



Informatica® Mass Ingestion  
May 2024

# Mass Ingestion Applications

Informatica Mass Ingestion Mass Ingestion Applications  
May 2024

© Copyright Informatica LLC 2021, 2024

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, Informatica Cloud, Informatica Intelligent Cloud Services, PowerCenter, PowerExchange, and the Informatica logo 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](mailto: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: 2024-05-17

# Table of Contents

<b>Preface .....</b>	<b>6</b>
Informatica Resources. ....	6
Informatica Documentation. ....	6
Informatica Intelligent Cloud Services web site. ....	6
Informatica Intelligent Cloud Services Communities. ....	6
Informatica Intelligent Cloud Services Marketplace. ....	6
Data Integration connector documentation. ....	7
Informatica Knowledge Base. ....	7
Informatica Intelligent Cloud Services Trust Center. ....	7
Informatica Global Customer Support. ....	7
 <b>Chapter 1: Mass Ingestion Applications.....</b>	<b>8</b>
Use cases. ....	8
System requirements. ....	9
Mass Ingestion Applications architecture. ....	9
General limitations and guidelines. ....	10
Supported sources. ....	10
Guidelines for Google Analytics sources. ....	11
Guidelines for Marketo sources. ....	12
Guidelines for Microsoft Dynamics 365 sources. ....	13
Guidelines for NetSuite sources. ....	13
Guidelines for Oracle Fusion Cloud sources. ....	14
Guidelines for Salesforce sources. ....	16
Guidelines for Salesforce Marketing Cloud sources. ....	16
Guidelines for SAP ECC and SAP S4/HANA sources using the SAP ODP Extractor connector. .	17
Guidelines for SAP sources using the SAP Mass Ingestion connector. ....	19
Guidelines for ServiceNow sources. ....	32
Guidelines for Workday sources. ....	33
Guidelines for Zendesk sources. ....	34
Supported targets. ....	36
Guidelines for Amazon Redshift targets. ....	37
Guidelines for Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Fabric OneLake, and Oracle Cloud Object Storage targets. ....	38
Guidelines for Apache Kafka targets. ....	48
Guidelines for Databricks Delta targets. ....	48
Guidelines for Google BigQuery targets. ....	50
Guidelines for Microsoft Azure Synapse Analytics targets. ....	51
Guidelines for Microsoft Azure SQL Database targets. ....	52
Guidelines for Oracle targets. ....	53
Guidelines for PostgreSQL targets. ....	53

Guidelines for Snowflake targets. . . . .	54
Avro data types. . . . .	56
Handling source schema changes. . . . .	57
Ability to apply deletes as soft deletes on the target . . . . .	59
Apply an audit history of all source table change operations to target tables. . . . .	59
Configuring application ingestion tasks. . . . .	61
Before you begin. . . . .	61
Defining basic task information. . . . .	61
Configuring the source. . . . .	63
Configuring the target. . . . .	105
Configuring schedule and runtime options. . . . .	138
Deploying an application ingestion task. . . . .	141
Running an application ingestion job. . . . .	142
Stopping an application ingestion job . . . . .	142
Resuming an application ingestion job. . . . .	143
Restart and recovery for incremental load jobs. . . . .	143
Overriding schema drift options when resuming an application ingestion job. . . . .	144
Redeploying an application ingestion job. . . . .	145
Undeploying an application ingestion job. . . . .	146
Aborting an application ingestion job. . . . .	146
Resynchronizing source and target objects. . . . .	147
Default Data Type Mappings. . . . .	148
Adobe Analytics Source and Amazon Redshift Target. . . . .	148
Adobe Analytics Source and Google BigQuery Target. . . . .	149
Adobe Analytics Source and Microsoft Azure Synapse Analytics Target. . . . .	149
Adobe Analytics Source and Oracle Target. . . . .	150
Adobe Analytics Source and Snowflake Target. . . . .	150
NetSuite Source and Amazon Redshift Target. . . . .	151
NetSuite Source and Databricks Delta Target. . . . .	151
NetSuite Source and Google BigQuery Target. . . . .	152
NetSuite Source and Microsoft Azure Synapse Analytics Target. . . . .	153
NetSuite Source and Oracle Target. . . . .	153
NetSuite Source and Snowflake Target. . . . .	154
Salesforce Source and Amazon Redshift Target. . . . .	154
Salesforce Source and Databricks Delta Target. . . . .	155
Salesforce Source and Google BigQuery Target. . . . .	156
Salesforce Source and Microsoft Azure Synapse Analytics Target. . . . .	158
Salesforce Source and Microsoft Azure SQL Database Target. . . . .	159
Salesforce Source and Oracle Target. . . . .	160
Salesforce Source and PostgreSQL Target. . . . .	161
Salesforce Source and Snowflake Target. . . . .	162
Salesforce Marketing Cloud Source and Amazon Redshift Target. . . . .	163

Salesforce Marketing Cloud Source and Databricks Delta Target. . . . .	163
Salesforce Marketing Cloud Source and Google BigQuery Target. . . . .	164
Salesforce Marketing Cloud Source and Microsoft Azure Synapse Analytics Target. . . . .	164
Salesforce Marketing Cloud Source and Oracle Target. . . . .	165
Salesforce Marketing Cloud Source and Snowflake Target. . . . .	165
SAP Source and Amazon Redshift Target. . . . .	166
SAP Source and Databricks Delta Target. . . . .	166
SAP Source and Google BigQuery Target. . . . .	167
SAP Source and Microsoft Azure Synapse Analytics Target. . . . .	167
SAP Source and Oracle Target. . . . .	168
SAP Source and Snowflake Target. . . . .	169
SAP Mass Ingestion connection Source and Amazon Redshift Target. . . . .	170
SAP Mass Ingestion connection Source and Databricks Delta Target. . . . .	171
SAP Mass Ingestion connection Source and Google BigQuery Target. . . . .	173
SAP Mass Ingestion connection source and Microsoft Azure Synapse Analytics Target. . . . .	174
SAP Mass Ingestion connection Source and Oracle Target. . . . .	176
SAP Mass Ingestion connection Source and Snowflake Target. . . . .	177
SAP ODP Source and Amazon Redshift Target. . . . .	178
SAP ODP Source and Databricks Delta Target. . . . .	180
SAP ODP Source and Google BigQuery Target. . . . .	181
SAP ODP Source and Microsoft Azure Synapse Analytics Target. . . . .	182
SAP ODP Source and Oracle Target. . . . .	184
SAP ODP Source and Snowflake Target. . . . .	185
<b>Index. . . . .</b>	<b>187</b>

# Preface

Read *Mass Ingestion Applications* to learn how to configure application ingestion tasks in the Mass Ingestion service.

## Informatica Resources

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

### Informatica Documentation

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

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com).

### Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>. This site contains information about Informatica Cloud integration services.

### Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

### Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

<https://marketplace.informatica.com/>

## Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

## Informatica Knowledge Base

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

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at [KB\\_Feedback@informatica.com](mailto:KB_Feedback@informatica.com).

## Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, on the [Informatica Intelligent Cloud Services Status](#) page, click **SUBSCRIBE TO UPDATES**. You can choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

## Informatica Global Customer Support

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

To find online support resources on the Informatica Network, click **Contact Support** in the Informatica Intelligent Cloud Services Help menu to go to the **Cloud Support** page. The **Cloud Support** page includes system status information and community discussions. Log in to Informatica Network and click **Need Help** to find additional resources and to contact Informatica Global Customer Support through email.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

# CHAPTER 1

## Mass Ingestion Applications

Mass Ingestion Applications can transfer data from Software-as-a-Service (SaaS) and on-premise applications to cloud-based data lakes, data warehouses, and event streaming platforms.

The SaaS and on-premise applications used in your business or organization store large amounts of business-critical data on a daily basis. You can use Mass Ingestion Applications to transfer the data stored by your applications to targets that can handle large volumes of data. After you transfer the data to the target, you can consolidate the data and use it for various purposes, such as advanced data analytics and data warehousing.

Mass Ingestion Applications can perform the following types of load operations:

- *Initial load.* Loads source data read at a single point in time to a target. After the data is loaded, the ingestion job ends. You can use this load type to materialize a target to which incremental changes will be sent later.
- *Incremental load.* Loads data changes continuously or until the ingestion job is stopped or ends. The ingestion job loads the changes that have occurred since the last time it ran or from a specific start point. You can use this load type to keep data in the target up to date so that you can make informed decisions for your business or organization based on the latest data.
- *Initial and Incremental load.* Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source objects on a continuous basis.

For more information about the sources and targets supported for each load type, see [“Supported sources” on page 10](#) and [“Supported targets” on page 36](#).

## Use cases

You can use Mass Ingestion Applications to solve multiple business problems.

Following are some of the use cases of Mass Ingestion Applications:

- **Data warehousing:** Organize the data of SaaS and on-premise applications by transferring it to a cloud-based data warehouse system. After an initial batch load of data to the data warehouse, Mass Ingestion Applications can replicate data changes continuously from a source application to keep the data up to date in the data warehouse.
- **Advanced data analytics:** Consolidate your application data in data lakes and data warehouses for extensive analysis that helps in making informed business decisions.
- **Utilizing application data in other data processing applications:** Keep data lakes and data warehouses synchronized with SaaS and on-premise sources and provide up-to-date data to other applications for processing.

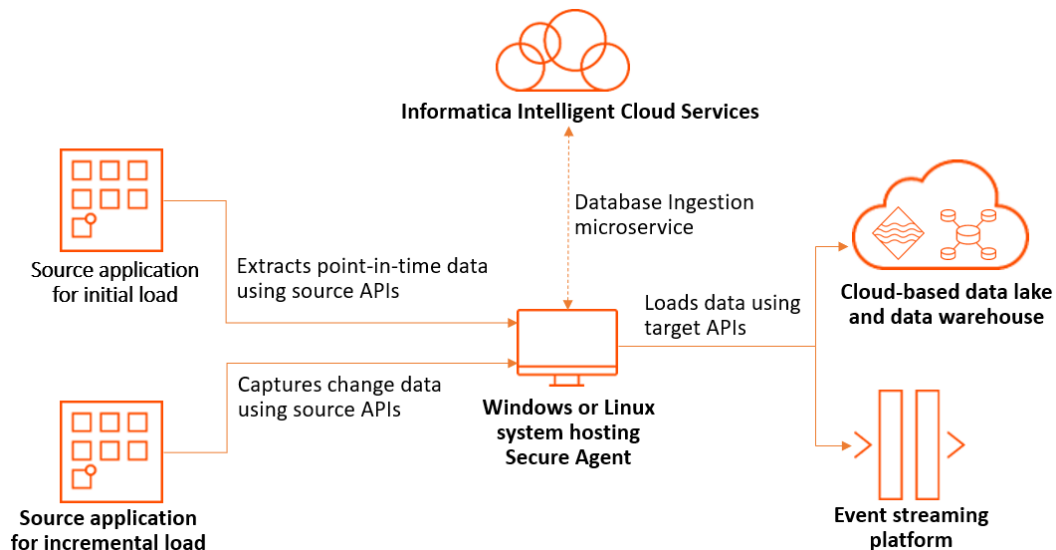
# System requirements

The following table lists the Mass Ingestion Applications minimum system requirements for the Secure Agent:

Component	Minimum requirement
Cores per CPU	8 minimum, 16 recommended if you need to process a large number of source objects in an initial load
Memory	32 GB
Disk space	5 GB per job, based on a row size of 2 KB

## Mass Ingestion Applications architecture

The following diagram shows the Mass Ingestion Applications architecture:



Mass Ingestion Applications requires the following components to run an application ingestion job:

- **Secure Agent:** Secure Agent is the program that runs tasks and enables secure communication between your organization and Informatica Intelligent Cloud Services. You must download the Secure Agent on all the systems from which you want to use Mass Ingestion Applications. When you run an application ingestion job, the metadata of the ingestion task is pushed to the Secure Agent which enables the ingestion job to process the data.
- **Database Ingestion service:** The Database Ingestion service is a microservice that is used by Secure Agent to run both application ingestion and database ingestion tasks. The Database Ingestion service is automatically downloaded to the system when you download the Secure Agent.
- **Informatica Intelligent Cloud Services interface:** Informatica Intelligent Cloud Services provides a web-based interface to create and deploy an application ingestion task. When you deploy the task, an executable ingestion job is created for the task.

- **Source and target APIs:** The application ingestion job uses the source APIs to retrieve the data from the source objects and the target APIs to load the data to the target. For an incremental load operation, the ingestion job identifies the changes that are made to the source objects after a specific date and time and retrieves the changes at intervals defined in the associated ingestion task.

