



Informatica™

Informatica® Big Data Management
10.2.1

Mass Ingestion Guide

Informatica Big Data Management Mass Ingestion Guide

10.2.1

May 2018

© Copyright Informatica LLC 2018, 2019

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, Big Data Management, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2019-01-28

Table of Contents

Preface	6
Informatica Resources.	6
Informatica Network.	6
Informatica Knowledge Base.	6
Informatica Documentation.	6
Informatica Product Availability Matrixes.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
Chapter 1: Introduction to Mass Ingestion.....	8
Overview.	8
Use Case.	9
Architecture.	12
Clients and Tools.	12
Application Services.	13
Repositories.	13
Process.	13
Roles.	15
Administrator Role.	15
Mass Ingestion Tool.	15
Logging In to the Mass Ingestion Tool.	16
Mass Ingestion Tool User Interface.	17
Chapter 2: Prepare.....	19
Prepare Overview.	19
Creating a Mass Ingestion Service.	19
Configuring Monitoring.	21
Step 1. Configure Monitoring Settings.	21
Step 2. Configure Reports and Statistics.	22
Assigning the Administrator Role.	23
Configuring the JDK Home Directory.	23
Chapter 3: Create.....	24
Create Overview.	24
Definition.	25
Source.	26
Target.	27
Configuring a Hive Target.	27
Configuring an HDFS Target.	30

Mass Transformations.	31
Filter Clauses.	33
Regular Expressions.	34
Replace Criteria.	36
Transformation Override.	37
Transformation Override for Hive Targets.	38
Transformation Override for HDFS Targets.	40
Creating a Mass Ingestion Specification.	40
Chapter 4: Deploy.	42
Deploy Overview.	42
Deploying a Mass Ingestion Specification.	43
Redeploying a Mass Ingestion Specification.	44
Deploy to an Application Archive File.	45
Migrate a Mass Ingestion Specification.	45
Chapter 5: Run.	47
Run Overview.	47
Viewing Run Instances for a Mass Ingestion Specification.	48
Run Instance Status.	48
Running a Mass Ingestion Specification.	49
Chapter 6: Monitor.	50
Monitor Overview.	50
Monitor a Mass Ingestion Specification in the Mass Ingestion Tool.	50
Overview Page.	50
Execution Statistics Page.	54
Monitoring Execution Statistics.	57
Filtering Ingestion Objects by Status.	59
Restart Ingestion Jobs.	60
Monitor a Mass Ingestion Specification in the Administrator Tool.	63
Monitor the Application and Ingestion Mapping Jobs.	63
Canceling Ingestion Jobs.	65
Rules and Guidelines.	65
Appendix A: infacmd mi Command Reference.	67
abortRun.	67
createService.	68
deploySpec.	71
exportSpec.	72
extendedRunStats.	73
getSpecRunStats.	74
listSpecRuns.	75

listSpecs.	76
restartMapping.	77
runSpec.	78
Index.	80

Preface

The Informatica *Big Data Management*™ *Mass Ingestion Guide* provides information about mass ingestion jobs. The guide contains information that you need to configure the Informatica domain to support mass ingestion and the information that you need to perform mass ingestion jobs. This guide assumes that you are familiar with the Informatica domain.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit <https://network.informatica.com>.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- 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 search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit <https://kb.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

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

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

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrixes>.

Informatica Velocity

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

If you are an Informatica Network member, you can access 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 augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

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

If you are an Informatica Network member, you can use Online Support at <http://network.informatica.com>.

CHAPTER 1

Introduction to Mass Ingestion

This chapter includes the following topics:

- [Overview, 8](#)
- [Use Case, 9](#)
- [Architecture, 12](#)
- [Process, 13](#)
- [Roles, 15](#)
- [Mass Ingestion Tool, 15](#)

Overview

Mass ingestion is the ingestion or replication of large amounts of data for use or storage in a database or a repository. The database can be a data lake, a cloud repository, or a Hadoop cluster.

To ingest or replicate large amounts of data between a relational database and a Hive or HDFS target, use the Mass Ingestion tool. In the Mass Ingestion tool, you can create a mass ingestion specification.

A mass ingestion specification is a configuration that determines how a data source is ingested into a specific location in the Hive or HDFS target. In the specification, you configure the relational source and the Hive or HDFS target. You can also configure parameters to perform a light transformation on the ingested data. For example, you can filter the data to ingest only certain columns or you can mask the data to protect private information.

Deploy and run the mass ingestion specification to ingest all of the data at once. The specification is deployed to a Data Integration Service. When you run the specification, the Data Integration Service connects to the Hadoop environment. In the Hadoop environment, the Blaze, Spark, and Hive engines ingest the data to the target. As the mass ingestion specification runs, you can begin monitoring the ingestion process.

A mass ingestion specification replaces the need to manually create and run mappings. You create one mass ingestion specification that ingests all of the data at once.

Use Case

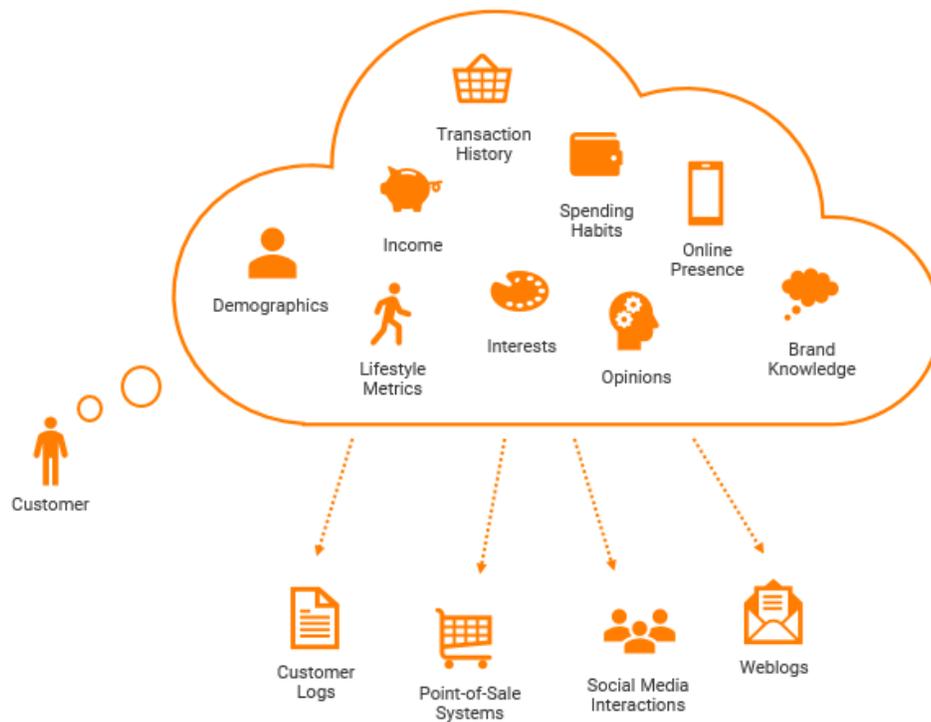
You work in an IT group for a commercial bank. Your team is taking on a new project to personalize the rewards program that you offer to customers who open checking and savings accounts at your bank.

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.

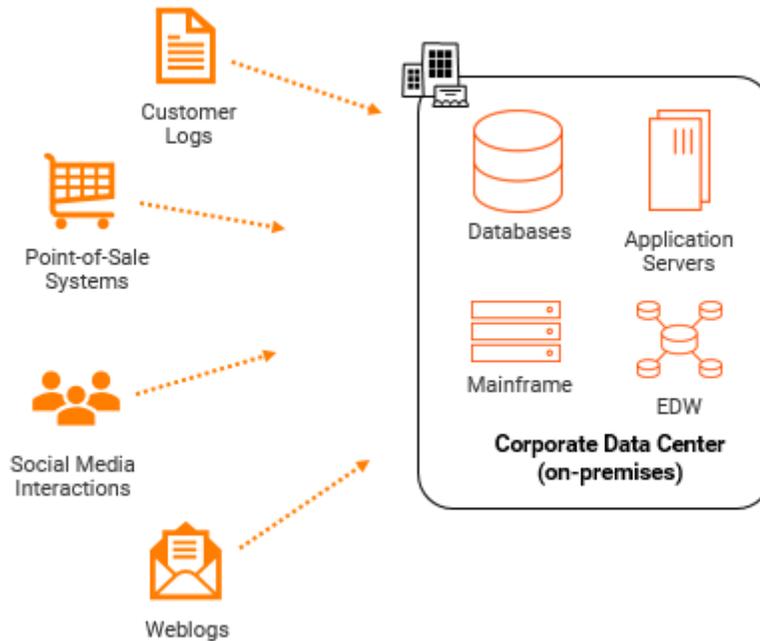
You collect the data through different media, including customer logs on file with the bank, point-of-sale systems at companies that the bank partners with, social media interactions, and customer weblogs.

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 different relational databases.

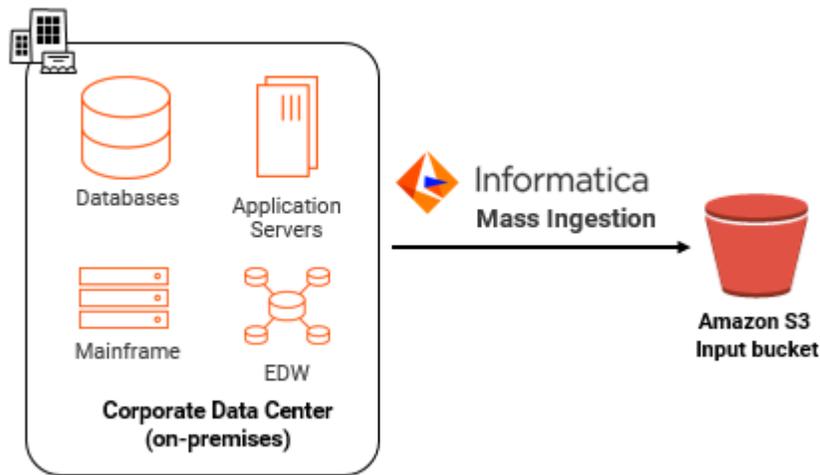
The following image shows how the data is stored in the bank's corporate data center and the databases that the bank might use to store the data:



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.

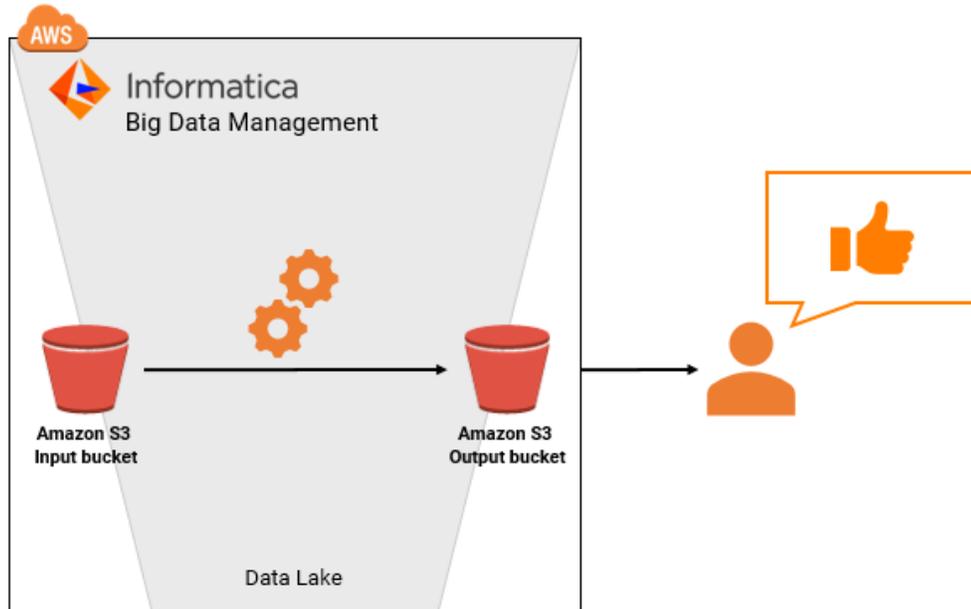
Instead of manually creating and running mappings, you can use mass ingestion. You create one mass ingestion specification that ingests all of the data at once. You have to specify only the source, target, and the parameters that you want to configure. You deploy the specification and all of the data is ingested at once.

