default is `full`. If incremental load is not enabled in the mass ingestion specification, you cannot use an incremental load to ingest the data.

Option	Argument	Description
-DISServiceName -dis	data_integration_service	Required. Name of the Data Integration Service where the mass ingestion specification is deployed.
-OperatingSystemProfile -osp	operating_system_profile_name	Optional. Name of the operating system profile configured for the Data Integration Service.

INDEX

P

R

S

V