## General limitations and guidelines

Before you configure application ingestion tasks for initial, incremental, or combined initial and incremental load operations, review the following limitations and guidelines:

- Mass Ingestion Applications does not support the Cloud Hosted Agent, a runtime environment on a cloud platform, or the serverless runtime environment. Use a local Secure Agent installation.
- If a Secure Agent in a runtime environment that is assigned to a application ingestion task terminates, you cannot undeploy the associated ingestion job, update the task to specify another runtime environment, and deploy the task again. In this situation, perform one of the following actions:
  - Assign a different Secure Agent to the runtime environment. Ensure that the new Secure Agent is running. Then restart the associated ingestion job.
  - Copy the task. In the task copy, specify another runtime environment that has an active Secure Agent. Then deploy the task and run the associated ingestion job.
- If you run multiple incremental load or combined initial and incremental load jobs against the same source, you cannot move capture processing of some source objects from one job to another job or merge the jobs into a single job, without resynchronizing the source and target objects being processed by the new job. Each job maintains a unique recovery checkpoint. If you try to move capture processing of an object from one job to another job, the correct recovery information for that table is not maintained, potentially causing data loss or duplicate events on the target.

## Supported sources

The sources that Mass Ingestion Applications support depend on whether the application ingestion tasks transfer a point-in-time snapshot of data in a batch initial load operation or load incremental change data from a specific start point.

The following table lists the source types that Mass Ingestion Applications support along with the types of load operations supported for each source type:

Source type	Supported load operations
Adobe Analytics	Initial load, incremental load, and combined initial and incremental load
Google Analytics	Initial load, incremental load, and combined initial and incremental load
Marketo	Initial load, incremental load, and combined initial and incremental load
Microsoft Dynamics 365	Initial load, incremental load, and combined initial and incremental load

Source type	Supported load operations
NetSuite	Initial load, incremental load, and combined initial and incremental load
Oracle Fusion Cloud Applications	REST: Initial load, incremental load, and combined initial and incremental load BICC: Initial load, incremental load, and combined initial and incremental load
Salesforce	Initial load, incremental load, and combined initial and incremental load
Salesforce Marketing Cloud	Initial load
SAP ERP Central Component (SAP ECC)	Initial load, incremental load, and combined initial and incremental load depending on the connection type.
SAP ODP	SAP ECC: Initial load, incremental load, and combined initial and incremental load SAP S4/HANA: Initial load, incremental load, and combined initial and incremental load
SAP Mass Ingestion	SAP ECC: Initial load, incremental load, and combined initial and incremental load for Oracle database. Certified on Snowflake target. SAP S4/HANA: Initial load and incremental load for HANA database. Certified on Snowflake target.
ServiceNow	Initial load, incremental load, and combined initial and incremental load
Workday	Initial load, incremental load, and combined initial and incremental load
Zendesk	Initial load, incremental load, and combined initial and incremental load

To determine the connectors to use for the source types, see *Connectors and Connections > Mass Ingestion Applications connectors*.

## Guidelines for Google Analytics sources

Consider the following guidelines when you use Google Analytics sources:

- The first time an incremental load job runs to capture change data for a Google Analytics source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- In Application ingestion initial load jobs, the Google Analytics report description is the input data. Each report contains a list of dimension and metrics that replicates to the target. There are 535 columns including dimensions and metrics. However, Google Analytics can fetch a maximum of 10 metrics and 9 dimensions in a single report. You can configure multiple reports to view all dimensions & metrics.
- Before you configure an application ingestion task, you must create the report description document and save it in a location that is accessible by the Secure Agent. You should have read-permission for the file.
- Google Analytics fetches data only for days and not for a given interval that is represented as a timestamp.

- Consider the following guidelines while updating the reports description document for application ingestion:
  - Don't change the order of the field while renaming it in the report. You must overwrite the new field name.
  - While adding new fields to the report, add the fields at the end of fields' list of the report.
  - To delete fields, remove the corresponding field names from the fields' list of the report.

## Guidelines for Marketo sources

Consider the following guidelines when you use Marketo sources:

- The first time an incremental load job runs to capture change data for a Marketo source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs configured for Marketo sources can replicate only insert and update operations performed on the source objects.
- Mass Ingestion Applications does not replicate the relationships defined between Marketo objects. The target does not contain the relationships that exists between the objects on the source.
- By default, Marketo instances are limited to 500 MB of downloaded data per day. Additionally, bulk API downloads are limited to a window of 31 day period.
- Some Marketo objects do not allow APIs to directly retrieve the data stored in them. To retrieve the data from such objects, APIs require certain filters as input parameters. The filters that the APIs use are fields of other Marketo objects. Mass Ingestion Applications does not replicate the data stored in such Marketo objects that are dependent on other objects.
- Application ingestion jobs do not propagate the data stored in custom objects that are not linked to any lead object. Additionally, application ingestion jobs do not propagate the data stored in custom object fields with the display name ID.
- By default, Mass Ingestion Applications can capture change data for the following Marketo objects:
  - Campaign
  - Email Templates
  - Folder
  - Landing Pages
  - Landing Page Templates
  - Program
  - Segmentations
  - Smart Campaign
  - Smart Lists
  - Snippets
  - Static Lists

**Note:** To capture change data for the Campaign, Email Templates, Folder, Landing Pages, Landing Page Templates, Segmentations, and Snippets objects, you must set the `marketo.include.additional.cdc.objects` custom property to TRUE. You can set the custom property on the **Source** tab of the application ingestion task wizard.

## Guidelines for Microsoft Dynamics 365 sources

Consider the following guidelines when you use Microsoft Dynamics 365 sources:

- The first time an incremental load job runs to capture change data for a Microsoft Dynamics 365 source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs for Microsoft Dynamics 365 sources don't capture the change data for objects without the `modifiedon` field.
- Incremental load jobs and combined initial and incremental load jobs for Microsoft Dynamics 365 sources don't detect and replicate the changes that are made to the source schema.
- Incremental load jobs and combined initial and incremental load jobs for Microsoft Dynamics 365 sources don't capture the DELETE data operation for a CDC cycle.
- If a source object contains locale ID and timezonecode values, Mass Ingestion Applications propagates the values to the target without converting them to the actual values that they represent.
- Mass Ingestion Applications does not propagate fields of the following data types:
  - CUSTOMERTYPE
  - LOOKUPTYPE
  - OWNERTYPE
  - PARTYLISTTYPE
  - VIRTUALTYPE
  - CALENDARRULESTYPE
  - UNKNOWN

## Guidelines for NetSuite sources

Consider the following guidelines when you use NetSuite sources:

- The first time an incremental load job runs to capture change data for a NetSuite source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs configured for NetSuite sources do not capture change data for the tables that do not contain a primary key.
- Incremental load jobs do not capture change data for the tables that do not contain at least one of the following timestamps:
  - The timestamp of when its records were created.
  - The timestamp of the last time its records were modified.

**Note:** For the tables that contain only the timestamp of when its records were created, incremental load jobs capture only the insert operations performed on the records.
- When an application ingestion job retrieves a document or image from the source, the job propagates the unique identifier of the file cabinet that contains the document or image instead of propagating the binary content of the file.
- Mass Ingestion Applications does not capture the deletion of records from the source tables.

- Incremental load jobs and combined initial and incremental load jobs configured for NetSuite sources do not detect and replicate the deletion and renaming of source fields.
- If a NetSuite Mass Ingestion connection configured for a NetSuite.com data source is already used in an application ingestion task and if the task is in any status other than Undeployed, you must not modify the connection to access a NetSuite2.com data source. Instead of modifying the existing connection, create a new connection for the NetSuite2.com data source and then configure a new application ingestion task using the new connection.
- Mass Ingestion Applications does not propagate data stored in the following NetSuite tables:
  - Account\_activity
  - Account\_period\_activity
  - Case\_types
  - Case\_origins
  - CompanyAddressbook
  - CompanyAddressbookEntityAddress
  - Deleted\_records
  - MAP\_wave\_custbody17
  - MAP\_wave\_custbody\_cseg4
  - MAP\_wave\_custbody\_csegcs\_multsel

## Guidelines for Oracle Fusion Cloud sources

Consider the following guidelines when you use Oracle Fusion Cloud sources:

- Mass Ingestion Applications can replicate data of the following Oracle Fusion Cloud Applications Suite modules and applications:

Module	Application
Enterprise Resource Planning (ERP)	Cloud Financials
	Procurement
	Project Management
	Risk Management and Compliance
Human Capital Management (HCM)	Human Resources
Industry	Higher Education
	Public Sector
	Digital Experience for Communications
	CX for Consumer Goods
	CX for Financial Services

Module	Application
Sales	Loyalty
	Partner Relationship Management
	Customer Data Management
	Sales Automation
	Subscription Management
Service	B2B Service
Supply Chain & Manufacturing (SCM)	AI Apps
	Cloud Service Logistics
	Fusion Cloud Inventory Management
	Maintenance
	Manufacturing
	Order Management
	Product Lifecycle Management
	Supply Chain Collaboration
	Supply Chain Planning

- Mass Ingestion Applications uses Business Intelligence Cloud Connector (BICC) or REST approach to transfer data from the source to the target.
- Incremental load jobs and combined initial load and incremental load jobs configured for Oracle Fusion Cloud sources do not capture and replicate the changes made to the source schema.
- Enable the **Support Incremental** option in the BICC console for the datastore to support the incremental load and combined initial and incremental load types. To enable the **Support Incremental** option, at least one of the columns in the datastore must be marked with **Incremental Filter** for the **TIMESTAMP** data type.
- By default, incremental load jobs capture change data only for the objects that contain the LASTUPDATEDATE or CREATIONDATE field. However, some objects in Oracle Fusion Cloud (REST) contain only the CREATIONDATE field instead of the LASTUPDATEDATE field. To capture the change data for such objects that contain only the CREATIONDATE field, set the `oraclefusion.cdc.allow.creationdate.objects` custom property to **true** on the **Source** tab of the application ingestion task wizard. If you set the custom property to **true** for a deployed task, make sure that you undeploy the job, and then deploy the task again for the change to take effect. If an object has only the CREATIONDATE field, only insertions are recorded, and updates and deletions are ignored as the LASTUPDATEDATE field isn't available.

## Guidelines for Salesforce sources

Consider the following guidelines when you use Salesforce sources:

- The first time an incremental load job runs to capture change data for a Salesforce source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- By default, incremental load jobs capture change data only for the objects that contain the `lastmodifieddate` or `systemModstamp` field. However, some objects in Salesforce contain only `CreatedDate` field instead of `lastmodifieddate` and `systemModstamp` fields. To capture the change data for such objects that contain only the `CreatedDate` field, set the `salesforce.cdc.allow.createddate.objects` custom property to true on the **Source** tab of the application ingestion task wizard. If you set the custom property to true for a deployed task, make sure that you undeploy the job and then deploy the task again for the change to take effect.
- If you want to replicate Data.com objects, your Salesforce account must be assigned with the appropriate Data.com license.
- Mass Ingestion Applications does not capture the relationships defined between Salesforce objects. The target does not contain the relationships between the source objects.
- If an object contains a compound field, Mass Ingestion Applications propagates the component fields that make up the compound field instead of propagating the compound field. For example, if the compound field `geolocation` is a combination of fields `geolocation_longitude` and `geolocation_latitude`, Mass Ingestion Applications separately loads the data stored in the `geolocation_longitude` and `geolocation_latitude` fields to the target instead of propagating the compound field `geolocation`.
- The masking configuration of source fields is retained on the target.
- All incremental load activities use only the standard REST API.
- The lookup fields are represented on the target by their unique alphanumeric identifiers.
- When a new field with a default value of the Function data type is added to an existing source object, Mass Ingestion Applications does not propagate the default value to any existing row on the target. However, when new rows are added to the object, the incremental jobs propagate the values stored in the field for the newly added rows.
- Application ingestion jobs do not replicate the data stored in big objects that are not indexed on the Salesforce platform. Before you deploy an application ingestion task for a Salesforce source, ensure that you index all the big objects which are selected by the object selection rules in the task.
- If a source field of Decimal data type contains a value whose scale exceeds the maximum limit defined for the field, application ingestion jobs might fail to propagate such fields. However, you can trim the scale of the value to the defined limit while propagating the field. You can trim the scale of the value by setting the custom property `salesforce.decimal.trim.scale` to true.

## Guidelines for Salesforce Marketing Cloud sources

Consider the following guidelines when you use Salesforce Marketing Cloud sources:

- Mass Ingestion Applications can ingest the data stored in the data extension objects of Salesforce Marketing Cloud.
- You can configure only initial load jobs for Salesforce Marketing Cloud sources.
- When an application ingestion job replicates a data extension object on the target, the job prepends `DATA_EXTENSION__` to the name of the object. For example, if an application ingestion job replicates the data extension object `Address`, the name of the object on the target is `DATA_EXTENSION__Address`.

- An application ingestion initial load job with Salesforce Marketing Cloud source and Microsoft Azure Synapse Analytics target might fail to deploy for some objects.