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:



After the data is ingested into Amazon S3 input buckets, you use Big Data Management™ to prepare the data and write the data to Amazon S3 output buckets.

The following image shows where you can use Big Data Management™ to prepare and write the data to Amazon S3 output buckets:

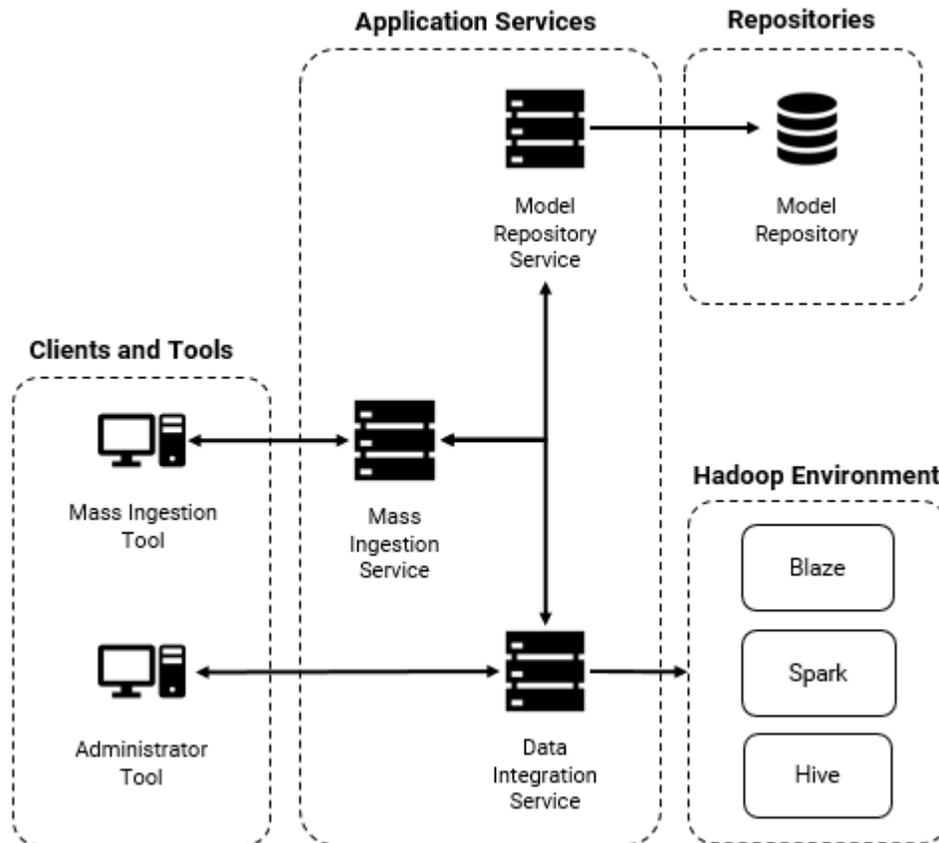


Mass ingestion saved you a lot of time and resources and your data analysts have much more time to analyze the data and develop a new system for the bank's rewards program.

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.

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 connects to the Model repository when you create a mass ingestion specification to store the specification in the Model repository.

Data Integration Service

Manages deployed mass ingestion specifications. Connects to the Hadoop environment to ingest the data configured in the mass ingestion specification.

Repositories

The mass ingestion components include the following repositories:

Model repository

The repository that stores the mass ingestion specification metadata when you create a mass ingestion specification.

Mass ingestion specifications are stored in the project `INFA_MI_PROJECT` in the Model repository. 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. You cannot view the project in the Developer tool, and you cannot view the specifications stored in the project from the Developer tool. You can only list the specifications that are stored in the project from the command line.

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 stores 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. You can deploy the application to a Data Integration Service.

Run

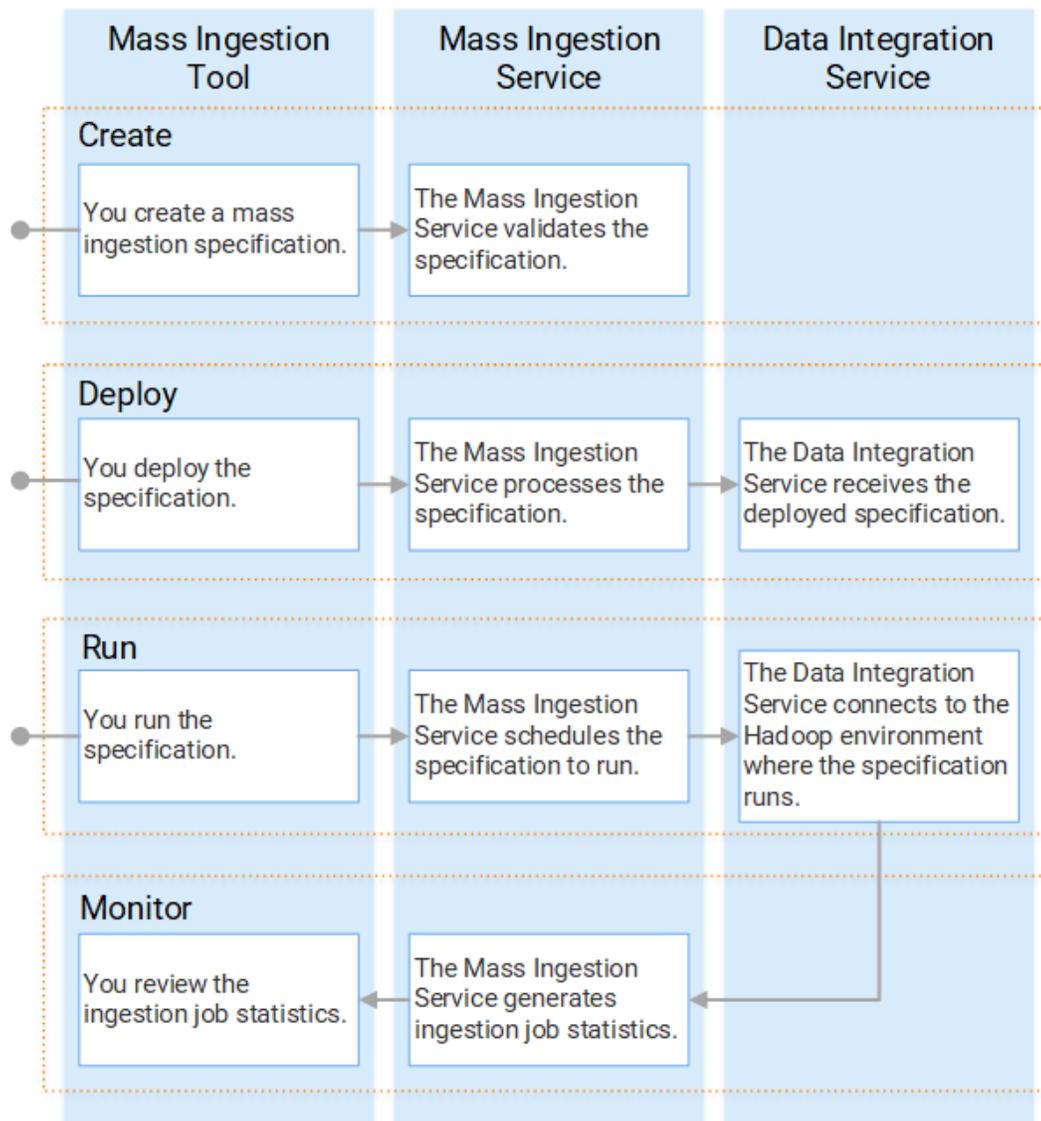
You run the mass ingestion specification to ingest data to Hive or HDFS. The Mass Ingestion Service schedules the specification to run. The Data Integration Service connects to the Hadoop environment. In the Hadoop environment, the Blaze, Spark, and Hive engines ingest the data to the target.

Monitor

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

You can also monitor the statistics in the Administrator tool to monitor the application and mappings that perform the ingestion job.

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



Roles

A role is a collection of privileges that you can assign to users and groups.

A role includes privileges for the Mass Ingestion Service and the Mass Ingestion tool. You assign roles to users or groups for the Mass Ingestion Service. If a user or group is assigned a role for the Mass Ingestion Service, the user or group can perform additional tasks defined by the role in the Mass Ingestion tool.

For example, you can create an Administrator role that includes privileges for the Mass Ingestion Service. You can assign the Administrator role to a user for the Development Mass Ingestion Service. The user can perform tasks defined by the Administrator role in the Mass Ingestion Service and the Mass Ingestion tool. You can revoke the role from the user for the Production Mass Ingestion Service and assign the Administrator role to a different user.

You can assign only the Administrator role to a user or group for the Mass Ingestion Service. You cannot assign custom roles. You cannot assign privileges and permissions.

Administrator Role

When you assign the Administrator role to a user or group for the Mass Ingestion Service, the user or group can complete tasks that are determined by the Administrator role, not by privileges or permissions.

You can assign a user or group all privileges for the domain and then grant the user or group full permissions on all domain or repository objects. However, this user or group cannot complete the tasks determined by the Administrator role.

For example, a user assigned the Administrator role for the Mass Ingestion Service can configure Mass Ingestion Service properties in the Administrator tool. A user assigned all domain privileges and permissions cannot configure Mass Ingestion Service properties.

The user assigned the Administrator role for the Mass Ingestion Service can perform the following tasks:

- Browse all mass ingestion specifications.
- Edit a mass ingestion specification.
- Run a mass ingestion specification.
- Delete a mass ingestion specification.

A user who is not assigned the Administrator role can perform the same tasks as the user who is assigned the Administrator role, but the user can browse, edit, run, and delete only the mass ingestion specifications that the user creates.

Note: You can assign only the Administrator role to a user or group. You cannot assign privileges and permissions for the Mass Ingestion Service.

Mass Ingestion Tool

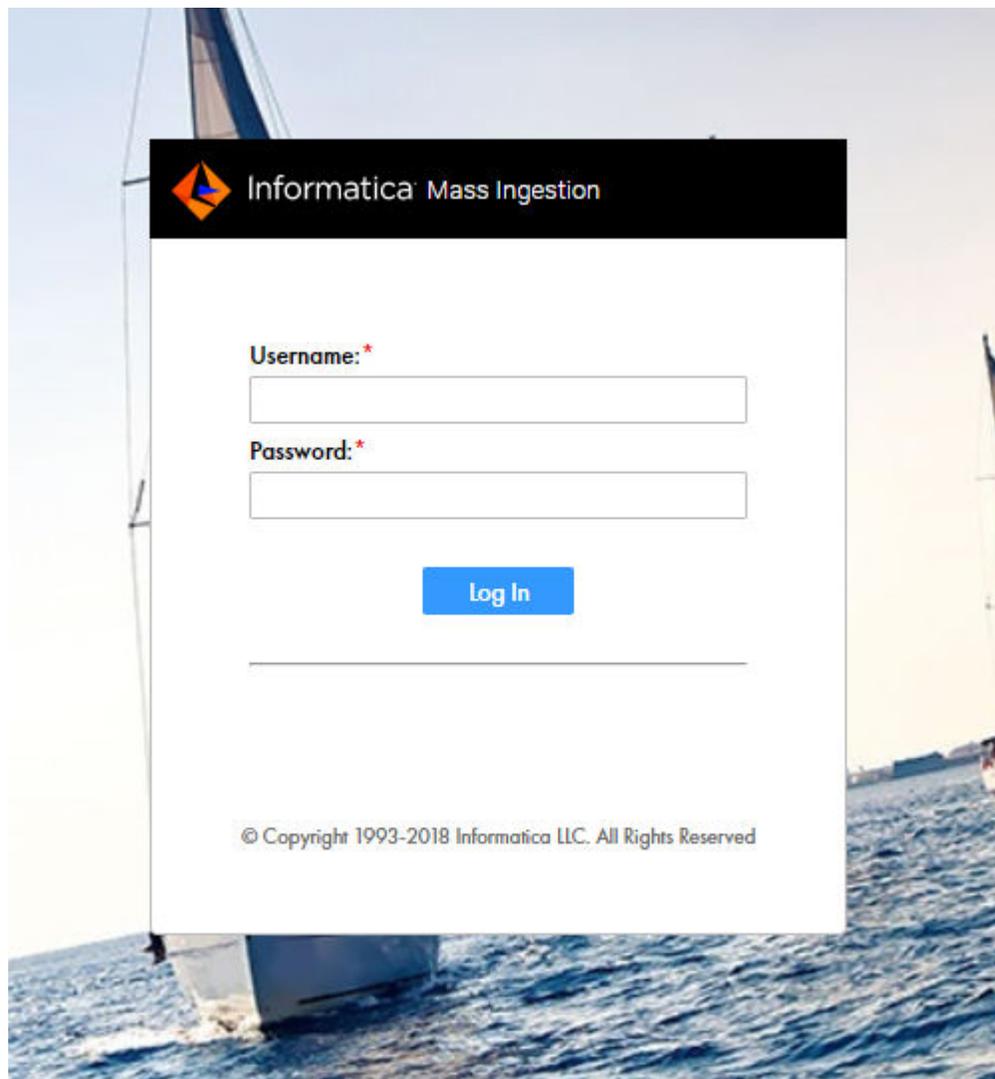
The Mass Ingestion tool is a web application that you use to run mass ingestion jobs.

You can use the Mass Ingestion tool to design a mass ingestion specification to specify how you want to ingest data. You can 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. The Mass Ingestion tool is a web application where you create, deploy, run, and monitor mass ingestion specifications. To get the URL for the Mass Ingestion tool, contact your system administrator.

1. Open Google Chrome.
2. In the Address field, enter the following URL provided by your system administrator:
`http://<host>:<port>/mi/login`
3. On the login page for the Mass Ingestion tool, enter your user name and password.



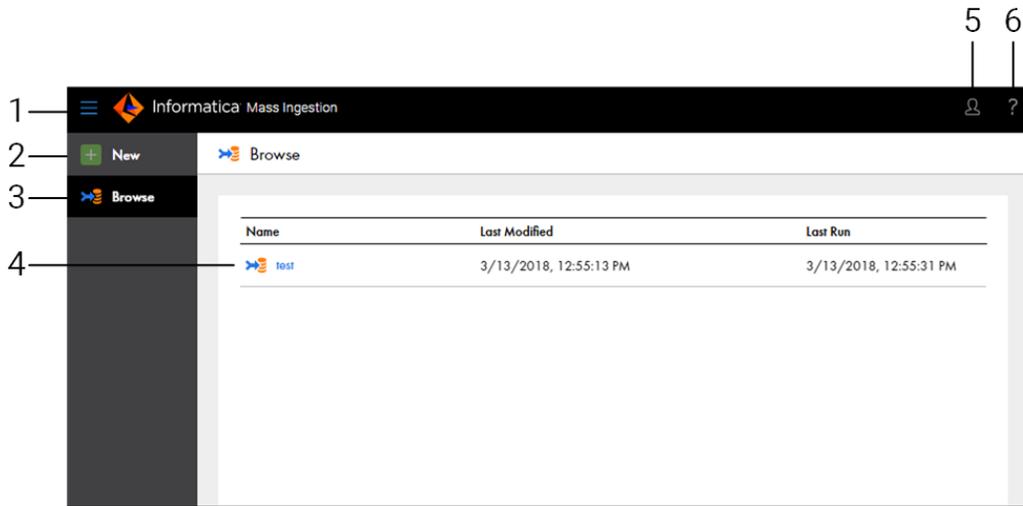
4. Click **Log In**.
The landing page for the Mass Ingestion tool appears.

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, 19](#)
- [Creating a Mass Ingestion Service, 19](#)
- [Configuring Monitoring, 21](#)
- [Assigning the Administrator Role, 23](#)
- [Configuring the JDK Home Directory, 23](#)

Prepare Overview

Before you can use the Mass Ingestion tool, you must use the Administrator tool to configure the Mass Ingestion Service.

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.
- Configure the JDK home directory on the Data Integration 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.

- Click **Actions > New > Mass Ingestion Service**.

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

- 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.

- Click **Next**.

- 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.

- Click **Next**.

The **New Mass Ingestion Service - Step 3 of 3** page appears.

- Enter the HTTP port number to use for the Mass Ingestion Service.

You cannot use HTTPS.

- To enable the Mass Ingestion Service, select **Enable Service**.

- Click **Finish**.

The domain creates the Mass Ingestion Service. If you selected **Enable Service**, the domain enables the Mass Ingestion Service.

- In the **Domain Navigator**, select the Mass Ingestion Service.

- 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 run mass ingestion specifications store statistics and reports in a Model repository. The statistics include historical information about objects that the Data Integration Services run.

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

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.

Option	Description
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**.

Configuring the JDK Home Directory

To run mass ingestion specifications using a Sqoop connection, you must install the Java Development Kit (JDK) on the machine that runs the Data Integration Service and configure the JDK Home Directory property on the Data Integration Service.

For more information, see the *Informatica Big Data Management Hadoop Integration Guide*.

CHAPTER 3

Create

This chapter includes the following topics:

- [Create Overview, 24](#)
- [Definition, 25](#)
- [Source, 26](#)
- [Target, 27](#)
- [Mass Transformations, 31](#)
- [Transformation Override, 37](#)
- [Creating a Mass Ingestion Specification, 40](#)

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 Hive or HDFS. You select the relational database that contains the data that you want to ingest and you configure the target where you want to ingest the specified data. When you create a mass ingestion specification, the Mass Ingestion Service validates the inputs and stores the specification in the Model repository.

When you create the mass ingestion specification, you use a wizard that helps you configure the ingestion job. In the wizard, you define the following properties:

1. **Definition.** Describe the type of specification that you want to create.
2. **Source.** Define the source database that you want to use to ingest data.
3. **Target.** Define the Hive or HDFS target where you want to ingest data.
4. **Mass Transformations.** Define the transformations to apply parameters across all source tables that you ingest.
5. **Transformation Override.** Override the mass transformations and specify parameters for specific tables.

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 how the specification ingests data. You can configure the specification to ingest data from a relational database to Hive or from a relational database to HDFS. The scenario that you specify defines the properties that you can configure in the rest of the specification.

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.
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 Hive or from a relational database to HDFS. You cannot edit the ingestion scenario after you create the mass ingestion specification.

Source

The source describes the source database that contains the data that you want to ingest to Hive or HDFS. 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.

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 database:

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.

Target

The target describes the Hive or HDFS target where you want to ingest the source tables from the relational database.

You specify different target properties depending on the scenario that you select in the specification definition. If you configure the specification to ingest data from a relational database to Hive, you configure a Hive connection and Hive table properties to define the target. If you configure the specification to ingest data from a relational database to HDFS, you configure an HDFS connection and an ingestion directory to define the target.

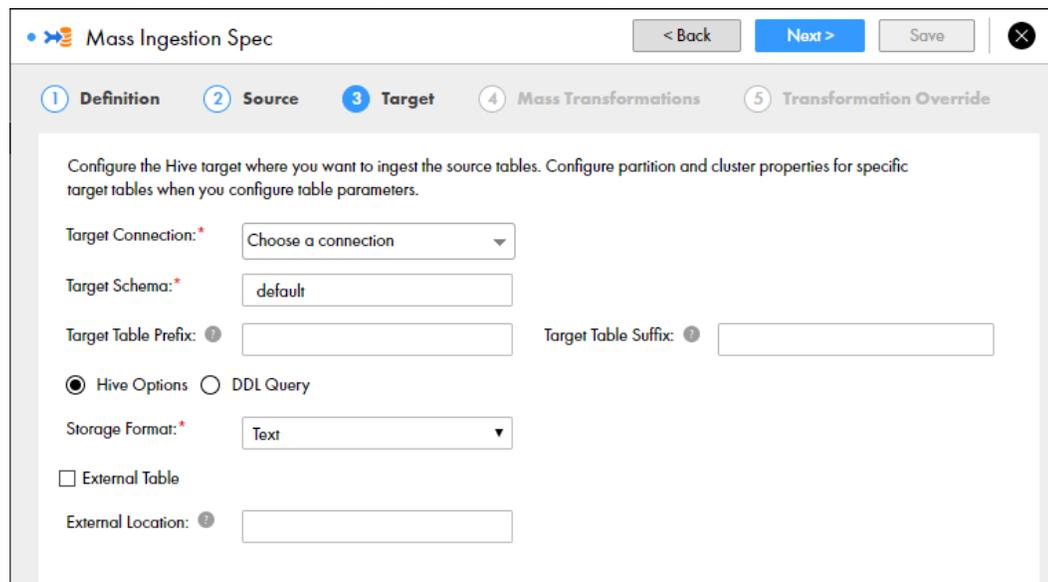
The mass ingestion solution ingests all data to the target. The mass ingestion solution does not provide the option to append data that has been recently updated. Each time that you run the mass ingestion specification, the existing data in the Hive or HDFS target is deleted and replaced with the data configured in the ingestion job.

Configuring a Hive Target

Configure a Hive target to ingest data to a Hive table. 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, Azure Blob, HDFS, WASB, or ADLS.

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



The screenshot shows a web application interface for configuring a Hive target. The page title is "Mass Ingestion Spec". At the top, there are navigation buttons: "< Back", "Next >" (highlighted in blue), "Save", and a close button (X). Below the title, there are five numbered tabs: 1 Definition, 2 Source, 3 Target (highlighted in blue), 4 Mass Transformations, and 5 Transformation Override. The main content area contains the following fields and options:

- Target Connection:** A dropdown menu with the text "Choose a connection".
- Target Schema:** A text input field containing "default".
- Target Table Prefix:** An empty text input field.
- Target Table Suffix:** An empty text input field.
- Options:** Two radio buttons: "Hive Options" (selected) and "DDL Query".
- Storage Format:** A dropdown menu with "Text" selected.
- External Table:** A checkbox that is currently unchecked.
- External Location:** An empty text input field.

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 Text, Avro, Parquet, or ORC. Default is Text.
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 <code>/temp</code> . A source table named <code>PRODUCT</code> is ingested to the external location <code>/temp/PRODUCT/</code>

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

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:

The screenshot shows the 'Mass Ingestion Spec' configuration window, specifically the 'Target' tab. The window has a title bar with a back button, a 'Next >' button, a 'Save' button, and a close button. Below the title bar are five tabs: '1 Definition', '2 Source', '3 Target', '4 Mass Transformations', and '5 Transformation Override'. The 'Target' tab is active. The main content area contains the following fields and options:

- Target Connection:** A dropdown menu with the text 'Choose a connection'.
- Target Schema:** A text input field containing 'default'.
- Target Table Prefix:** A text input field.
- Target Table Suffix:** A text input field.
- Options:** Two radio buttons: 'Hive Options' (unselected) and 'DDL Query' (selected).
- DDL Query:** A text area containing the query: `CREATE TABLE {{INFA_TABLE_NAME}} ({{INFA_COLUMN_LIST}})`.

Configuring an HDFS Target

Configure an HDFS target to ingest 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.

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 <code>/temp</code> . A source table named <code>PRODUCT</code> is ingested to the directory <code>/temp/PRODUCT/</code>
Compression	Required. The compressed file format that stores the target files. You can select None, Gzip, Bzip2, LZ0, 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 Delimiter. Enter a custom delimiter.