## Guidelines for SAP ECC and SAP S4/HANA sources using the SAP ODP Extractor connector

Consider the following guidelines when you use SAP ECC or SAP S4/HANA sources with SAP ODP Extractor connector:

- When you use an SAP source for the first time, perform the following steps before you configure an application ingestion task for the source:
  1. Verify that the appropriate SAP Notes are available in the SAP server. For more information, see the *SAP Connector Guide*.
  2. Download the SAP Java Connector (SAP JCo) library from the [SAP](#) website.
  3. Copy the native and JAR files from the downloaded SAP JCo library to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\infa.odp\`
  4. Copy the `sap-adapter-common.jar` file from `<Secure_Agent>\downloads\package-ICSAGENT_<version>\package\ICS\main\bin\rdtm\javalib\sap` to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\infa.odp\`
  5. Restart the Secure Agent.
- Before you configure an application ingestion task for an SAP source, ensure that you release the data sources from which you want to transfer data. You can use the following functions to release the required data sources:
  - BS\_ANLY\_DS\_RELEASE\_ODP
  - RODPS\_OS\_EXPOSE
- The first time an incremental load job runs to capture change data for an SAP source, the job retrieves and loads the change records from the latest position in the data stream that contains source data. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs for SAP sources do not capture the change data for data sources that do not have a primary key. Additionally, the incremental load jobs do not capture change data for the following SAP data sources:
  - The data sources for which SAP does not support delta.
  - The data sources that are configured to provide change data in the form of additive images.
- When an application ingestion task encounters NULL values in the SAP table source column, it does not replicate NULL with blank data to the target table.
- Mass Ingestion Applications does not detect and propagate schema changes in SAP sources.
- Mass Ingestion Applications does not propagate fields of the following obsolete SAP data types:
  - D16S
  - D34S
  - PREC
  - VARC

Additionally, application ingestion jobs do not propagate fields of the RSTR data type.

## Configuring SAP User Authorization

Configure the SAP user account to process SAP table data.

The following table describes the required authorization to read from SAP tables:

Read Object Name	Required Authorization
S_BTCH_JOB	DELE, LIST, PLAN, SHOW. Set Job Operation to RELE.
S_PROGRAM	BTCSUBMIT, SUBMIT
S_RFC	SYST, SDTX, SDIFRUNTIME, /INFADI/TBLRDR, RFC1
S_TABU_DIS / S_TABU_NUM	Provide SAP table name from which you want to read data.

## Installing SAP ODP Connector Transport Files

Install the SAP ODP Extractor transport files on the SAP machines that you want to access. Before you install the transports on your production system, install and test the transports in a development system.

Verify that you install the latest transport files to extract data from the SAP ODP object.

Install the following data file and cofile to read data from the SAP ODP object:

Data and Cofile Names	Transport Request	Functionality
- K900426.IN7 - R900426.IN7	IN7K900426	Install the transports only when you want to read from an SAP ODP object that supports hierarchy. You can use the SAP ODP Extractor Connector without installing the SAP ODP Extractor transport files for objects that do not support hierarchy.

## Installing Transport Files

Install the latest transport files from a Secure Agent directory to read from a Unicode SAP system. The transport files are for SAP version ERP 6.0 EHP7 system or later.

- Find the transport files in the following directory on the Secure Agent machine:  
`<Informatica Secure Agent installation directory>\downloads\package-SAPODP.<Latest version>\package\sapodp\sap-transport`
- Copy the cofile transport file to the `Cofile` directory in the SAP transport management directory on each SAP machine that you want to access.  
The cofile transport file uses the following naming convention: `<number>.<sap system>`.
- Copy the data transport file to the `Data` directory in the SAP transport management directory on each SAP machine that you want to access.  
The data transport file uses the following naming convention: `<number>.<sap-system>`.
- To import the transports to SAP, in the STMS, click **Extras > Other Requests > Add** and add the transport request to the system queue.
- In the **Add Transport Request to Import Queue** dialog box, enter the request number for the cofile transport.

The request number inverts the order of the renamed cofile as follows: <sap-system><number>.

6. In the Request area of the import queue, select the transport request number that you added, and click **Import**.
7. If you are upgrading from a previous version of the Informatica Transports, select the **Overwrite Originals** option.

## Guidelines for SAP sources using the SAP Mass Ingestion connector

The SAP Mass Ingestion connector supports the following databases:

- SAP ECC (Oracle database)
- SAP S/4 HANA (Trigger based)

### Source preparation

To use Oracle sources in application ingestion tasks, first prepare the source database and review the usage considerations

- Define the following system environment variables on the Linux or Windows system where the Secure Agent runs:
  - ORACLE\_HOME environment variable. Points to the Oracle client installation directory, one level up from the bin directory on Windows or lib directory on Linux. This environment variable is not required. However, if you do not define it, you'll need to specify the full Oracle client installation path when you define other environment variables or Secure Agent properties that specify a path to a subdirectory.
  - TNS\_ADMIN environment variable. If you specify a TNS name in the **Database Connect String** property of the Oracle Database Ingestion connection properties, use this environment variable to point to the directory location of the tnsnames.ora file when the file is not in the default \$ORACLE\_HOME/network/admin directory. The tnsnames.ora file, along with the Oracle Call Interface (OCI), is used to communicate with the Oracle source database.

**Note:** In Administrator, you can set the ociPath property for the Database Ingestion agent service (DBMI agent) to point to the OCI library that contains the oci.dll or libclntsh.so file. The OCI library is used by application ingestion CDC tasks to connect to Oracle. By default, Oracle uses the ociPath value of \$ORACLE\_HOME/lib on Linux or %ORACLE\_HOME%\bin on Windows.

- Make sure the Mass Ingestion Applications user has the Oracle privileges that are required for the application ingestion load type to be performed.  
For more information, see ["Oracle privileges" on page 24](#).
- Application ingestion jobs that use log-based CDC require read access to Oracle online and archive redo logs to read incremental change data. If the redo logs are remote from the on-premises system where the Secure Agent runs, make sure that read access to the logs is provided, for example, by using Oracle Automatic Storage Management (ASM), mounting the logs to a network file system (NFS).
- If you use Oracle 11.2.04, set the Oracle COMPATIBLE initialization parameter to 11.2.04 to ensure that Oracle has all of the most current Redo Logs fixes for that release.
- Ensure that the Oracle Database Client or Instant Client is installed and configured on the Secure Agent server for the Secure Agent to communicate with Oracle. If you do not already have an Oracle client installed, you can download a client and access installation information from the Oracle web site, or ask your Oracle DBA to download and configure an Oracle client.

- For incremental load or combined initial and incremental load operations that use log-based CDC, perform the following prerequisite tasks in Oracle:

- Enable ARCHIVELOG mode for the Oracle database. If the database is not in an Amazon RDS environment, issue the following SQL statements:

```
SHUTDOWN IMMEDIATE;
STARTUP MOUNT;
ALTER DATABASE ARCHIVELOG;
ALTER DATABASE OPEN;
SHUTDOWN IMMEDIATE;
STARTUP;
```

- Enable Oracle minimal global supplemental logging on the source database.

## Usage considerations

- For incremental load operations with Oracle sources, Mass Ingestion Applications provides alternative capture methods for capturing change data from the source and applying the data to a target. Mass Ingestion Applications reads data changes from Oracle redo logs. This method requires users to have extended privileges.
- Because Mass Ingestion Applications expects each source table row to be unique, Informatica recommends that each source table have a primary key. Mass Ingestion Applications does not honor unique indexes in place of a primary key. If no primary key is specified, Mass Ingestion Applications treats all columns as if they are part of the primary key. Exception: For Oracle query-based CDC, a primary key is required in each source table.
- If Oracle source CHAR or VARCHAR columns contain nulls, the application ingestion job does not delimit the null values with double-quotation (") marks or any other delimiter when writing data to a Amazon S3, Flat File, Microsoft Azure Data Lake, or Microsoft Azure Synapse Analytics target.
- Mass Ingestion Applications can process data across a RESETLOGS boundary. To avoid the source and targets becoming out of sync, Informatica recommends that you stop capture processing before performing a RESETLOGS and then restart capture processing after the RESETLOGS event. Otherwise, the capture process might send data to the target that is subsequently reverted by the RESETLOGS event, causing the source and target to become out of sync.
- Alternative strategies for accessing the Oracle redo logs are available.
- If an application ingestion incremental load or combined initial and incremental load task contains an Oracle source table name or one or more column names that are longer than 30 characters, Oracle suppresses supplemental logging for the entire table, including primary keys and foreign keys. As a result, most operations on the table fail. This problem is caused by an Oracle restriction. In this situation, exclude the table from capture processing or rename the long table and column names to names of 30 characters or less.
- Do not use custom data type mapping rules to map an Oracle source RAW column to a target CHAR or VARCHAR column. Otherwise, the deployment of the application ingestion task might fail.
- Mass Ingestion Applications does not support invisible columns in Oracle source columns, regardless of the target type. For these columns, application ingestion incremental load jobs and combined initial and incremental load jobs propagate nulls to the corresponding target columns.
- Application ingestion incremental load and combined initial and incremental load jobs that have an Oracle source and use a Secure Agent group with multiple agents can switch over to another agent in the group when the active agent stops running, subject to following limitations:
  - The job cannot have a Kafka target.
  - The job cannot have persistent storage enabled.
  - The job cannot use the query-based CDC method to capture changes by querying a timestamp column.

- You must Stop and Resume the job for the switch to occur.
- Application ingestion combined initial and incremental load jobs that have Oracle sources can read changes from copies of the archive redo logs. You must set the **Reader Mode** property to ARCHIVECOPY in the Oracle Database Ingestion connection properties and also set the source custom property `pxw.cdcreader.oracle.reader.additional` with the `dir` and `file` parameters. The `dir` parameter points to the name of the base directory that the CDC log reader scans for the archive log copies, and the `file` parameter specifies a mask that is used to filter the log copies.

### Guidelines for SAP ECC (Oracle database) sources with the SAP Mass Ingestion connector

Consider the following guidelines when you use SAP ECC sources with the SAP Mass Ingestion connector:

- An SAP ECC (Oracle database) source with an SAP Mass Ingestion connector supports database views (initial load), transparent tables, pool tables, and cluster tables.
- When you use an SAP ECC source with an SAP Mass Ingestion connector for the first time, perform the following steps before you configure an application ingestion task for the source:
  1. You must install the latest version of the SAP transport file on the SAP machine you want to access. For more information about accessing the SAP transport file, contact Informatica Global Customer Support.

The following table describes the transport on the SAP development system:

Data and Cofile Names	Data and Cofile Names	Transport Request	Functionality
R<number>.EP6 K<number>.EP6	Required	EP6K<number>	Install the transports to read from SAP transparent tables, cluster tables, pool tables, and views.

For more information about installing the transport files, see [“Installing transport files” on page 24](#).

2. Verify that the appropriate SAP Notes are available in the SAP server. For more information, see the *SAP Connector Guide*.
  3. Download the SAP Java Connector (SAP JCo) library from the [SAP](#) website.
  4. Copy the native and JAR files from the downloaded SAP JCo library to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\informatica.sapami\`
  5. Restart the Secure Agent.
- Mass Ingestion Applications does not propagate fields of the following obsolete SAP data types:
    - D16S
    - D34S
    - PREC
    - VARC

Additionally, application ingestion jobs do not propagate fields of the RSTR data type.

- An SAP ECC source with an SAP Mass Ingestion connector doesn't support schema drift.

### Guidelines for SAP S/4 HANA sources with the SAP Mass Ingestion connector

Consider the following guidelines when you use SAP S/4 HANA sources with the SAP Mass Ingestion connector:

- An SAP S/4 HANA (Trigger based) source with an SAP Mass Ingestion connector supports database views (initial load) and transparent tables.

- When you use an SAP S/4 HANA source with an SAP Mass Ingestion connector for the first time, perform the following steps before you configure an application ingestion task for the source:

1. You must install the latest version of the SAP transport file on the SAP machine that you want to access.  
For more information about accessing the SAP transport file, contact Informatica Global Customer Support.

The following table describes the transport on the SAP development system:

Data and Cofile Names	Data and Cofile Names	Transport Request	Functionality
R<number>.EP6 K<number>.EP6	Required	EP6K<number>	Install the transports to read from SAP transparent tables, cluster tables, pool tables, and views.

For more information about installing the transport files, see [“Installing transport files” on page 24](#).

2. Verify that the appropriate SAP Notes are available in the SAP server. For more information, see the *SAP Connector Guide*.
3. Download the SAP Java Connector (SAP JCo) library from the [SAP](#) website.
4. Copy the native and JAR files from the downloaded SAP JCo library to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\informatica.sapami\`
5. The SAP Mass Ingestion connector uses JDBC to connect to SAP S/4 HANA to read data and metadata and to test connection properties. You must download the SAP HANA JDBC driver file, `ngdbc.jar`, and copy it to a specific subdirectory of the Secure Agent installation directory on the machine where the Secure Agent runs.
  - a. Download the SAP HANA JDBC driver jar file, `ngdbc.jar`, to the Linux or Windows machine where the Secure Agent runs.  
Verify that you download the most recent version of the file. If you encounter issues with downloading the file, contact SAP Customer Support.
  - b. Copy the `ngdbc.jar` file to the following directory:  
`<Secure Agent installation directory>\ext\connectors\thirdparty\informatica.hanami`
  - c. Restart the Secure Agent.
6. Create a Mass Ingestion Database user. Connect to the source database as a user with admin privilege and execute the following statement:  

```
CREATE USER dbmi_user password "<password>" NO FORCE_FIRST_PASSWORD_CHANGE;
```

 This statement creates a user in the database with the default permissions, which permit basic data dictionary views to be read and the required CDC objects to be created in the user's own schema.
7. To deploy and run a database ingestion task that includes an SAP S/4 HANA source, the source connection must specify a Mass Ingestion Databases user (`dbmi_user`) who has the privileges to read metadata and other information from the following system views:
  - `SYS.M_DATABASE` - Used to fetch the database version.
  - `SYS.M_CS_PARTITIONS` - Used to identify if a table is partitioned. (Not applicable for SAP HANA Cloud)
  - `SYS.SCHEMAS` - Used to fetch the list of schemas for a database.