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
- LZ0
- Snappy
- Custom

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

Mass Transformations

Configure mass transformations to define the parameters to apply across all tables ingested from the source 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	<p>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 <code>STATE='California'</code>. You can use any transformation language functions in the filter clause. The filter clause must evaluate to TRUE or FALSE.</p> <p>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 <code>STATE='California' AND STATUS='Single'</code>.</p>
Drop Columns	<p>Drops columns in the target table. Enter each column as a string and separate column names with a comma, or specify a regular expression.</p> <p>For example, if you enter <code>COL1</code>, the column COL1 will be dropped in the target table.</p> <p>If you enter a regular expression such as <code>.*SSN.*</code>, columns that have SSN in the column name will be dropped.</p>
Trim	<p>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.</p> <p>For example, if you enter <code>COL1</code>, values in the column COL1 will be trimmed in the target table.</p> <p>If you enter a regular expression such as <code>.*SSN.*</code>, values in columns that have SSN in the column name will be trimmed.</p>
Convert to Uppercase	<p>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.</p> <p>For example, if you enter <code>COL1</code>, values in the column COL1 will be converted to upper case in the target table.</p> <p>If you enter a regular expression such as <code>.*SSN.*</code>, values in columns that have SSN in the column name will be converted to upper case.</p>
Convert to Lowercase	<p>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.</p> <p>For example, if you enter <code>COL1</code>, values in the column COL1 will be converted to lower case in the target table.</p> <p>If you enter a regular expression such as <code>.*SSN.*</code>, values in columns that have SSN in the column name will be converted to lower case.</p>
Replace Columns	<p>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.</p> <p>For example, if you enter <code>COL1</code>, values in the column COL1 will be replaced in the target table.</p> <p>If you enter a regular expression such as <code>.*SSN.*</code>, values in columns that have SSN in the column name will be replaced.</p> <p>If you specify columns to replace, you must specify the replace criteria.</p>
Replace Criteria	<p>Required if you specify columns to replace. Determines how to replace column values in the target table. You can select <code>Pattern</code> or <code>Entire String</code>.</p> <p>If you select <code>Entire String</code>, all values in the columns that you specify are replaced by the value that you configure.</p> <p>If you select <code>Pattern</code>, enter the pattern to be replaced. Then enter the value to replace the pattern.</p>

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

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 ingest the following source table:

EMPLOYEEID	PHONENUMBER
607014	(630) 4468851
620368	(904) 3854084

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

You want to filter the rows that you ingest to the target. You can 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 on 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`.

The following table describes the metacharacters that you can use in a regular expression:

Metacharacter	Description
.	Matches any single character.
[]	Indicates a character class. Matches any character inside the brackets. For example, [abc] matches "a," "b," and "c."
^	If this metacharacter occurs at the start of a character class, it negates the character class. A negated character class matches any character except those inside the brackets. For example, [^abc] matches all characters except "a," "b," and "c." If this metacharacter occurs at the beginning of the regular expression, it matches the beginning of the input. For example, ^[abc] matches the input that begins with "a," "b," or "c."
-	Indicates a range of characters in a character class. For example, [0-9] matches any of the digits "0" through "9."

Metacharacter	Description
?	Indicates that the preceding expression to this metacharacter is optional. It matches the preceding expression zero or one time. For example, [0-9][0-9]? matches "2" and "12."
+	Indicates that the preceding expression matches one or more times. For example, [0-9]+ matches "1," "13," "666," and similar combinations.
*	Indicates that the preceding expression matches zero or more times. For example, the input <abc*> matches <abc>, <abc123>, and similar combinations that contains <abc> as the preceding expression.
??, +?, *?	Modified versions of ?, +, and *. These match as little as possible, unlike the versions that match as much as possible. For example, the input "<abc><def>," <.*?> matches "<abc>" and the input <.*> matches "<abc><def>."
()	Grouping operator. For example, (\d+)*\d+ matches a list of numbers separated by commas such as "1" or "1,23,456."
{ }	Indicates a match group.
\	An escape character, which interprets the next metacharacter literally. For example, [0-9]+ matches one or more digits, but [0-9]\+ matches a digit followed by a plus character. Also used for abbreviations such as \a for any alphanumeric character. If \ is followed by a number <i>n</i> , it matches the <i>n</i> th match group, starting from 0. For example, <{.*?}>.*?</\0> matches "<head>Contents</head>". In C++ string literals, two backslashes must be used: "\\+", "\\a," "<{.*?}>.*?</\0>."
\$	At the end of a regular expression, this character matches the end of the input. For example, [0-9]\$ matches a digit at the end of the input.
	Alternation operator that separates two expressions, one of which matches. For example, T the matches "The" or "the."
!	Negation operator. The expression following ! does not match the input. For example, a!b matches "a" not followed by "b."

The following table describes the abbreviations that you can use in the regular expressions:

Abbreviation	Definition
\a	Any alphanumeric character, ([a-zA-Z0-9]).
\b	White space (blank), ([\t]).
\c	Any alphabetic character, ([a-zA-Z]).
\d	Any decimal digit, ([0-9]).
\h	Any hexadecimal digit, ([0-9a-fA-F]).
\n	Newline, (\r \r?\n).
\q	Quoted string, (\("[^"]*" '['']*').

Abbreviation	Definition
\w	Simple word, ([a-zA-Z]+).
\z	Integer, ([0-9+]).

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:

Mass Ingestion Spec

< Back Next > Save ✕

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

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

Filter By:

Drop Columns:

Trim:

Convert to Uppercase:

Convert to Lowercase:

Replace Columns:

⚠ Columns COL1 will be replaced.

Replace Criteria: Value:

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

Mass Ingestion Spec

< Back Next > Save

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

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

Filter By:

Drop Columns:

Trim:

Convert to Uppercase:

Convert to Lowercase:

Replace Columns:

⚠ Columns COL1 will be replaced.

Replace Criteria: Value:

Transformation Override

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

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. Or, clear the mass transformations and set specific parameters for each table.

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

You configure different parameters depending on whether you ingest data to a Hive or HDFS target.

Transformation Override for Hive Targets

When you configure a transformation override for a Hive target, you can edit the mass transformations that are applied to specific source tables when they are ingested to the Hive target. You can additionally edit the Hive target table properties for each source table that you ingest.

The following image shows the transformation override that you can configure for a Hive target:

Mass Ingestion Spec

< Back Next > Save

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

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

Ingestion Objects (1)

Sources	Filter By	Drop Columns	Trim	Convert to Uppercase	Convert to Lowercase	Replace Columns	HIVE
EMPLOYEE							

You can edit the following properties for each target table:

- 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. Edit the Hive properties for the target table.

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

Hive Options for: TEST

Hive Options DDL Query

Storage Format: Text

External Location: None

Partition Key: ?

Clustered By: ?

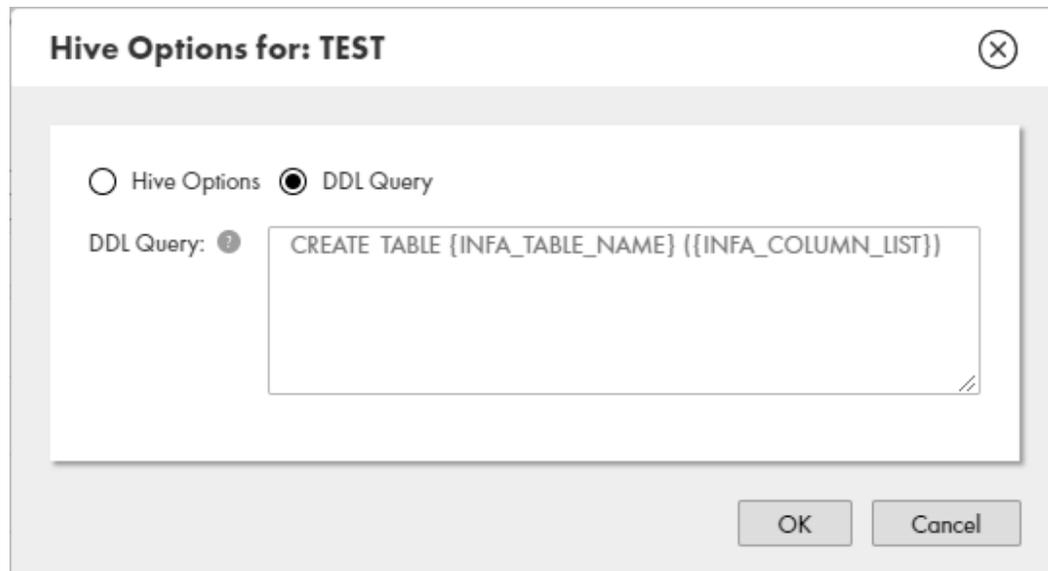
Number of buckets:

OK Cancel

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 <code>/temp</code> . A source table named <code>PRODUCT</code> is ingested to the external location <code>/temp/PRODUCT/</code>
Partition Key	The partition key for the target Hive table.
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, you can 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.

Transformation Override for HDFS Targets

When you configure a transformation override for an HDFS target, you can edit the mass transformations that are applied to specific source tables when they are ingested to the HDFS target.

The following image shows the transformation override that you can configure for an HDFS target:

Sources	Filter By	Drop Columns	Trim	Convert to Uppercase	Convert to Lowercase	Replace Columns
EMPLOYEE						

You can edit the following properties for each target table:

- 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.

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. On the **Definition** page, configure the definition for the mass ingestion specification.
3. Click **Next**.
4. On the **Source** page, configure the source database and the source tables that you want to ingest.
5. Click **Next**.
6. On the **Target** page, configure the Hive or HDFS target.
7. Click **Next**.
8. On the **Mass Transformations** page, configure the parameters that you want to apply to all source tables.
9. Click **Next**.
10. On the **Transformation Override** page, edit the parameters for specific tables.

If you need to enter a column or a regular expression, enter the column name or the regular expression in the text box, or click the arrow to select columns from a list of columns that exist in the table.

If the mass ingestion specification ingests data to a Hive target, you can also edit the Hive target properties for individual tables.

11. Click **Save**.

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

CHAPTER 4

Deploy

This chapter includes the following topics:

- [Deploy Overview, 42](#)
- [Deploying a Mass Ingestion Specification, 43](#)
- [Redeploying a Mass Ingestion Specification, 44](#)
- [Deploy to an Application Archive File, 45](#)
- [Migrate a Mass Ingestion Specification, 45](#)

Deploy Overview

After you create the mass ingestion specification, you can deploy the specification to a Data Integration Service and specify a Hadoop connection. The Data Integration Service uses the Hadoop connection to connect to the Hadoop environment when you run the deployed specification.

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.

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.

Deploying a Mass Ingestion Specification

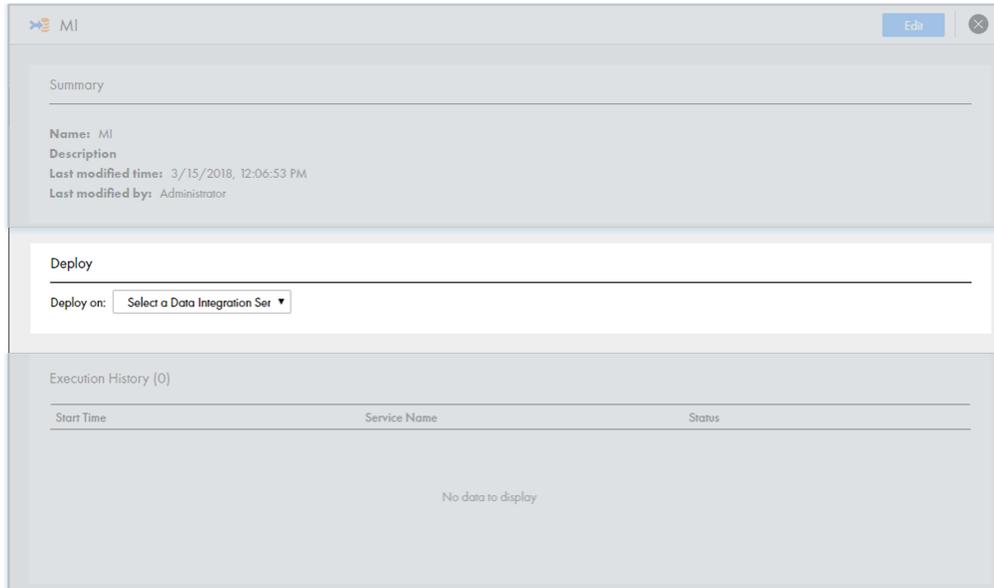
Deploy a mass ingestion specification in the Mass Ingestion tool.