- SYS.TABLES - Used to fetch a list of table names for a schema.
- SYS.TABLE\_COLUMNS - Used to fetch column metadata for a table.
- SYS.INDEXES - Used to fetch index information for a table.
- SYS.INDEX\_COLUMNS - Used to fetch index information for a table.

8. For incremental load jobs, grant the following privileges:

- For triggers to write rows to the PKLOG and shadow \_CDC tables of thedbmi\_user, grant INSERT access on thedbmi\_user's schema to the user that owns the schema of the source tables (schema\_user):

```
GRANT INSERT ON SCHEMA dbmi_user TO schema_user;
```

- To capture change data from source tables in the schema\_user's schema by using triggers, execute one of the following statements:

```
GRANT TRIGGER ON SCHEMA schema_user TO dbmi_user;
```

This statement grants trigger access to all tables in the schema.

- or -

```
GRANT TRIGGER ON database.table_name TO dbmi_user;
```

This statement grants trigger access to a specific source table. Use this statement when you want to capture data from just a few selected tables. Repeat the grant for each source table of CDC interest.

9. For initial load jobs, execute one of the following grant statements to read data from source tables:

```
GRANT SELECT ON SCHEMA schema_user TO dbmi_user;
```

This statement grants SELECT access to all tables in the schema.

- or -

```
GRANT SELECT ON database.table_name TO dbmi_user;
```

This statement grants SELECT access on a specific source table. Repeat this grant for each source table from which you want to read data.

10. For SAP S/4 HANA sources, the connection requires an SAP HANA JDBC custom connection property setting for encryption. Enter the following properties in the **Advanced Connection Properties** field in the SAP Mass Ingestion connection properties:

```
encrypt=true&validateCertificate=false
```

- Mass Ingestion Applications does not propagate fields of the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC
- DATN
- TIMN
- DECFLOAT\_16
- DECFLOAT\_34
- UTCLONG
- STRG datatypes (For incremental load)

Additionally, application ingestion jobs do not propagate fields of the RSTR data type.

- An SAP ECC source with an SAP Mass Ingestion connector doesn't support schema drift.

For more information about configuring SAP user authorization, see [“Configuring SAP User Authorization” on page 18](#).

## Installing transport files

Install the latest transport files from a Secure Agent directory to read from a Unicode SAP system. The transport files are for SAP version ECC 5.0 or later.

1. Find the transport files in the following directory on the Secure Agent machine:  
`<Secure_Agent>\downloads\package-SAPAmi.<version>\package\sapami\SAPTableReader\`
2. Copy the cofile transport file to the `Cofile` directory in the SAP transport management directory on each SAP machine that you want to access.  
  
The cofile transport file uses the following naming convention: `K<number>.EP6`.
3. Copy the data transport file to the `Data` directory in the SAP transport management directory on each SAP machine that you want to access.  
  
The data transport file uses the following naming convention: `R<number>.EP6`.
4. To import the transports to SAP, in the STMS, click **Extras > Other Requests > Add** and add the transport request to the system queue.
5. In the **Add Transport Request to Import Queue** dialog box, enter the request number for the cofile transport.  
  
The request number inverts the order of the renamed cofile as follows: `EP6K<number>`.  
For example, for a cofile transport file renamed as `K900215.EP6`, enter the request number as `EP6K900215`.
6. In the Request area of the import queue, select the transport request number that you added, and click **Import**.
7. If you are upgrading from a previous version of the Informatica Transports, select the **Overwrite Originals** option.

## Oracle privileges

To deploy and run an application ingestion task that has an Oracle source, the source connection must specify a Mass Ingestion Applications user who has the privileges required for the ingestion load type.

### Privileges for incremental load processing with log-based CDC

**Note:** If the Oracle logs are managed by ASM, the user must have SYSASM or SYSDBA authority.

For an application ingestion task that performs an incremental load or combined initial and incremental load using the log-based CDC method, ensure that the Mass Ingestion Applications user (*cmid\_user*) has been granted the following privileges:

```
GRANT CREATE SESSION TO <cmid_user>;

GRANT SELECT ON table TO <cmid_user>;      -- For each source table created by user
GRANT EXECUTE ON DBMS_FLASHBACK TO <cmid_user>;

-- The following grant is required for combined initial and incremental loads only. Do not
-- use ANY TABLE unless your security policy allows it.
GRANT FLASHBACK ON table|ANY TABLE TO <cmid_user>;

-- Include the following grant only if you want to Execute the CDC script for enabling
-- supplemental logging from the user interface. If you manually enable supplemental
-- logging, this grant is not needed.
GRANT ALTER table|ANY TABLE TO <cmid_user>;

GRANT SELECT ON DBA_CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON DBA_CONS_COLUMNS TO <cmid_user>;
```

```

GRANT SELECT ON DBA_INDEXES TO <cmid_user>;
GRANT SELECT ON DBA_LOG_GROUPS TO <cmid_user>;
GRANT SELECT ON DBA_LOG_GROUP_COLUMNS TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO <cmid_user>;
GRANT SELECT ON DBA_OBJECT_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_TABLESPACES TO <cmid_user>;
GRANT SELECT ON DBA_USERS TO <cmid_user>;

GRANT SELECT ON "PUBLIC".V$ARCHIVED_LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$CONTAINERS TO <cmid_user>; -- For Oracle multitenant environments
GRANT SELECT ON "PUBLIC".V$DATABASE TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$DATABASE_INCARNATION TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$ENCRYPTION_WALLET TO <cmid_user>; -- For Oracle TDE access
GRANT SELECT ON "PUBLIC".V$LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$LOGFILE TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$PARAMETER TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$PDBS TO <cmid_user>; -- For Oracle multitenant environments
GRANT SELECT ON "PUBLIC".V$SPPARAMETER TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$STANDBY_LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$THREAD TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$TRANSACTION TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$TRANSPORTABLE_PLATFORM TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$VERSION TO <cmid_user>;

GRANT SELECT ON SYS.ATTRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.COL$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
GRANT SELECT ON SYS.INDPART$ TO <cmid_user>;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.PARTOBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.TAB$ TO <cmid_user>;
GRANT SELECT ON SYS.TABCOMPART$ TO <cmid_user>;
GRANT SELECT ON SYS.TABPART$ TO <cmid_user>;
GRANT SELECT ON SYS.TABSUBPART$ TO <cmid_user>;

-- Also ensure that you have access to the following ALL_* views:
ALL_CONSTRAINTS
ALL_CONS_COLUMNS
ALL_ENCRYPTED_COLUMNS
ALL_INDEXES
ALL_IND_COLUMNS
ALL_OBJECTS
ALL_TABLES
ALL_TAB_COLS
ALL_TAB_PARTITIONS
ALL_USERS

```

### Privileges for incremental load processing with query-based CDC

For an application ingestion task that performs an incremental load or combined initial and incremental load using the query-based CDC method, ensure that the user has the following privileges at minimum:

```

GRANT CREATE SESSION TO <cmid_user>;

GRANT SELECT ON DBA_INDEXES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECT_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO <cmid_user>;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_USERS TO <cmid_user>;
GRANT SELECT ON DBA_VIEWS TO <cmid_user>; -- Only if you unload data from views

GRANT SELECT ANY TABLE TO <cmid_user>;
-or-
GRANT SELECT ON table TO <cmid_user>; -- For each source table created by user

```

```

GRANT SELECT ON ALL CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON ALL_CONS_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_ENCRYPTED_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_IND_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_INDEXES TO <cmid_user>;
GRANT SELECT ON ALL_OBJECTS TO <cmid_user>;
GRANT SELECT ON ALL_TAB_COLS TO <cmid_user>;
GRANT SELECT ON ALL_USERS TO <cmid_user>;

```

```

GRANT SELECT ON "PUBLIC"."V$DATABASE" TO cmid_user;
GRANT SELECT ON "PUBLIC"."V$CONTAINERS" TO cmid_user;
GRANT SELECT ON SYS.ATRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.COL$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO cmid_user;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.TAB$ TO <cmid_user>;

```

### Privileges for initial load processing

For an application ingestion task that performs an initial load, ensure that the user has the following privileges at minimum:

```

GRANT CREATE SESSION TO <cmid_user>;

GRANT SELECT ON DBA_INDEXES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECT TABLES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO cmid_user;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_USERS TO <cmid_user>;
GRANT SELECT ON DBA_VIEWS TO <cmid_user>; -- Only if you unload data from views

GRANT SELECT ANY TABLE TO <cmid_user>;
-or-
GRANT SELECT ON table TO <cmid_user>; -- For each source table created by user

GRANT SELECT ON ALL CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON ALL_CONS_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_ENCRYPTED_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_IND_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_INDEXES TO <cmid_user>;
GRANT SELECT ON ALL_OBJECTS TO <cmid_user>;
GRANT SELECT ON ALL_TAB_COLS TO <cmid_user>;
GRANT SELECT ON ALL_USERS TO <cmid_user>;

GRANT SELECT ON "PUBLIC"."V$DATABASE" TO cmid_user;
GRANT SELECT ON "PUBLIC"."V$CONTAINERS" TO cmid_user;
GRANT SELECT ON SYS.ATRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.COL$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO cmid_user;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.TAB$ TO <cmid_user>;n

```

## Oracle privileges for Amazon RDS for Oracle sources

If you have an Amazon RDS for Oracle source, you must grant certain privileges to the Mass Ingestion Applications user.

**Important:** You must log in to Amazon RDS under the master username to run GRANT statements and procedures.

Grant the SELECT privilege, at minimum, on objects and system tables that are required for CDC processing to the Mass Ingestion Applications user (*cmid\_user*). Some additional grants are required in certain situations.

Use the following GRANT statements:

```
GRANT SELECT ON "PUBLIC"."V$ARCHIVED_LOG" TO "cmid_user";

GRANT SELECT ON "PUBLIC"."V$DATABASE" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$LOG" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$LOGFILE" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$TRANSPORTABLE_PLATFORM" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$THREAD" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$DATABASE_INCARNATION" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$TRANSACTION" TO "cmid_user";

GRANT SELECT ON "SYS"."DBA_CONS_COLUMNS" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_CONSTRAINTS" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_INDEXES" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_LOG_GROUP_COLUMNS" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_TABLESPACES" TO "cmid_user";

GRANT SELECT ON "SYS"."OBJ$" TO "cmid_user";
GRANT SELECT ON "SYS"."TAB$" TO "cmid_user";
GRANT SELECT ON "SYS"."IND$" TO "cmid_user";
GRANT SELECT ON "SYS"."COL$" TO "cmid_user";

GRANT SELECT ON "SYS"."PARTOBJ$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABPART$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABCOMPART$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABSUBPART$" TO "cmid_user";
COMMIT;

/* For combined load jobs:*/
GRANT EXECUTE ON DBMS_FLASHBACK TO "cmid_user";

/*To provide read access to the Amazon RDS online and archived redo logs:*/
GRANT READ ON DIRECTORY ONLINELOG_DIR TO "cmid_user";
GRANT READ ON DIRECTORY ARCHIVELOG_DIR TO "cmid_user";
```

Additionally, log in as the master user and run the following Amazon RDS procedures to grant the SELECT privilege on some more objects:

```
begin
  rdsadmin.rdsadmin_util.grant_sys_object(
    p_obj_name => 'DBA_TABLES',
    p_grantee => 'cmid_user',
    p_privilege => 'SELECT');
end;
/
begin
  rdsadmin.rdsadmin_util.grant_sys_object(
    p_obj_name => 'DBA_OBJECTS',
    p_grantee => 'cmid_user',
    p_privilege => 'SELECT');
end;
/
begin
  rdsadmin.rdsadmin_util.grant_sys_object(
    p_obj_name => 'DBA_OBJECT_TABLES',
    p_grantee => 'cmid_user',
    p_privilege => 'SELECT');
end;
/
begin
  rdsadmin.rdsadmin_util.grant_sys_object(
    p_obj_name => 'DBA_VIEWS',
    p_grantee => 'cmid_user',
    p_privilege => 'SELECT');
end;
/
```

```

/ begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'DBA_USERS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$CONTAINERS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$PARAMETER',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$SPPARAMETER',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$STANDBY_LOG',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$VERSION',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_CONS_COLUMNS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_CONSTRAINTS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_OBJECTS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_TABLES',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;

```

```

/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_TAB_PARTITIONS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_USERS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_TABLES',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL_TAB_PARTITIONS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ATTRCOL$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'CCOL$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'COLTYPE$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'INDPART$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'IDNSEQ$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'CDEF$',
p_grantee => 'cmid_user',

```

```

p_privilege => 'SELECT');
end;
/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'RECYCLEBIN$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/
/* Only required for RDS21 which supports PDB*/
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$PDBS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
/

```

## Configuring BFILE access to Oracle redo logs in the Oracle file system

If you store redo logs in the local Oracle server file system and want to access the logs by using Oracle directory objects with BFILES, perform the following configuration tasks:

Complete the following usual Oracle source preparation tasks, which are not specific to BFILE access:

- Define the ORACLE\_HOME environment variable on the Linux or Windows system where the Secure Agent runs for Mass Ingestion Applications to use the Oracle Call Interface (OCI) to communicate with the Oracle source database.
- Make sure the Mass Ingestion Applications user has the Oracle privileges that are required for the application ingestion incremental load processing. For more information, see [“Oracle privileges” on page 24](#).
- Enable ARCHIVELOG mode for the Oracle database.
- Define the archive log destination.

**Note:** For BFILE access, use a specific archive log destination directory. Do not use the Oracle USE\_DB\_RECOVERY\_FILE\_DEST keyword to store archive logs by date in a Fast Recovery Area (FRA).

- Enable Oracle minimal global supplemental logging on the source database.
- If your Oracle source tables have primary keys, ensure that supplemental logging is enabled for all primary key columns. For source tables that do not have primary keys, ensure that supplemental logging is enabled for all columns from which change data will be captured.

**Note:** When you create an application ingestion task, you have the option of generating a script that implements supplemental logging for all columns or only primary key columns for the selected source tables.

- Ensure that the Oracle MAX\_STRING\_SIZE initialization parameter is *not* set to EXTENDED. If it is set to EXTENDED, Mass Ingestion Applications will not be able to replicate inserts and updates for tables containing columns defined with large (extended size) VARCHAR2, NVARCHAR2, or RAW columns.

Additionally, for BFILE access, perform the following steps:

1. Query the Oracle database for the online and archived redo log locations in the Oracle server file system. You can use the following example queries:

To get location of the online redo logs:

```
select * from v$logfile;
```

To get the archive log destination:

```
select dest_id, dest_name, destination, status from V$ARCHIVE_DEST;
```

2. Create the ONLINELOG\_DIR and ARCHIVELOG\_DIR directory objects that point to the locations of log files from step 1. An Oracle directory object specifies a logical alias name for a physical directory on the Oracle server file system under which the log files to be accessed are located. For example:

```
CREATE DIRECTORY ONLINELOG_DIR AS '/u01/oracle/data';
CREATE DIRECTORY ARCHIVELOG_DIR AS '/u01/oracle/archivedata';
```

**Note:** If you plan to set the reader mode to ARCHIVEONLY in the Oracle Database Ingestion connection to read changes only from archive logs, you do not need to create an ONLINELOG\_DIR directory or directory object.

The Oracle database does not verify that the directories you specify exist. Make sure you specify valid directories that exist in the Oracle file system.

3. To verify that the directory objects were created with the correct file system paths for the redo logs, issue a select statement such as:

```
select * from all_directories;
OWNER      DIRECTORY_NAME      DIRECTORY_PATH
-----
SYS         ARCHIVELOG_DIR         /u01/oracle/data/JO112DTL
SYS         ONLINELOG_DIR          /u01/oracle/data/JO112DTL
```

4. Grant read and write access on the ONLINELOG\_DIR and ARCHIVELOG\_DIR directory objects to the Mass Ingestion Applications user who is specified in the Oracle Database Ingestion connection properties. For example:

```
grant read on directory "ARCHIVELOG_DIR" to "cmid_user";
grant read on directory "ONLINELOG_DIR" to "cmid_user";
```

5. In the Oracle Database Ingestion connection properties, select the **BFILE Access** check box.

## Oracle archive log retention considerations

Application ingestion incremental and combined initial and incremental load jobs must be able to access transaction data in Oracle online and archive redo logs. If the logs are not available, application ingestion jobs end with an error.

Typically, the Oracle DBA sets the archive log retention period based on your organization's particular business needs and Oracle environment. Make sure that the source archive logs are retained for the longest period for which you expect change capture to be stopped or latent, plus about 1 hour, so that the logs will be available for restart processing.

To determine if the current log retention policy in your environment is sufficient to accommodate application ingestion change capture processing, consider the following factors:

- How long are Oracle transactions typically open on a source?
- What is the longest period of time that change capture is allowed to be down or latent, accounting for weekends and holidays?
- What is the replication latency from source to target?
- Do you run application ingestion jobs based on a schedule? If yes, what type of schedule?
- Is the `pxw.cdcreader.oracle.option.additional ageOutPeriod=minutes` custom property set on the **Source** page of task wizard?

**Note:** This property specifies the age at which outstanding UOWs without change records of CDC interest are removed from the calculation of the next restart point. You can use the property to prevent CDC failures that might occur if you shut down and then restart capture processing while the transaction is outstanding and the redo log in which the UOW started is not available.

- What is the redo generation rate?
- Do you ship copies of archive logs to a secondary system?

If the archive logs are not available when you need to restart capture processing in the logs, you can ask your DBA to restore them and to modify the retention period if necessary. Otherwise, perform another initial load to re-materialize the target and then start incremental change data processing again. However, in this case, you might lose some changes.

## Guidelines for ServiceNow sources

Consider the following guidelines when you use ServiceNow sources:

- The first time an incremental load job runs to capture change data for a ServiceNow source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- The data stored in some ServiceNow objects can be retrieved only if the API client is configured with the maint and nobody roles. Mass Ingestion Applications does not retrieve data from such source objects that require the ServiceNow Mass Ingestion connection to be configured with the maint and nobody roles.
- Incremental load jobs configured for ServiceNow sources do not capture change data for the objects that do not contain the sys\_updated\_on field.
- Incremental load jobs and combined initial and incremental load jobs configured for ServiceNow sources do not detect and replicate the renaming of source fields.
- Mass Ingestion Applications retains the masking configuration of source fields on the target. For example, if a masked field on the source contains the value 1234\*\*\*\*, the corresponding field on the target also contains the value 1234\*\*\*\*.
- In the Secure Agent configuration properties, set the metadata\_manager.jvm.ai.servicenow.enable.cdc.using.inherited.columns and task\_container.jvm.ai.servicenow.enable.cdc.using.inherited.columns custom properties to **true** to include the child tables inheriting CDC columns from the parent table in the task.
- Set the ai.servicenow.enable.additional.datatypes custom property to true, to enable the following ServiceNow data types:
  - Audio
  - Basic Image
  - Encrypted Text
  - FX Currency
  - Image
  - Journal List
  - List
  - Name-Value Pairs
  - Reference
  - Video
  - Wiki

**Note:** For the **Reference** data type, when the ai.servicenow.enable.additional.datatypes custom property is set to true, the columnName\_value, columnName\_display\_value, and columnName\_link attributes are available in the target. If the ai.servicenow.enable.additional.datatypes custom property is set to false, only the columnName attribute is available in the target.

## Guidelines for Workday sources

Consider the following guidelines when you use Workday sources:

Mass Ingestion Applications provides you the option to extract Workday data through the following web services:

- **Workday Web Services:** Provides access to Workday data through SOAP APIs.
- **Workday Report-as-a-Service (RaaS):** Provides access to data in the custom objects and fields through custom reports.

The option to select the required web service appears on the **Source** tab of the application ingestion task wizard.

### Guidelines for Workday Web Services

- You can use Workday Web Services to ingest the data of Workday Human Capital Management (HCM). On the **Source** tab of the application ingestion task wizard, you can select the specific HCM services that you want to replicate on your target.
- Mass Ingestion Applications replicates the data of only the source operations whose name start with `Get_`.
- Application ingestion jobs retrieve the source data in an XML structure and then writes the data to the target as a single object in JSON or XML format. When you configure an application ingestion job, on the **Source** tab of the task wizard, you can specify the format of the target data. On the target, each record contains the following fields:
  - **WID:** Stores the unique identifier or primary key of the record.
  - **Data:** Stores the content of the record in JSON or XML format.
- Mass Ingestion Applications retains the hierarchical structure of source data on the target.
- Mass Ingestion Applications does not propagate the data stored in custom objects.
- When an application ingestion job ingests an operation to the target, the operation is renamed in the following format: `<Service_Name>__<Operation_Name>`
- Workday might not allow initial load jobs to replicate the data of the `Get_Budget_Pools` operation.
- Initial load jobs configured for a Workday source and a Microsoft Azure Data Lake Storage Gen2 target do not generate schema files for the following operations if they do not contain any record:
  - `Benefits_Administration__Get_ACA_1095-C_Forms_Data`
  - `Compensation_Review__Get_Budget_Pools`
  - `Compensation_Review__Get_Employee_Awards`
  - `Recruiting__Get_Assess_Candidate`
  - `Recruiting__Get_Interview_Feedbacks`
- The first time an incremental load job runs to capture change data for a Workday source, the job retrieves and loads only the change records that are created after the date and time specified in the application ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs can automatically capture the insert and update operations performed on the source. The jobs perform upsert operations on the target to replicate the insert and update operations performed on the source data. The application ingestion task wizard does not provide the schema drift options for Workday sources because the upsert operations automatically replicate all changes made to the source schema.

- Mass Ingestion Applications can capture change data for the following Workday services and operations:

Service	Operations
Human_Resources	Get_Job_Profiles
	Get_Organizations
	Get_Workers
Recruiting	Get_Evergreen_Requisitions
	Get_Job_Requisitions
	Get_Organizations
	Get_Positions
Staffing	Get_Organizations
	Get_Positions
	Get_Workers

### Guidelines for Workday RaaS

- You can use Workday RaaS only in initial load tasks.
- On the **Source** tab of the application ingestion task wizard, you can choose to read data from one or more custom reports. If you choose to extract multiple reports, create a CSV file on the Secure Agent host and list the name of the reports in the first column of the file.
- Application ingestion jobs configured to use Workday RaaS cannot detect the changes made to the source schema.

## Guidelines for Zendesk sources

Consider the following guidelines when you use Zendesk sources:

- The first time an incremental load job runs to capture change data for a Zendesk source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs and combined initial and incremental load job configured for Zendesk sources do not detect and replicate the changes that are made to the source schema.
- If a source record contains multiple custom fields, application ingestion jobs store the data of all the custom fields as a JSON object in a single column of the target table.
- Application ingestion jobs can retrieve data from the custom fields that are present in the following Zendesk objects:
  - Organizations
  - Requests
  - Tickets

- Users
- Initial load jobs propagate null values for the data stored in the following Zendesk fields:
  - dns results field in the Support Addresses object.
  - raw request and raw response fields in the Target Failures object.
  - reason code field in the Satisfactory Ratings object.
  - URL field in the Sharing Agreements object.
- Initial load jobs do not propagate the data stored in the following Zendesk objects:
  - Attachments
  - Channel Framework
  - Dynamic Content Item Variants
  - End users
  - Incremental Skill-based Routing
  - NPS<sup>®</sup> Invitations
  - NPS<sup>®</sup> Recipients
  - NPS<sup>®</sup> Responses
  - OAuth Tokens for Grant Types
  - Push Notification Devices
  - Search
  - Side Conversations
  - Side Conversation Attachment
  - Side Conversation Events
  - Skill-based Routing
  - Ticket Comments
  - Ticket Import
  - User Identities
  - User Passwords
- Mass Ingestion Applications can capture change data for the following Zendesk standard objects:
  - NPS Recipients
  - Organizations
  - Side Conversation Events
  - Tickets
  - Tickets Metric Events
  - Users
- Incremental load jobs can capture the deletion of records that are stored in the following objects:
  - Organizations
  - Tickets
  - Users

- Mass Ingestion Applications does not replicate the hierarchical structure of the source fields of the type Object. On the target table, all fields are at the same hierarchical level. When an application ingestion job replicates an Object field with multiple hierarchical levels, the job creates the corresponding columns at the same hierarchical level.

## Supported targets

The targets that Mass Ingestion Applications support depend on the sources specified for the application ingestion tasks.

The following table lists the targets that Mass Ingestion Applications support for each source type:

Source type	Supported target type
Adobe Analytics	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
Google Analytics	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
Marketo	Amazon Redshift, Amazon S3, Apache Kafka (incremental load), Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
Microsoft Dynamics 365	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
NetSuite	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
Oracle Fusion Cloud	<ul style="list-style-type: none"> <li>- Using REST API - Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake</li> <li>- Using BICC - Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake</li> </ul>
Salesforce	Amazon Aurora PostgreSQL, Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Oracle, Microsoft Azure SQL Database, Microsoft Fabric OneLake, and Snowflake
Salesforce Marketing Cloud	Amazon Redshift, Amazon S3, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake

Source type	Supported target type
SAP	<ul style="list-style-type: none"> <li>- Using SAP ODP Extractor connector - Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake</li> <li>- Using SAP Mass Ingestion connector - Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Azure SQL Database (initial load), Microsoft Fabric OneLake, Oracle, and Snowflake</li> </ul>
ServiceNow	Amazon Redshift, Apache Kafka, Databricks Delta, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake
Workday	<ul style="list-style-type: none"> <li>- SOAP. Amazon Redshift, Amazon S3, Apache Kafka (incremental load), Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake</li> <li>- RaaS. Amazon Redshift, Amazon S3, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake</li> </ul>
Zendesk	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, and Snowflake

To determine the connectors to use for the target types, see *Connectors and Connections > Mass Ingestion Applications connectors*.