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:



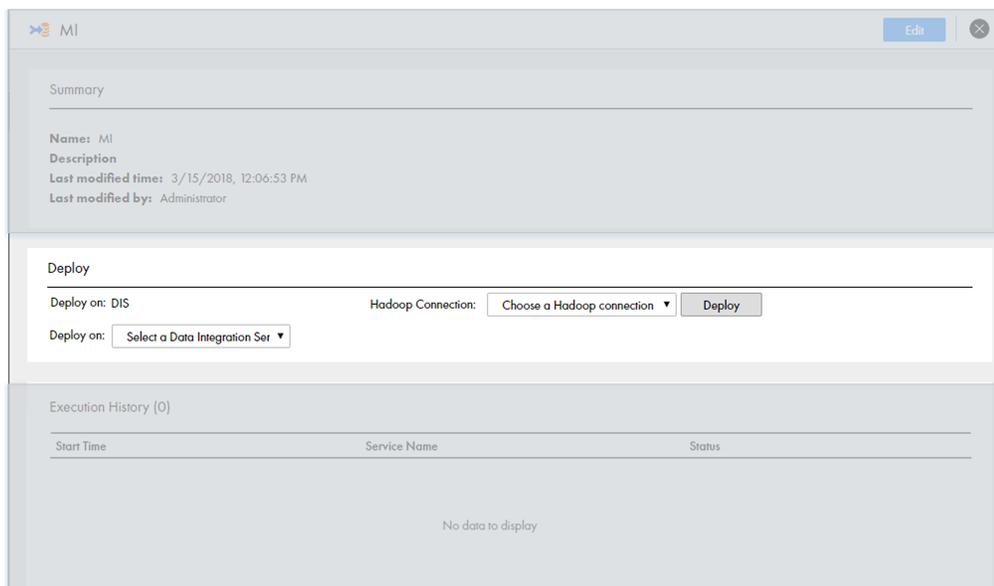
The screenshot shows a web interface for a Mass Ingestion (MI) specification. At the top, there is a header with the MI icon and the text 'MI', along with 'Edit' and 'Close' buttons. Below this is a 'Summary' section containing the following information: Name: MI, Description, Last modified time: 3/15/2018, 12:06:53 PM, and Last modified by: Administrator. The main section is titled 'Deploy' and contains a 'Deploy on:' dropdown menu with the text 'Select a Data Integration Ser'. Below the Deploy section is an 'Execution History (0)' section with a table header: Start Time, Service Name, and Status. The table is currently empty, displaying 'No data to display'.

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:



This screenshot shows the same interface as the previous one, but with additional options in the 'Deploy' section. The 'Deploy on:' dropdown is still present. To its right, there is a 'Hadoop Connection:' label followed by a dropdown menu with the text 'Choose a Hadoop connection' and a 'Deploy' button. The 'Execution History (0)' section remains empty with 'No data to display'.

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:

The screenshot displays the 'Deploy' view of a mass ingestion specification. It features a 'Summary' section with the following details: Name: MI, Description, Last modified time: 3/15/2018, 12:06:53 PM, and Last modified by: Administrator. The 'Deploy' section includes a 'Deployed on' field showing 'DIS on 3/15/2018, 11:56:15 AM', a 'Hadoop Connection' dropdown menu with the option 'Choose a Hadoop connection', a 'Redeploy' button, an 'Operating System Profile' dropdown menu with the option 'None', and a 'Run now' button. There is also a 'Deploy on' dropdown menu with the option 'Select a Data Integration Ser'. At the bottom, there is an 'Execution History (0)' section with a table header (Start Time, Service Name, Status) and a message 'No data to display'.

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 oie 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 on `infacmd`, see the *Informatica Command Reference*.

For more information on 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 oie exportObjects`. Import the mass ingestion specification to the project `INFA_MI_PROJECT` in the target repository using `infacmd oie 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 oie exportObjects`. Import the mass ingestion specification to the project `INFA_MI_PROJECT` in the target repository using `infacmd oie 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 oie`, see the "infacmd oie" 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, 47](#)
- [Viewing Run Instances for a Mass Ingestion Specification, 48](#)
- [Running a Mass Ingestion Specification, 49](#)

Run Overview

After you deploy the mass ingestion specification, you run the mass ingestion 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 specification and connects to the Hadoop environment. In the Hadoop environment, the Blaze, Spark, and Hive engines run individual jobs to ingest source tables into the target.

When you run the mass ingestion specification, the job ingests all data to the target. The mass ingestion solution does not provide the option to append only the data that has been recently updated to the target. When the Blaze, Spark, and Hive engines run the ingestion job, the existing data in the Hive or HDFS target is deleted and replaced with the data configured in the mass ingestion specification.

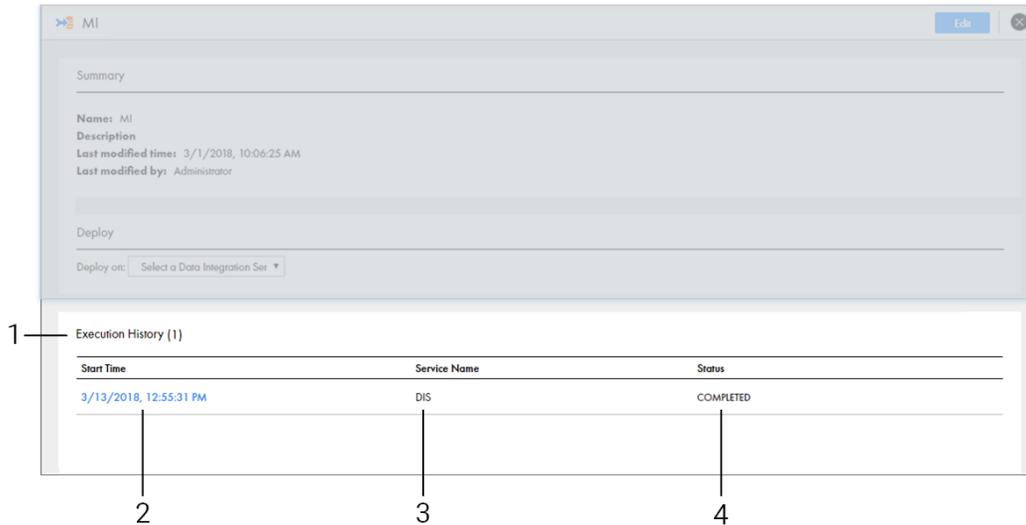
If the ingestion job stops responding or takes an excessive amount of time to complete, you can cancel the job in the Monitor tool.

For more information on canceling ingestion jobs, see [“Canceling Ingestion Jobs” on page 65](#) in the "Monitor" chapter.

Viewing Run Instances for a Mass Ingestion Specification

Each time that you run a mass ingestion specification, the Mass Ingestion Service generates a run instance. 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 on the Overview page:



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. 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 75](#) 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 <code>Unknown</code> in the following circumstances: <ul style="list-style-type: none"> - A Model repository is not configured in the monitoring settings for the Mass Ingestion Service. - The Data Integration Service is disabled.

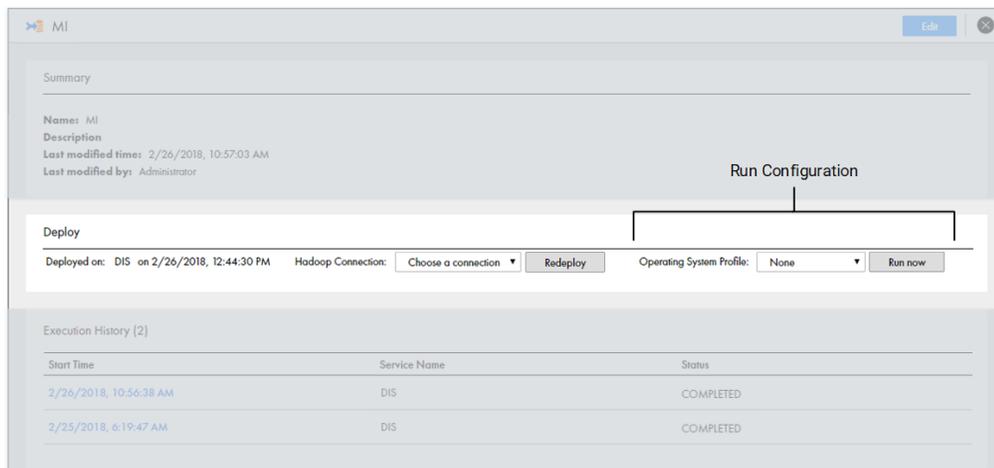
Running a Mass Ingestion Specification

Run a mass ingestion specification in the Mass Ingestion tool.

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

The **Overview** page appears.

The following image shows the run configurations in the **Deploy** view on the **Overview** page:



2. In the **Deploy** view, select an operating system profile. Default is None.
3. 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, 50](#)
- [Monitor a Mass Ingestion Specification in the Mass Ingestion Tool, 50](#)
- [Monitor a Mass Ingestion Specification in the Administrator Tool, 63](#)
- [Rules and Guidelines, 65](#)

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.

Monitor a Mass Ingestion Specification 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.

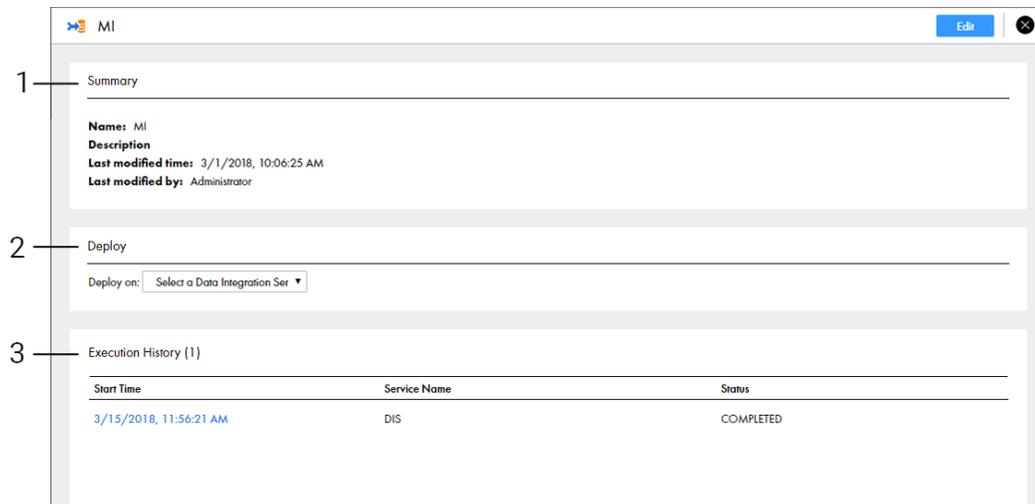
The Overview page provides general details on the specification, the deployment history, and the execution history. The Execution Statistics page provides detailed statistics on the ingestion job for a specification run instance.

Overview Page

On the Overview page, review details on a mass ingestion specification, the deployment and run options, and the execution history. The execution history lists the run instances for the mass ingestion specification. Each

run instance lists the start time of the ingestion job and the name of the Data Integration Service where the specification runs.

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.