## Guidelines for Amazon Redshift targets

Consider the following guidelines when you use Amazon Redshift targets:

- Before writing data to Amazon Redshift target tables, application ingestion jobs stage the data in an Amazon S3 bucket. You must specify the name of the bucket when you configure the application ingestion task. The ingestion jobs use the COPY command to load the data from the Amazon S3 bucket to the Amazon Redshift target tables. For more information about the COPY command, see the Amazon Web Services documentation.
- When you define a connection for an Amazon Redshift target, provide the access key and secret access key for the Amazon S3 bucket in which you want the application ingestion jobs to stage the data before loading it to the Amazon Redshift target tables.
- When you ingest data from a source to an Amazon Redshift target, the application ingestion job fails if the data source contains more than 32 data fields or columns that are defined as primary keys.
- Incremental load jobs and combined initial and incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema of the target tables.

# Guidelines for Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Fabric OneLake, and Oracle Cloud Object Storage targets

Consider the following guidelines when you use Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Fabric OneLake, and Oracle Cloud Infrastructure (OCI) Object Storage targets:

- When you configure an application ingestion task for an Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Fabric OneLake, or Oracle Cloud Object Storage target, you can select CSV, Avro, or Parquet as the format for the output files that contain the source data to be applied to the target.
- If you select **CSV** as the output file format, Mass Ingestion Applications creates the following files on the target for each source field:
  - schema.ini file that describes the schema of the field. The file also includes some settings for the output file on the target.
  - Output files that contain the data stored in the source field. Mass Ingestion Applications names the output files based on the name of the source field with an appended date and time.

The schema.ini file lists the sequence of columns for the rows in the corresponding output file. The following table describes the columns in the schema.ini file:

Column	Description
ColNameHeader	Indicates whether the source data files include column headers.
Format	Format of the output files. Mass Ingestion Applications uses a comma (,) to delimit column values.
CharacterSet	Character set that is used for the corresponding output file. By default, Mass Ingestion Applications generates the files in the UTF-8 character set.
COL<sequence_number>	Name and data type of the source field. <b>Notes:</b> <ul style="list-style-type: none"><li>- If you selected any of the <b>Add Operation...</b> properties under <b>Advanced</b> on the <b>Target</b> page of the task wizard, the list of columns includes metadata columns for the operation type, time, owner, or transaction ID.</li><li>- If you selected the <b>Add Before Images</b> check box, for each source column, the job creates a <i>column_name_OLD</i> column for UNDO data and <i>column_name_NEW</i> column for REDO data.</li></ul>

**Note:** You must not edit the schema.ini file.

- If you select the Avro output format, you can select an Avro format type, a file compression type, an Avro data compression type, and the directory that stores the Avro schema definitions generated for each source table. The schema definition files have the following naming pattern: *schemaname\_tablename.txt*.
- If you select the Parquet output format, you can optionally select a compression type that Parquet supports.
- For application ingestion tasks configured for Microsoft Azure Data Lake Storage Gen2 and Microsoft Fabric OneLake targets, Mass Ingestion Applications creates an empty directory on the target for each empty source field.
- For Amazon S3 targets, if you do not specify an access key and secret key in the connection properties, Mass Ingestion Applications tries to find the AWS credentials by using the default credential provider

chain that is implemented by the `DefaultAWSCredentialsProviderChain` class. For more information, see the Amazon Web Services documentation.

- When an incremental load job or combined initial and incremental load job configured for a target that uses the CSV output format propagates an Update operation that changed primary key values on the source, the job performs a Delete operation on the associated target row and then performs an Insert operation on the same row to replicate the change made to the source object. The Delete operation writes the before image to the target and the subsequent Insert image writes the after image to the target.

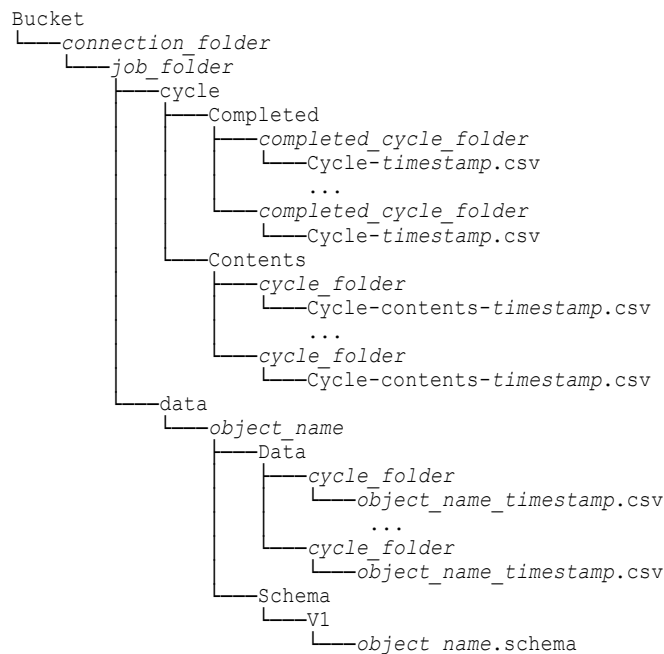
For Update operations that do not change primary key values, application ingestion jobs process each Update operation as a single operation and writes only the after image to the target.

**Note:** If a source object does not contain any primary key, Mass Ingestion Applications considers all fields of the object to be a part of the primary key. In such scenarios, Mass Ingestion Applications processes each Update operation performed on the source as a Delete operation followed by an Insert operation on the target.

## Default directory structure of CDC files on Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, and Microsoft Fabric OneLake targets

Application ingestion jobs create directories on Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, and Microsoft Fabric OneLake targets to store information about change data processing.

The following directory structure is created by default on the targets:



The following table describes the directories in the default structure:

Folder	Description
<i>connection_folder</i>	Contains the Mass Ingestion Applications objects. This folder is specified in the <b>Folder Path</b> field of the Amazon S3 connection properties, in the <b>Directory Path</b> field of the Microsoft Azure Data Lake Storage Gen2 connection properties, or in the <b>Lakehouse Path</b> field of the Microsoft Fabric OneLake connection properties. <b>Note:</b> This folder is not created for Google Cloud Storage targets.
<i>job_folder</i>	Contains job output files. This folder is specified in the <b>Directory</b> field on the <b>Target</b> page of the application ingestion task wizard.
cycle/Completed	Contains a subfolder for each completed CDC cycle. Each cycle subfolder contains a completed cycle file.
cycle/Contents	Contains a subfolder for each CDC cycle. Each cycle subfolder contains a cycle contents file.
data	Contains output data files and schema files for each object.
data/object_name/ Schema/V1	Contains a schema file. <b>Note:</b> Mass Ingestion Applications does not save a schema file in this folder if the output files use the Parquet format.
data/object_name/ Data	Contains a subfolder for each CDC cycle that produces output data files.

## Cycle directories

Mass Ingestion Applications uses the following pattern to name cycle directories:

```
[dt=]yyyy-mm-dd-hh-mm-ss
```

The "dt=" prefix is added to cycle folder names if you select the **Add Directory Tags** check box on the **Target** page of the application ingestion task wizard.

## Cycle contents files

Cycle contents files are located in cycle/Contents/*cycle\_folder* subdirectories. Cycle contents files contain a record for each object that has had a DML event during the cycle. If no DML operations occurred on an object in the cycle, the object does not appear in the cycle contents file.

Mass Ingestion Applications uses the following pattern to name cycle content files:

```
Cycle-contents-timestamp.csv
```

A cycle contents csv file contains the following information:

- Object name
- Cycle name
- Path to the cycle folder for the object
- Start sequence for the object
- End sequence for the object
- Number of Insert operations
- Number of Update operations
- Number of Delete operations

- **Combined load jobs only.** Number of Insert operations encountered during the initial load phase
- **Combined load jobs only.** Number of Delete operations encountered during the initial load phase
- Schema version
- Path to the schema file for the schema version

**Note:** If the output data files use the Parquet format, Mass Ingestion Applications does not save a schema file at the path that is specified in the cycle contents file. Instead, use the schema file in the folder that is specified in the **Avro Schema Directory** field on the **Target** page of the application ingestion task wizard.

## Completed cycle files

Completed cycle files are located in `cycle/Completed/completed_cycle_folder` subdirectories. An application ingestion job creates a cycle file in this subdirectory after a cycle completes. If this file is not present, the cycle has not completed yet.

Mass Ingestion Applications uses the following pattern to name completed cycle files:

`Cycle-timestamp.csv`

A completed cycle csv file contains the following information:

- Cycle name
- Cycle start time
- Cycle end time
- Current sequence number at the time the cycle ended
- Path to the cycle contents file
- Reason for the end of cycle

Valid reason values are:

- **NORMAL\_COMMIT.** A commit operation was encountered after the cycle had reached the DML limit or the end of the cycle interval. A cycle can end only on a commit boundary.
- **NORMAL\_EXPIRY.** The cycle ended because the cycle interval expired. The last operation was a commit.
- **Combined load jobs only: BACKLOG\_COMPLETED.** The cycle ended because CDC backlog processing completed. The CDC backlog consists of events captured during the initial load phase of the combined job. The backlog includes potential DML changes captured at the beginning or end of the initial load phase and during the transition from the initial load phase to the main CDC incremental processing.
- **Combined load jobs only: INITIAL\_LOAD\_COMPLETED.** The cycle ended because the initial load completed.
- **Combined load jobs only: RESYNC\_STARTED.** The cycle ended because the object resync initiated.

## Output data files

The data files contain records that include the following information:

- Operation type. Valid values are:
  - **I** for Insert operations
  - **U** for Update operations
  - **D** for Delete operations
  - **Combined load jobs only: X** for Delete operations encountered during the initial load phase of a combined load job
  - **Combined load jobs only: Y** for Insert operations encountered during the initial load phase of a combined load job

- Sortable sequence number. In combined initial and incremental load jobs, the sortable sequence number contains a 20-digit prefix that can be used to align rows with the resync version and the load job. The prefix is a combination of the following attributes:
  1. Incarnation. This nine-digit number is incremented each time the object is resynced. The initial value is 1.
  2. Schema version. This nine-digit number is incremented each time a schema drift change is propagated for the object. The initial value is 1.
  3. Phase. This two-digit number changes when transition from unload, to backlog, to CDC is performed. Valid values are:
    - 00 for Truncation, which is the first data record written during initial load or resync
    - 01 for a normal insert during initial load or resync
    - 02 for a change detected during the initial load
    - 03 for a change detected after the initial load or resync is completed but before the transition back to the main CDC phase
    - 04 for a change detected during the normal CDC phase
- Data fields
 

**Note:** Insert and Delete records contain only after images. Update records contain both before and after images.

## Custom directory structure for output files on Amazon S3, Google Cloud Storage, Microsoft Fabric OneLake, and ADLS Gen2 targets

You can configure a custom directory structure for the output files that initial load, incremental load, and combined initial and incremental load jobs write to Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage (ADLS) Gen2, and Microsoft Fabric OneLake targets if you do not want to use the default structure.

### Initial loads

By default, initial load jobs write output files to *tablename\_timestamp* subdirectories under the parent directory. For Amazon S3 and ADLS Gen2 targets, the parent directory is specified in the target connection properties if the **Connection Directory as Parent** check box is selected on the **Target** page of the task wizard.

- In an Amazon S3 connection, this parent directory is specified in the **Folder Path** field.
- In an ADLS Gen2 connection, the parent directory is specified in the **Directory Path** field.

For Google Cloud Storage targets, the parent directory is the bucket container specified in the **Bucket** field on the **Target** page of the task wizard.

For Microsoft Fabric OneLake targets, the parent directory is the path specified in the **Lakehouse Path** field in the Microsoft Fabric OneLake connection properties.

You can customize the directory structure to suit your needs. For example, for initial loads, you can write the output files under a root directory or directory path that is different from the parent directory specified in the connection properties to better organize the files for your environment or to find them more easily. Or you can consolidate all output files for a table directly in a directory with the table name rather than write the files to separate timestamped subdirectories, for example, to facilitate automated processing of all of the files.

To configure a directory structure, you must use the **Data Directory** field on the **Target** page of the ingestion task wizard. The default value is `{TableName}_{Timestamp}`, which causes output files to be written to *tablename\_timestamp* subdirectories under the parent directory. You can configure a custom directory path

by creating a directory pattern that consists of any combination of case-insensitive placeholders and directory names. The placeholders are:

- {TableName} for a target table name
- {Timestamp} for the date and time, in the format yyyyymmdd\_hhmissms, at which the initial load job started to transfer data to the target
- {Schema} for the target schema name
- {YY} for a two-digit year
- {YYYY} for a four-digit year
- {MM} for a two-digit month value
- {DD} for a two-digit day in the month

A pattern can also include the following functions:

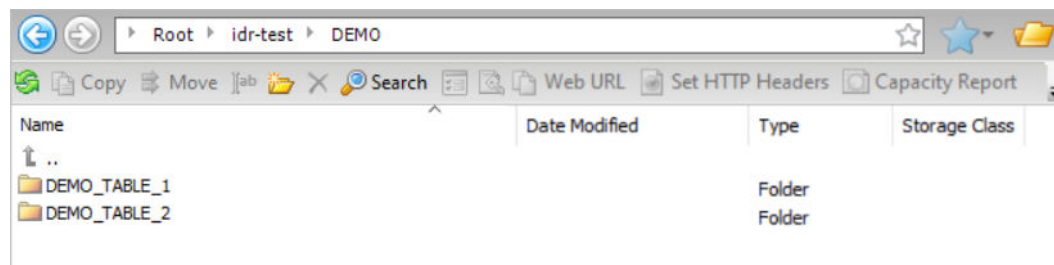
- toLower() to use lowercase for the values represented by the placeholder in parentheses
- toUpper() to use uppercase for the values represented by the placeholder in parentheses

By default, the target schema is also written to the data directory. If you want to use a different directory for the schema, you can define a directory pattern in the **Schema Directory** field.

#### Example 1

You are using an Amazon S3 target and want to write output files and the target schema to the same directory, which is under the parent directory specified in the **Folder Path** field of the connection properties. In this case, the parent directory is `idr-test/DEMO/`. You want to write all of the output files for a table to a directory that has a name matching the table name, without a timestamp. You must complete the **Data Directory** field and select the **Connection Directory as Parent** check box.

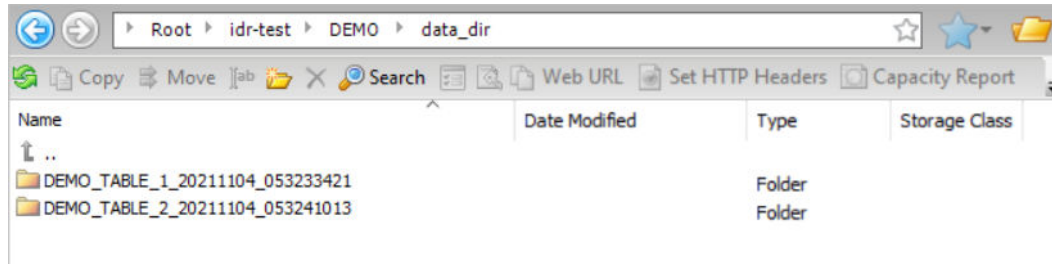
Based on this configuration, the resulting directory structure is:



#### Example 2

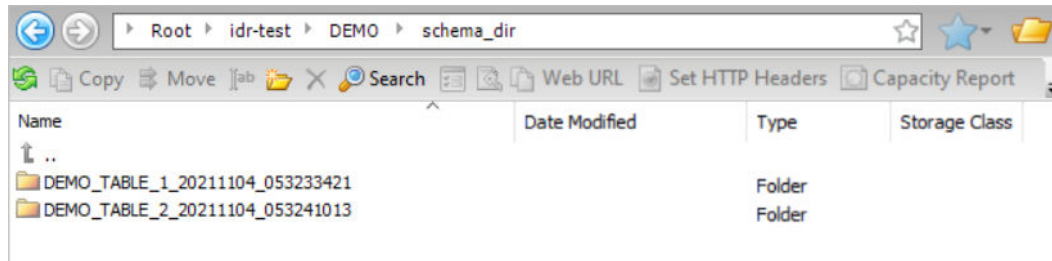
You are using an Amazon S3 target and want to write output data files to a custom directory path and write the target schema to a separate directory path. To use the directory specified in the **Folder Path** field in the Amazon S3 connection properties as the parent directory for the data directory and schema directory, select **Connection Directory as Parent**. In this case, the parent directory is `idr-test/DEMO/`. In the **Data Directory** and **Schema Directory** fields, define directory patterns by using a specific directory name, such as `data_dir` and `schema_dir`, followed by the default `{TableName}_{Timestamp}` placeholder value. The placeholder creates `tablename_timestamp` destination directories.

Based on this configuration, the resulting data directory structure is:



Name	Date Modified	Type	Storage Class
↑ ..			
DEMO_TABLE_1_20211104_053233421		Folder	
DEMO_TABLE_2_20211104_053241013		Folder	

And the resulting schema directory structure is:



Name	Date Modified	Type	Storage Class
↑ ..			
DEMO_TABLE_1_20211104_053233421		Folder	
DEMO_TABLE_2_20211104_053241013		Folder	

### Incremental loads and combined initial and incremental loads

By default, incremental load and combined initial and incremental load jobs write cycle files and data files to subdirectories under the parent directory. However, you can create a custom directory structure to organize the files to best suit your organization's requirements.

This feature applies to application ingestion incremental load jobs that have a Salesforce source and Amazon S3, Google Cloud Storage, Microsoft Fabric OneLake, or Microsoft Azure Data Lake Storage (ADLS) Gen2 targets.

For Amazon S3 and ADLS Gen2 targets, the parent directory is set in the target connection properties if the **Connection Directory as Parent** check box is selected on the **Target** page of the task wizard.

- In an Amazon S3 connection, the parent directory is specified in the **Folder Path** field.
- In an ADLS Gen2 connection, the parent directory is specified in the **Directory Path** field.

For Google Cloud Storage targets, the parent directory is the bucket container specified in the **Bucket** field on the **Target** page of the task wizard.

For Microsoft Fabric OneLake targets, the parent directory is the path specified in the **Lakehouse Path** field in the Microsoft Fabric OneLake connection properties.

You can customize the directory structure to suit your needs. For example, you can write the data and cycle files under a target directory for the task instead of under the parent directory specified in the connection properties. Alternatively, you can 1) consolidate table-specific data and schema files under a subdirectory that includes the table name, 2) partition the data files and summary contents and completed files by CDC cycle, or 3) create a completely customized directory structure by defining a pattern that includes literal values and placeholders. For example, if you want to run SQL-type expressions to process the data based on time, you can write all data files directly to timestamp subdirectories without partitioning them by CDC cycle.

To configure a custom directory structure for an incremental load task, define a pattern for any of the following optional fields on the **Target** page of the ingestion task wizard:

Field	Description	Default
Task Target Directory	<p>Name of a root directory to use for storing output files for an incremental load task.</p> <p>If you select the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory. It will be appended to the parent directory to form the root for the data, schema, cycle completion, and cycle contents directories.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>	None
Connection Directory as Parent	<p>Select this check box to use the parent directory specified in the connection properties.</p> <p>This field is not available for the Microsoft Fabric OneLake target.</p>	Selected
Data Directory	<p>Path to the subdirectory that contains the data files.</p> <p>In the directory path, the {TableName} placeholder is required if data and schema files are <i>not</i> partitioned by CDC cycle.</p>	{TaskTargetDirectory}/data/{TableName}/data
Schema Directory	<p>Path to the subdirectory in which to store the schema file if you do not want to store it in the data directory.</p> <p>In the directory path, the {TableName} placeholder is required if data and schema files are not partitioned by CDC cycle.</p>	{TaskTargetDirectory}/data/{TableName}/schema
Cycle Completion Directory	Path to the directory that contains the cycle completed file.	{TaskTargetDirectory}/cycle/completed
Cycle Contents Directory	Path to the directory that contains the cycle contents files.	{TaskTargetDirectory}/cycle/contents
Use Cycle Partitioning for Data Directory	<p>Causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>	Selected
Use Cycle Partitioning for Summary Directories	Causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.	Selected
List Individual Files in Contents	<p>Lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>	<p>Not selected if <b>Use Cycle Partitioning for Summary Directories</b> is selected.</p> <p>Selected if you cleared <b>Use Cycle Partitioning for Summary Directories</b>.</p>

A directory pattern consists of any combination of case-insensitive placeholders, shown in curly brackets {}, and specific directory names. The following placeholders are supported:

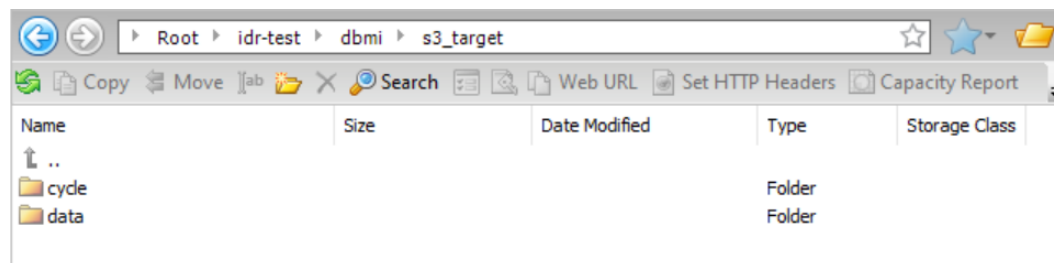
- {TaskTargetDirectory} for a task-specific base directory on the target to use instead of the directory the connection properties
- {TableName} for a target table name
- {Timestamp} for the date and time, in the format yyyyymmdd\_hhmissms
- {Schema} for the target schema name
- {YY} for a two-digit year
- {YYYY} for a four-digit year
- {MM} for a two-digit month value
- {DD} for a two-digit day in the month

**Note:** The timestamp, year, month, and day placeholders indicate when the CDC cycle started when specified in patterns for data, contents, and completed directories, or indicate when the CDC job started when specified in the schema directory pattern.

### Example 1

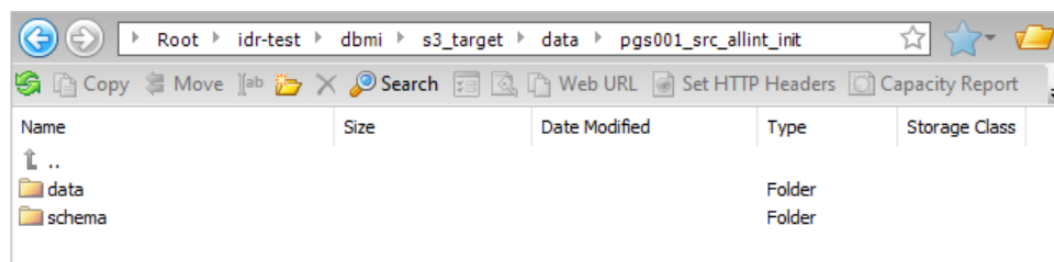
You want to accept the default directory settings for incremental load jobs as displayed in the task wizard. The target type is Amazon S3. Because the **Connection Directory as Parent** check box is selected by default, the parent directory path that is specified in the **Folder Path** field of the Amazon S3 connection properties is used. This parent directory is `idr-test/dbmi/`. You also must specify a task target directory name, in this case, `s3_target`, because the {TaskTargetDirectory} placeholder is used in the default patterns in the subsequent directory fields. The files in the data directory and schema directory will be grouped by table name because the {TableName} placeholder is included in their default patterns. Also, because cycle partitioning is enabled, the files in the data directory, schema directory, and cycle summary directories will be subdivided by CDC cycle.

Based on this configuration, the resulting data directory structure is:



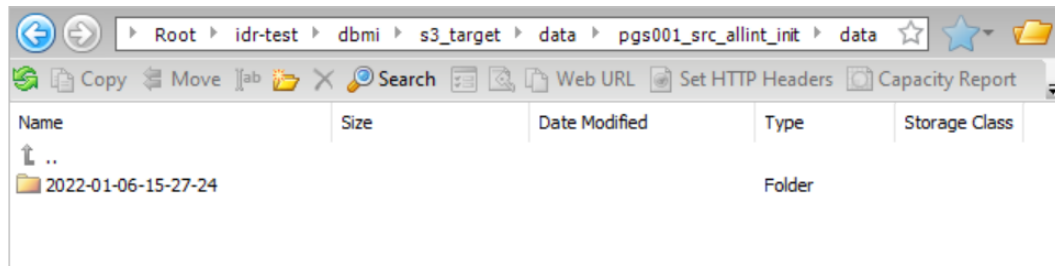
Name	Size	Date Modified	Type	Storage Class
↑ ..				
cycle			Folder	
data			Folder	

If you drill down on the data folder and then on a table in that folder (`pgs001_src_allint_init`), you can access the data and schema subdirectories:

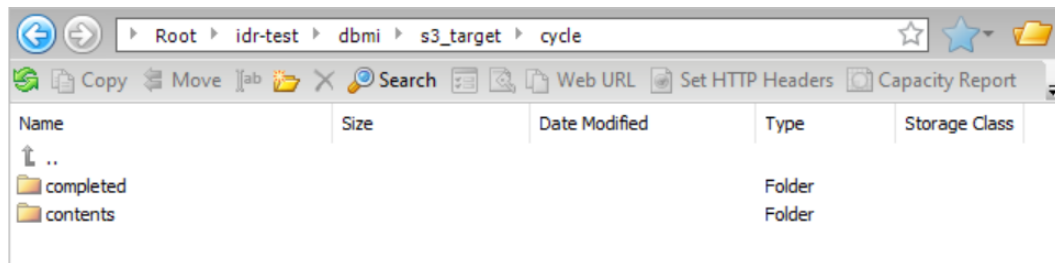


Name	Size	Date Modified	Type	Storage Class
↑ ..				
data			Folder	
schema			Folder	

If you drill down on the data folder, you can access the timestamp directories for the data files:



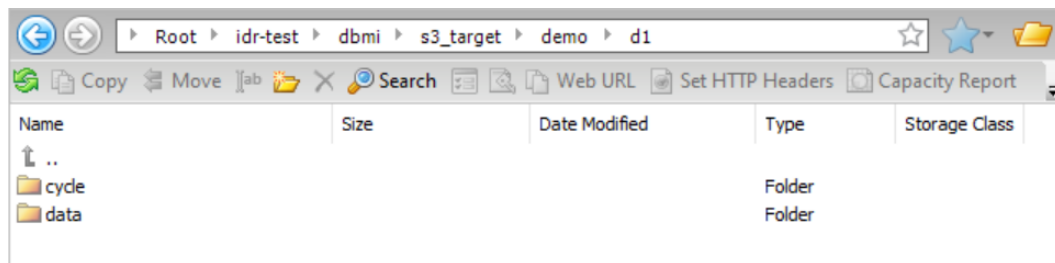
If you drill down on cycle, you can access the summary contents and completed subdirectories:



## Example 2

You want to create a custom directory structure for incremental load jobs that adds the subdirectories "demo" and "d1" in all of the directory paths except in the schema directory so that you can easily find the files for your demos. Because the **Connection Directory as Parent** check box is selected, the parent directory path (idr-test/dbmi/) that is specified in the **Folder Path** field of the Amazon S3 connection properties is used. You also must specify the task target directory because the {TaskTargetDirectory} placeholder is used in the patterns in the subsequent directory fields. The files in the data directory and schema directory will be grouped by table name. Also, because cycle partitioning is enabled, the files in the data, schema, and cycle summary directories will be subdivided by CDC cycle.