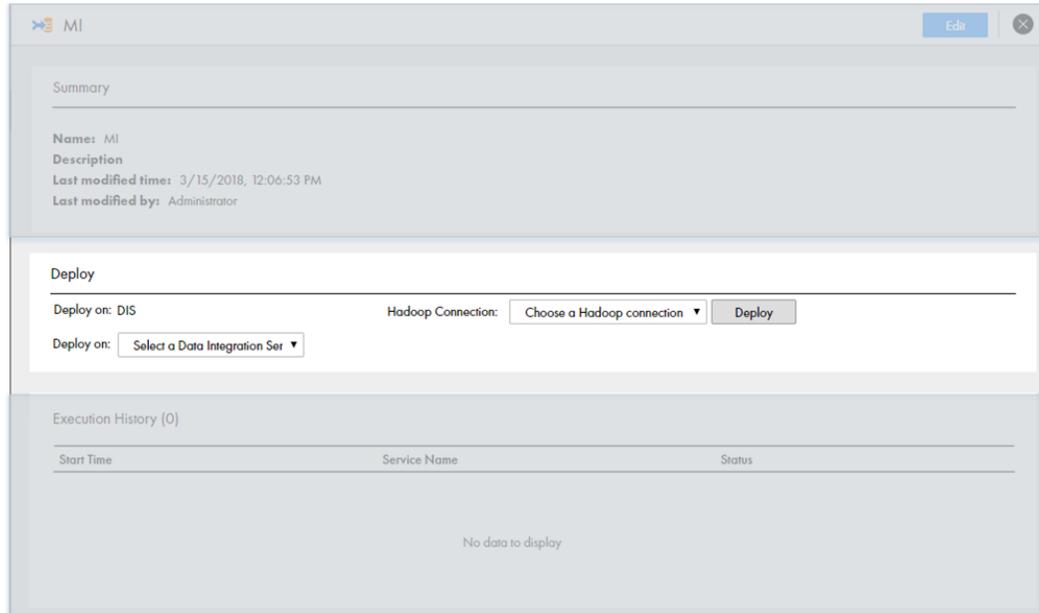
The following table describes the specification properties that you can review:

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.

The following image shows the options that you can select in the Deploy view before you deploy a mass ingestion specification:

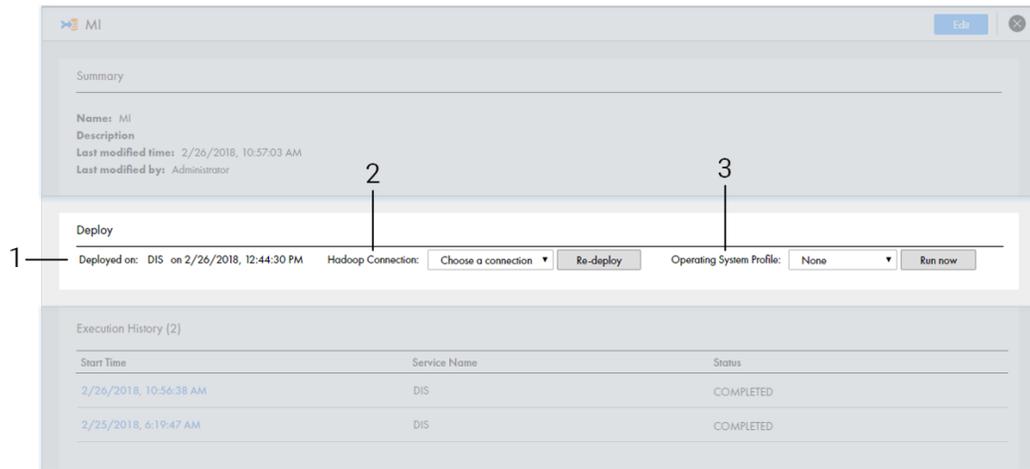


The screenshot shows a web interface for a mass ingestion (MI) specification. At the top, there is a title bar with an icon and the text 'MI', and a blue 'Edit' button with a close icon. Below this is a 'Summary' section with a horizontal line separator. The summary includes the following text: 'Name: MI', 'Description', 'Last modified time: 3/15/2018, 12:06:53 PM', and 'Last modified by: Administrator'. Below the summary is a 'Deploy' section, also separated by a horizontal line. It contains two rows of controls. The first row has 'Deploy on: DIS' on the left, 'Hadoop Connection: Choose a Hadoop connection' in the middle with a dropdown arrow, and a 'Deploy' button on the right. The second row has 'Deploy on: Select a Data Integration Set' with a dropdown arrow. Below the deploy section is an 'Execution History (0)' section, which contains a table with three columns: 'Start Time', 'Service Name', and 'Status'. The table is currently empty, with the text 'No data to display' centered below it.

To deploy the specification, you select a Data Integration Service. When you select a Data Integration Service, the option to select a Hadoop connection appears.

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. When you run the specification, you have the option to select 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. Select a Hadoop connection to redeploy the mass ingestion specification.
3. Operating System Profile. Select an operating system profile to run the mass ingestion 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.
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.
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 <code>Unknown</code> in the following circumstances: <ul style="list-style-type: none"> - 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 mass ingestion specification run instance.

The following image shows the Execution Statistics page:

The screenshot shows the 'Execution Statistics' page for a mass ingestion job. It includes a 'Job Properties' section with details like 'Mass Ingestion Specification: MI', 'Started By: Administrator', and 'Start Time: 2/6/2018, 4:10:21 PM'. The 'Ingestion Status' is shown as a donut chart with 112 completed (green), 16 failed (red), and 0 canceled (gray) items. The 'Ingestion Objects' table lists 131 source tables, all with a 'Completed' status. A detailed view for the 'TEST_PRODUCT' object shows it is 'COMPLETED' with 10,000 successful rows and 0 errors.

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 describes general properties of the specification run instance.

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.
Started By	The user who initiated the mass ingestion specification run.
Start Time	The time that the ingestion job is initiated.
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.

Status	Description
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.
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 <code>Unknown</code> in the following circumstances: <ul style="list-style-type: none"> - 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 specification . If you modify the specification and run the specification again, review the statistics on a unique Execution Statistics page for the 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
Ingestion Objects	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 might be Blaze, Spark, or Hive.
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. View the number of rows that were ingested successfully and the number of rows that contain errors.
Target	The statistics for the table rows that are ingested in the target. View the number of rows that were ingested successfully and the number of rows that contain errors.

Monitoring Execution Statistics

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

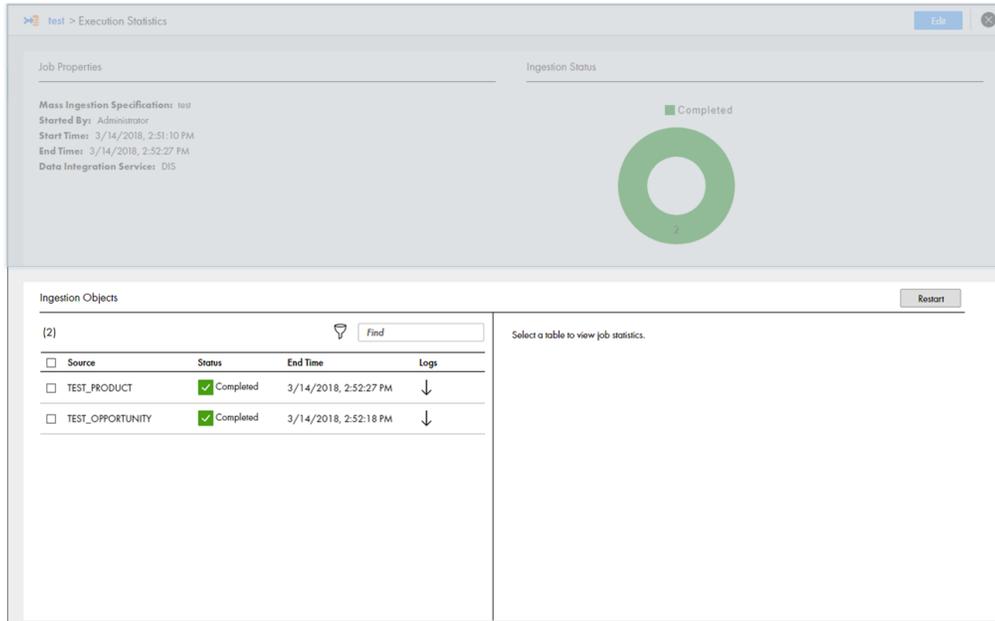
1. Browse to a mass ingestion specification in the Mass Ingestion tool.
The **Overview** page appears.
2. Select a run instance in the Execution History view to view the ingestion job statistics.
The **Execution Statistics** page appears for the run instance that you select.
3. On the **Execution Statistics** page, review the ingestion job.

- In the Ingestion Objects view, select a source table to view statistics for the job that ingests the source table.

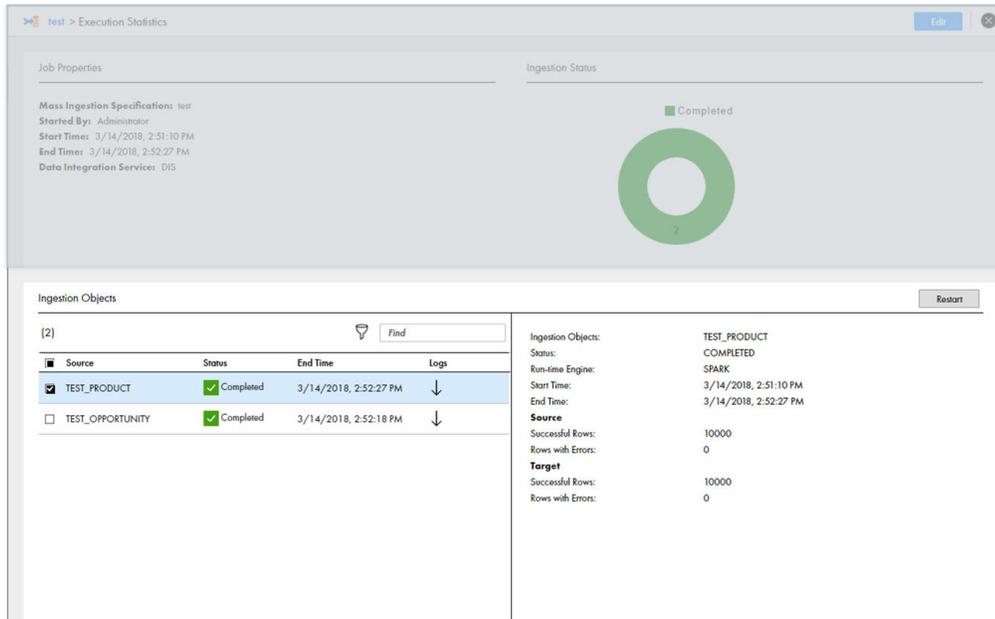
You can view ingestion statistics for one table at a time.

The statistics appear in the Ingestion Statistics view.

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:

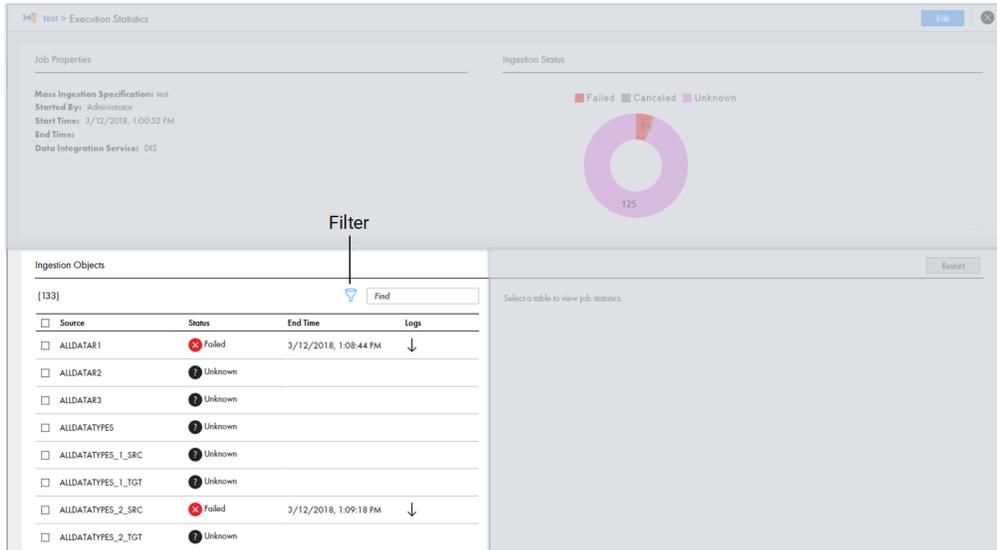


Filtering Ingestion Objects by Status

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.

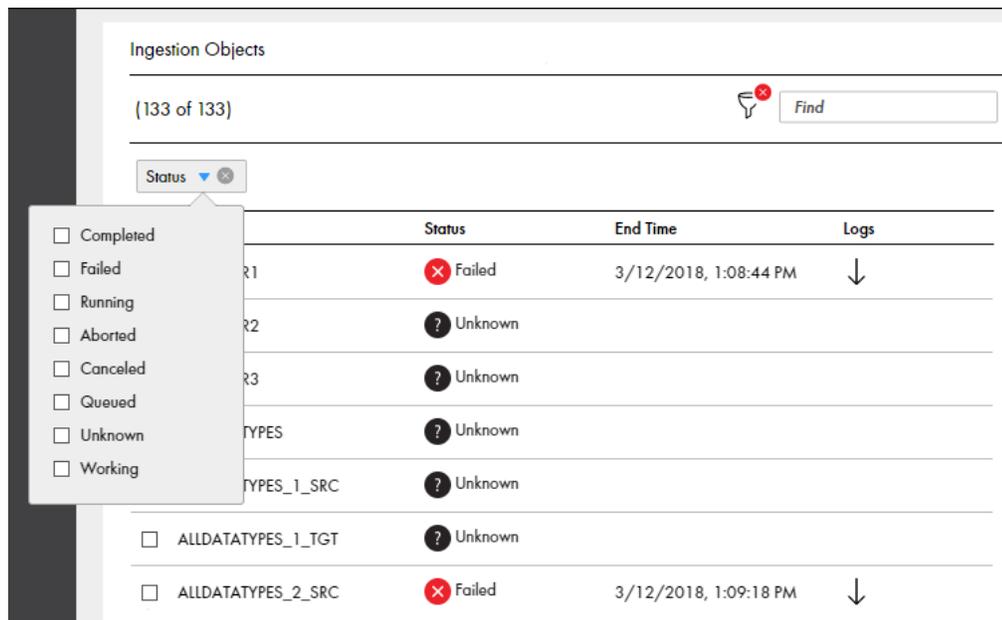
1. In the Ingestion Objects view on the Execution Statistics page, select the **Filter** icon.

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



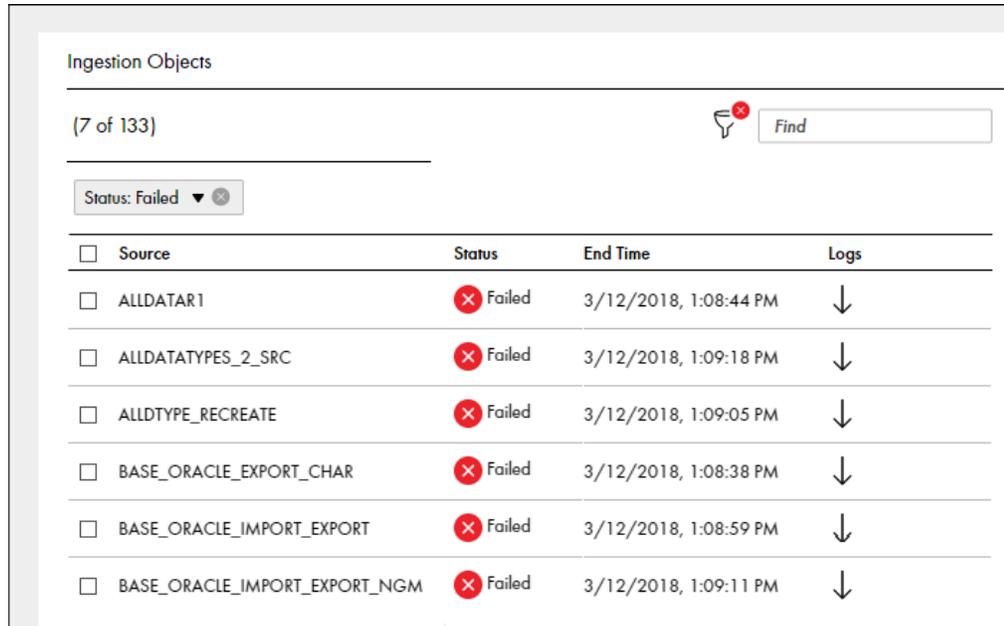
2. Select **Status** to filter the ingestion objects by the ingestion status.

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:



The screenshot shows the 'Ingestion Objects' interface. At the top, it displays '(7 of 133)' and a search bar with a 'Find' button. Below the search bar, there is a filter dropdown menu set to 'Status: Failed'. The main content is a table with the following columns: Source, Status, End Time, and Logs. The table lists seven failed ingestion objects, each with a checkbox, a red 'X' icon, the word 'Failed', the end time, and a downward arrow for logs.

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

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.

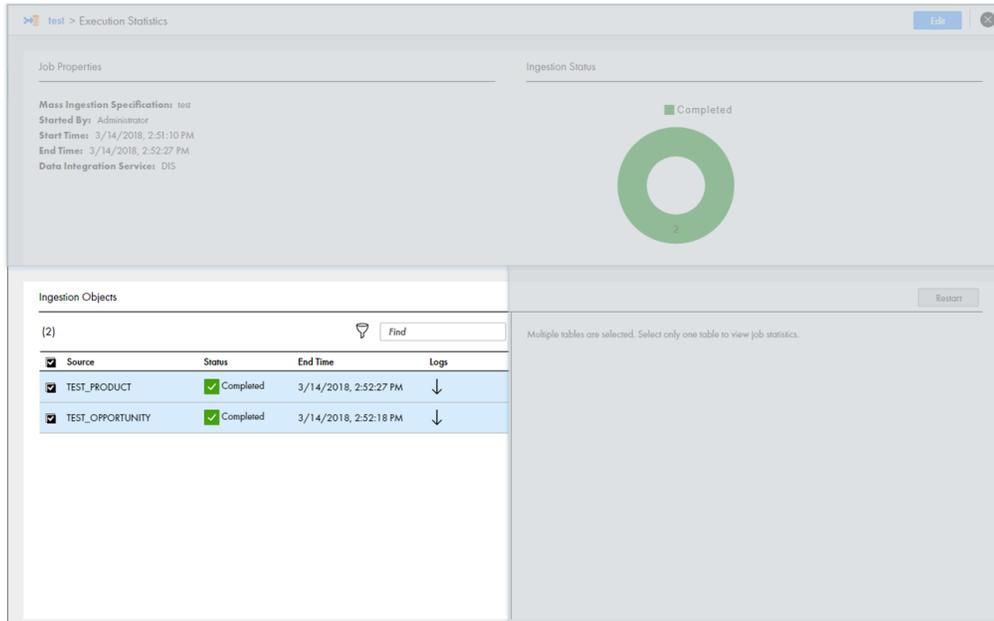
When you restart the ingestion jobs, you do not redeploy or rerun the entire mass ingestion specification. The ingestion status for the job refreshes and the status displays Working. The status continues to display Working until the Mass Ingestion Service fetches the ingestion status from the Data Integration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

Restarting Selected Ingestion Objects

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

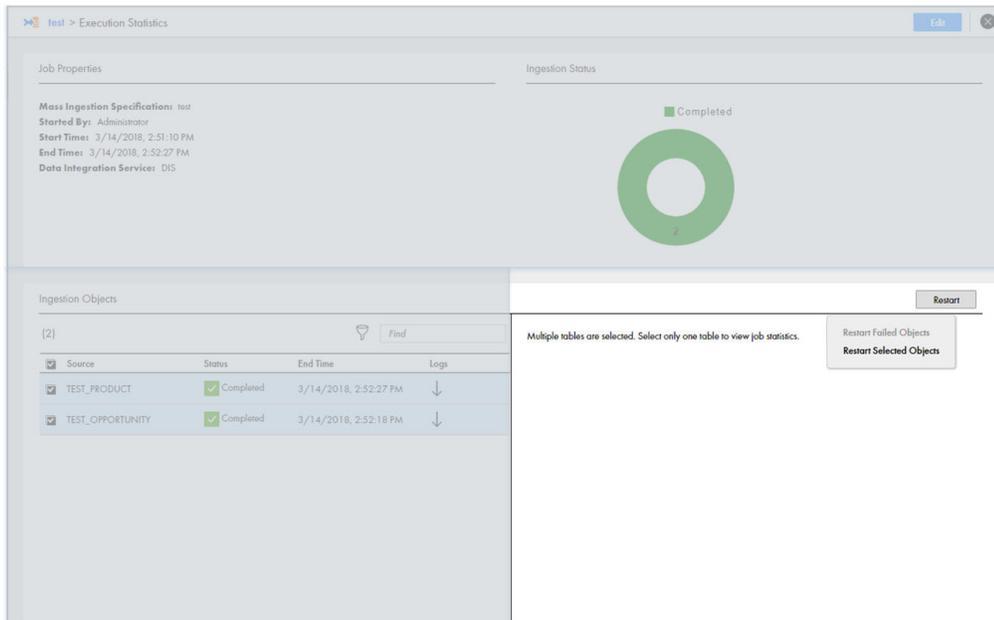
1. In the Ingestion Objects view on the Execution Statistics page, select the ingestion objects that you want to restart.

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.

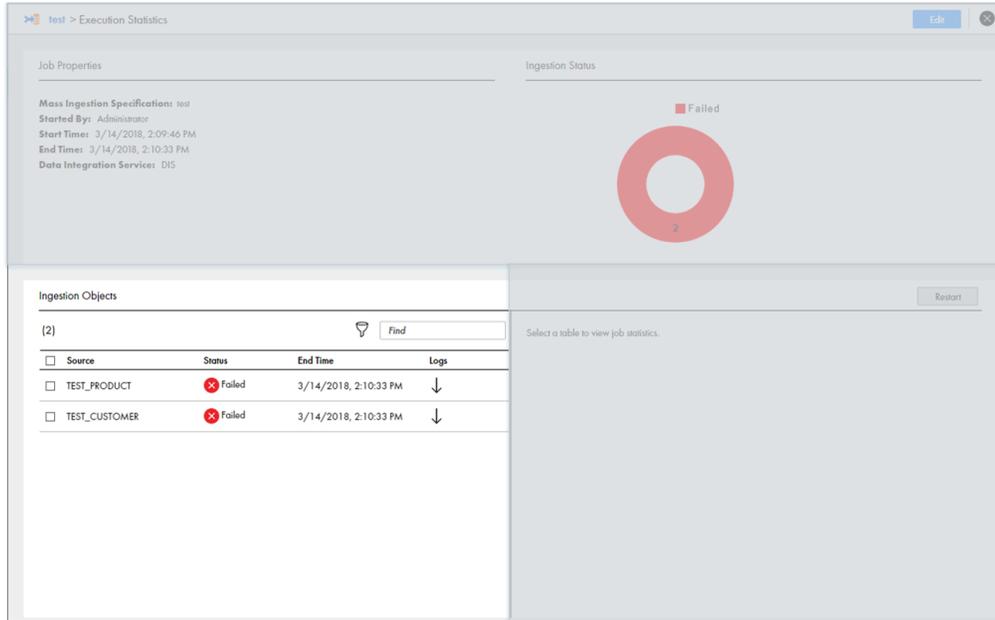
Restarting Failed Ingestion Objects

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**.

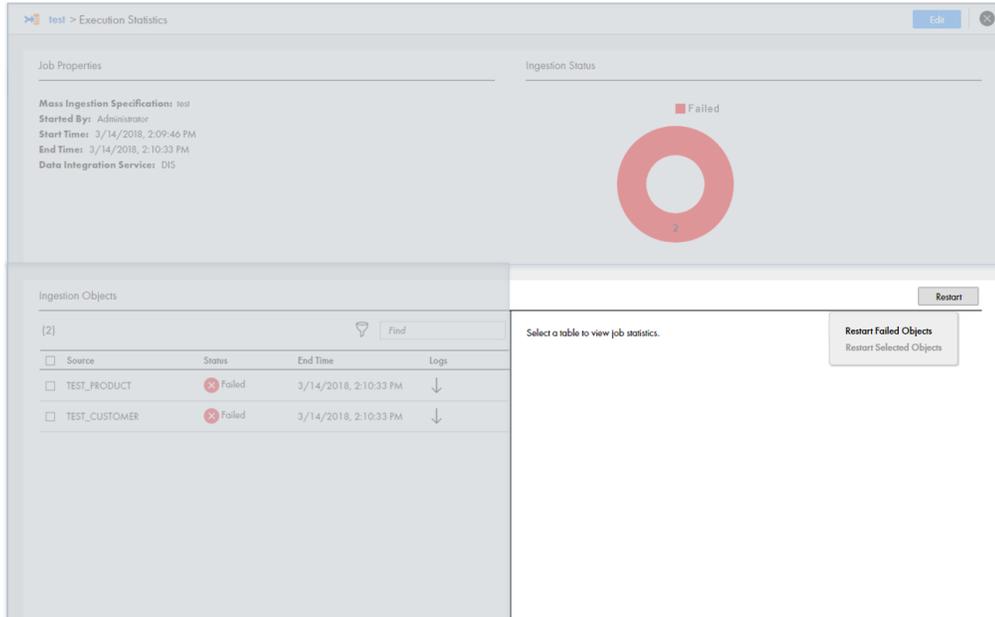
The following image shows several objects that failed:



The screenshot displays the 'test > Execution Statistics' page. On the left, under 'Job Properties', it shows 'Mass Ingestion Specifications: test', 'Started By: Administrator', 'Start Times: 3/14/2018, 2:09:46 PM', 'End Times: 3/14/2018, 2:10:33 PM', and 'Data Integration Service: DIS'. On the right, the 'Ingestion Status' section features a donut chart with a red segment representing 'Failed' objects, with the number '2' below it. Below the chart is a 'Restart' button. The 'Ingestion Objects' table lists two failed objects:

Source	Status	End Time	Logs
<input type="checkbox"/> TEST_PRODUCT	✘ Failed	3/14/2018, 2:10:33 PM	↓
<input type="checkbox"/> TEST_CUSTOMER	✘ Failed	3/14/2018, 2:10:33 PM	↓

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



This screenshot is identical to the previous one, but with a callout box highlighting the 'Restart Failed Objects' button. The callout box contains the text 'Restart Failed Objects' and 'Restart Selected Objects'.

2. Select **Restart Failed Objects**.

The ingestion job restarts for all failed and canceled objects.

Monitor a Mass Ingestion Specification 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

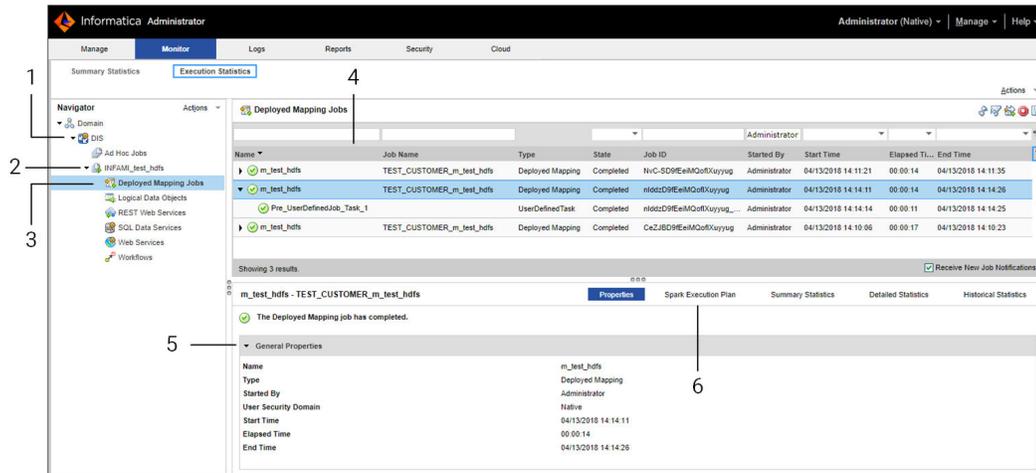
On the Execution Statistics tab, you can monitor the application and the ingestion mapping jobs that run on a Data Integration Service.

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 runs 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 Blaze, Spark, or Hive 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: <code>m_<mass ingestion specification name></code>
Job Name	The name of the source table that the mapping job ingests. The name appears in the following format: <code><source table name>_m_<mass ingestion specification name></code> 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 on execution statistics, see the *Informatica Administrator Guide*.

For more information on Blaze, Spark, and Hive execution plans, see the *Informatica Big Data Management 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 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.

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

For more information, see ["abortRun" on page 67](#) in the "infacmd mi Command Reference" appendix.

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**.

Rules and Guidelines

Some third-party rules and guidelines apply to processing data depending on the source or target connection that you use to ingest data.

Consider the following rules and guidelines when you ingest data:

- The source connection must be a JDBC connection. Use the JDBC connection to access a schema.
For example, to access an Oracle schema, configure a JDBC connection that uses an Oracle driver to connect to an Oracle database. You cannot use an Oracle connection.
- JDBC connections must be enabled with Sqoop connectivity.
- A source table cannot be ingested using a Sqoop connection if the table contains special characters in the table metadata.
- A source table cannot be ingested using a Sqoop connection if column names in the table contain spaces.
- Oracle long and number data types cannot be ingested using a Sqoop connection. To resolve the issue, append `NumberPrecisionScale=1` to the JDBC connection string.
- 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. To ingest timestamp data to an Avro file, the third-party Hive JDBC driver must use Hive version 1.1 or higher.

APPENDIX A

infacmd mi Command Reference

This appendix includes the following topics:

- [abortRun, 67](#)
- [createService, 68](#)
- [deploySpec, 71](#)
- [exportSpec, 72](#)
- [extendedRunStats, 73](#)
- [getSpecRunStats, 74](#)
- [listSpecRuns, 75](#)
- [listSpecs, 76](#)
- [restartMapping, 77](#)
- [runSpec, 78](#)

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
```

```

<-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. Port number for the Mass Ingestion Service. You cannot use HTTPS.
-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.
-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: <i>/parent_folder/child_folder</i> 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 <code>-dn</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN</code> . If you set a domain name with both methods, the <code>-dn</code> option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the <code>-un</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN_USER</code> . If you set a user name with both methods, the <code>-un</code> option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the <code>-pd</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN_PASSWORD</code> . If you set a password with both methods, the password set with the <code>-pd</code> 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 <code>-sdn</code> option or the environment variable <code>INFA_DEFAULT_SECURITY_DOMAIN</code> . If you set a security domain name with both methods, the <code>-sdn</code> 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 the mass ingestion specification is deployed.

Option	Argument	Description
-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 run the mass ingestion specification.

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 <code>-dn</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN</code> . If you set a domain name with both methods, the <code>-dn</code> option takes precedence.
-UserName -un	user_name	Required. User name to connect to the domain. You can set the user name with the <code>-un</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN_USER</code> . If you set a user name with both methods, the <code>-un</code> option takes precedence.
-Password -pd	password	Required. Password for the user name. The password is case sensitive. You can set a password with the <code>-pd</code> option or the environment variable <code>INFA_DEFAULT_DOMAIN_PASSWORD</code> . If you set a password with both methods, the password set with the <code>-pd</code> option takes precedence.

Option	Argument	Description
-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 an ingestion mapping job in the deployed mass ingestion specification. To get the extended statistics, you must specify the JobID for the mapping, and the mapping type. To find the JobID and the mapping type, get the ingestion run statistics using `infacmd mi getSpecRunStats`.

The extended run 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.

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
<-DISServiceName|-dsn> dis_service_name
<-JobId|-jid> mapping_job_id
<-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.
-DISServiceName -dis	data_integration_service	Required. Name of the Data Integration Service where the mass ingestion specification is deployed.
-JobId -jid	mapping_job_id	Required. The job ID of the ingestion mapping job in the mass ingestion specification. To find the job ID for the mapping, 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. To find the mapping type for the mapping, 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 <code>infacmd mi listSpecRuns</code> .

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 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

<-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.
-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

A

abortRun (infacmd mi) [67](#)

C

CreateService (infacmd mi) [68](#)

D

Data Integration Service

configuring [23](#)

JDK Home Directory [23](#)

deploying mi spec

application archive file [45](#)

deploySpec (infacmd mi) [71](#)

E

exportSpec

infacmd mi [72](#)

F

filter ingestion objects [59](#)

G

getSpecRunStats

infacmd mi [74](#)

I

infacmd mi

aborting a mass ingestion specification [67](#)

creating Mass Ingestion Service [68](#)

deploying mass ingestion specification [71](#)

deploying spec [72](#)

extendedRunStats [73](#)

getting the spec stats [74](#)

listing mi specs [76](#)

listSpecRuns [75](#)

restarting jobs [77](#)

running mi spec [78](#)

ingestion

monitoring [51](#)

L

listSpecs (infacmd mi) [76](#)

M

mass ingestion

administrator [15](#)

administrator role [15](#)

application services [13](#)

architecture [12](#)

assigning Administrator role [23](#)

assigning roles [23](#)

before you begin [19](#)

canceling job [65](#)

clients and tools [12](#)

configuring monitoring [21](#)

create [24](#)

data processing [65](#)

definition [25](#)

deploy

view [52](#)

diagram [13](#)

entire string [36](#)

execution history [53](#)

execution statistics [54](#), [63](#)

filter by [33](#)

filter clause [33](#)

guidelines [65](#)

ingestion objects [55](#)

ingestion statistics [56](#)

ingestion status [56](#)

job properties [55](#)

login [16](#)

mass parameters [31](#)

migrate [45](#)

monitor [50](#), [63](#)

monitoring [54](#)

overview [8](#)

pattern [36](#)

prepare [19](#)

regex [34](#)

regular expression [34](#)

replace columns [36](#)

replace criteria [36](#)

repositories [13](#)

restart [60](#)

restart failed [62](#)

restart ingestion [60](#)

restart ingestion jobs [60](#)

restart jobs [61](#), [62](#)

restart selected [61](#)

roles [15](#)

rules [65](#)

- mass ingestion (*continued*)
 - run
 - status [48](#)
 - run instances [48](#)
 - run statistics
 - accessing with infacmd mi [73](#)
 - source [26](#)
 - specification [25–27](#), [31](#), [37](#), [63](#)
 - summary [51](#)
 - summary statistics [63](#)
 - table parameters [37](#)
 - target [27](#), [30](#)
 - tool [15](#)
 - use case [9](#)
 - views [17](#)
- Mass Ingestion Service
 - assign to node [19](#)
 - creating [19](#), [68](#)
- mass ingestion specification
 - aborting [67](#)
 - compression codec [31](#)
 - create [40](#)
 - ddl query [28](#)
 - deploy [43](#)
 - filtering [59](#)
 - hdfs target
 - compression codec [31](#)
 - hive target
 - ddl query [28](#)
 - monitor
 - execution statistics [57](#)
 - run [49](#)
- mass ingestion tool
 - interface [17](#)
 - user interface [17](#)
- mi
 - admin role [15](#)
 - application services [13](#)
 - clients and tools [12](#)
 - migrate [45](#)
 - repositories [13](#)
 - roles [15](#)
 - use case [9](#)
- mi run
 - status [48](#)
- mi spec
 - canceling [65](#)
 - compression codec [31](#)
 - creating [24](#)
 - ddl query [28](#)
 - deploy [43](#)
 - deploying [42](#)
 - execution statistics [54](#)

- mi spec (*continued*)
 - hdfs target
 - compression codec [31](#)
 - HDFS target [30](#)
 - hive target
 - ddl query [28](#)
 - Hive target [27](#), [38](#), [40](#)
 - monitoring [50](#)
 - run statistics [73](#)
 - run stats [73](#)
 - running [47](#)
 - summary [51](#)
 - table parameters
 - Hive target [38](#), [40](#)
 - monitoring
 - preferences, configuring [22](#)

P

- preferences
 - monitoring [22](#)

R

- regex
 - in mi spec [34](#)
- regular expression
 - in mi spec [34](#)
- restartMapping (infacmd mi) [77](#)
- run summary
 - deployed mi spec [75](#)
- runSpec
 - infacmd mi [78](#)

S

- spec
 - deploying to an archive file [72](#)
- spec status
 - accessing with infacmd mi [74](#)
- specs deployed to a Data Integration Service
 - running [78](#)

V

- viewing mi spec
 - runs [48](#)