Based on this configuration, the resulting data directory structure is:



## Guidelines for Apache Kafka targets

Consider the following guidelines when you use Apache Kafka targets:

- Mass Ingestion Applications supports Apache Kafka as targets for incremental load jobs.  
To indicate the Kafka target type, you must specify Kafka producer properties in the task definition or Kafka connection properties. To specify these properties for a task, enter a comma-separated list of *key:value* pairs in the **Producer Configuration Properties** field on the **Target** page of the task wizard. To specify the producer properties for all tasks that use a Kafka connection, enter the list of properties in the **Additional Connection Properties** field in the connection properties. You can override the connection-level properties for specific tasks by also defining producer properties at the task level. For more information about producer properties, see the Apache Kafka documentation.
- If you select **AVRO** as the output format for a Kafka target, Mass Ingestion Applications generates a schema definition file for each table with a name in the following format:  

```
schema_name_tablename.txt
```

  
If a source schema change is expected to alter the target in an incremental load job, Mass Ingestion Applications regenerates the Avro schema definition file with a unique name that includes a timestamp:  

```
schema_name_tablename_YYYYMMDDhhmmss.txt
```

  
This unique naming pattern preserves older schema definition files for audit purposes.
- You can specify Kafka producer properties in either the **Producer Configuration Properties** field on the **Target** page of the task wizard or in the **Additional Connection Properties** field in the Kafka connection properties. Enter property=value pairs that meet your business needs and that are supported by your Kafka vendor.

## Guidelines for Databricks Delta targets

Consider the following guidelines when you use Databricks Delta targets:

- When you use a Databricks Delta target for the first time, perform the following steps before you configure an application ingestion task for the target:
  1. Download the DatabricksJDBC42 driver version 2.6.25 from the Databricks JDBC driver downloads website.
  2. Copy the DatabricksJDBC42.jar file to the following directory:  

```
Secure_Agent_installation_directory/apps/Database_Ingestion/ext/
```
  3. In the Databricks Delta connection properties, set the **JDBC Driver Class Name** property to `com.databricks.client.jdbc.Driver`.
  4. On Windows, install Visual C++ Redistributable Packages for Visual Studio 2013 on the computer where the Secure Agent runs.
- For incremental load jobs, you must enable Change Data Capture (CDC) for all source fields.
- You can access Databricks Delta tables created on top of the following storage types:
  - Microsoft Azure Data Lake Storage (ADLS) Gen2
  - Amazon Web Services (AWS) S3

The Databricks Delta connection uses a JDBC URL to connect to the Databricks cluster. When you configure the target, specify the JDBC URL and credentials to use for connecting to the cluster. Also define the connection information that the target uses to connect to the staging location in Amazon S3 or ADLS Gen2.

- Before writing data to Databricks Delta target tables, application ingestion jobs stage the data in an Amazon S3 bucket or ADLS directory. You must specify the directory for the data when you configure the application ingestion task.

**Note:** Mass Ingestion Applications does not use the **ADLS Staging Filesystem Name** and **S3 Staging Bucket** properties in the Databricks Delta connection properties to determine the directory.

- Mass Ingestion Applications uses jobs that run once to load data from staging files on AWS S3 or ADLS Gen2 to external tables.  
By default, Mass Ingestion Applications runs jobs on the cluster that is specified in the Databricks Delta connection properties. If you want to run the jobs on another cluster, set the `dbDeltaUseExistingCluster` custom property to `false` on the **Target** page in the application ingestion task wizard.
- If the cluster specified in the Databricks Delta connection properties is not up and running, the application ingestion job waits for the cluster to start. By default, the job waits for 10 minutes. If the cluster does not start within 10 minutes, the connection times out and deployment of the job fails.  
If you want to increase the timeout value for the connection, set the `dbClusterStartWaitingTime` custom property to the maximum time in seconds for which the ingestion job must wait for the cluster to be up and running. You can set the custom property on the **Target** page in the application ingestion task wizard.
- By default, Mass Ingestion Applications uses the Databricks Delta COPY INTO feature to load data from the staging file to Databricks Delta target tables. You can disable it for all load types by setting the `writerDatabricksUseSqlLoad` custom property to `false` on the **Target** page in the application ingestion task wizard.
- If you use an AWS cluster, you must specify the **S3 Service Regional Endpoint** value in the Databricks Delta connection properties. For example:

```
s3.us-east-2.amazonaws.com
```

Before you can test a Databricks Delta connection, you must specify the JDBC URL in the **SQL Endpoint JDBC URL** field in the Databricks Delta connection properties. After you test the connection, remove the **SQL Endpoint JDBC URL** value. Otherwise, when you define an application ingestion task that uses the connection, a design-time error occurs because Mass Ingestion Applications tries to use the JDBC URL as well as the required **Databricks Host**, **Cluster ID**, **Organization ID**, and **Databricks Token** values to connect to target, resulting in login failures.

- Processing of Rename Field operations on Databricks Delta target tables, without the need to rewrite the underlying Parquet files, requires the Databricks Delta Column Mapping feature with Databricks Runtime 10.2 or later. If you set the Rename Field option to **Replicate** on the **Schedule and Runtime Options** page in the task wizard, you must alter the generated target table to set the following Databricks table properties after task deployment and before you run the job:

```
ALTER TABLE <target_table> SET TBLPROPERTIES (  
  'delta.columnMapping.mode' = 'name',  
  'delta.minReaderVersion' = '2',  
  'delta.minWriterVersion' = '5')
```

These properties enable the Databricks Delta Column Mapping feature with the required reader and writer versions. If you do not set these properties, the application ingestion job will fail.

- Application ingestion jobs that have Databricks Delta targets can get schema information for generating the target tables from Databricks Unity Catalog. To enable access to information in a Unity Catalog, specify the catalog name in the **Catalog Name** field in the Databricks Delta connection properties. The catalog name is appended to the **SQL Warehouse JDBC URL** value for a data warehouse.

**Note:** Catalog use is optional for SQL warehouses and does not apply to job clusters.

If you use Unity Catalog, a personal storage location is automatically provisioned. To use the personal staging location, in the **Staging Environment** field of the connection properties, select **Personal Staging Location**. The Parquet data files for ingestion jobs can then be staged to the local personal storage location, which has a data retention period of 7 days. By default, the staging location is `stage://tmp/<user_name>` in the root AWS or Azure location. The `<user_name>` is taken from the **Database Token** connection property. This user must have read and write access to the personal staging location.

- If you generated Databricks Delta unmanaged tables as the target tables and no longer need a target unmanaged table, use a SQL DROP TABLE statement to delete the table from the target database. You must not manually delete the external directory in Amazon S3 or Azure Data Lake Storage for the unmanaged table. If you do so and try to deploy another job that uses that table, the deployment fails with a Metadata Handler error.

## Guidelines for Google BigQuery targets

Consider the following guidelines when you use Google BigQuery targets:

- When you use a Google BigQuery target for the first time, perform the following steps before you configure an application ingestion task for the target:
  1. Download the Google BigQuery JDBC driver version 1.2.25.1029 from the [Google Cloud](#) website.
  2. Copy the JDBC driver jar files to the following directory:
 

```
<Secure_Agent_installation_directory>/apps/Database_Ingestion/ext/
```
  3. Restart the Secure Agent.
- You must have a service account in your Google account to access Google BigQuery and Google Cloud Storage.
- Ensure that you have the client\_email, project\_id, private\_key, and region ID values for the service account. Enter the values in the corresponding **Service Account ID**, **Project ID**, **Service Account Key**, and **Region ID** connection properties when you create a Google BigQuery connection.
- If you want to configure a timeout interval for a Google BigQuery connection, specify the timeout interval property in the **Provide Optional Properties** field of the connection properties. Use the following format:
 

```
"timeout": "<timeout_interval_in_seconds>"
```
- You must have read and write access to the following entities:
  - Google BigQuery datasets that contain the target tables.
  - Google Cloud Storage path where Mass Ingestion Applications creates the staging file.
- Application ingestion jobs configured for Google BigQuery targets do not replicate the modification and renaming of source fields on the target.
- For incremental load tasks, you must enable source database Change Data Capture (CDC) on all source fields.
- You must have the following permissions to write data to a Google BigQuery table:
  - bigquery.datasets.get
  - bigquery.datasets.getIamPolicy
  - bigquery.models.\*
  - bigquery.routines.\*
  - bigquery.tables.create
  - bigquery.tables.delete
  - bigquery.tables.export
  - bigquery.tables.get
  - bigquery.tables.getData
  - bigquery.tables.list
  - bigquery.tables.update
  - bigquery.tables.updateData

- bigquery.tables.updateTag
- resourcemanager.projects.get
- resourcemanager.projects.list
- bigquery.jobs.create
- When you deploy application ingestion jobs, Mass Ingestion Applications generates Google BigQuery target tables clustered by primary key or unique key columns, by default. Each key column must have one of the following data types that Google BigQuery supports for clustering:
  - STRING
  - INT64
  - NUMERIC
  - BIGNUMERIC
  - DATE
  - DATETIME
  - TIMESTAMP
  - BOOL
  - GEOGRAPHY

If any column in the primary key or unique key has an unsupported data type, that column is skipped during clustering. For example, if the primary key contains the C1, C2, C3, C4, C5 columns and C2 has an unsupported data type, the target table is created with the C1, C3, C4, and C5 columns in the CLUSTER BY clause.

## Guidelines for Microsoft Azure Synapse Analytics targets

Consider the following guidelines when you use Microsoft Azure Synapse Analytics targets:

- To deploy and run an application ingestion task with a Microsoft Azure Synapse Analytics target, the target connection must specify a database user who has the CONTROL permission on the target database. To grant the CONTROL permission to the user, use the following SQL statements:

```
USE database_name;
GRANT CONTROL TO user_name;
```

The CONTROL permission is required for initial load, incremental load, and combined initial and incremental load jobs. The permission allows Mass Ingestion Applications to create target tables and database objects such as external data source, external file format, and database scoped credential objects if they do not exist in the database. The CONTROL permission is specifically required for creating external data source and database scoped credential objects.

**Note:** You must manually create the master key. To create the master key, you must have the CONTROL permission on the database.

- Application ingestion jobs first send data to a Microsoft Azure Data Lake Storage Gen2 staging file before writing the data to Microsoft Azure Synapse Analytics target tables. The staging file uses the hexadecimal x1d separator as the field delimiter. After the data is written to the target, the data stored in the table-specific directory that includes the staging files are deleted.
- If you use Microsoft Azure Data Lake Storage Gen2 with a Microsoft Azure Synapse Analytics connection, you must enable the **Hierarchical namespace** option in Microsoft Azure Data Lake Storage. With this setting, blob storage is not recommended.

- When you configure an application ingestion task for a Microsoft Azure Synapse Analytics target, ensure that each source object that you select for replication meets the following criteria:
  - The object must not contain more than 1024 fields and the size of each field must be less than 500 KB.
  - The object must not contain any record that is greater than 1 MB in size.
  - The object must not contain more than 32 primary keys.
  - The cluster index key must not contain data that is greater than 900 bytes in size.
  - The primary keys of the object must be of a data type that Microsoft Azure Synapse Analytics supports for primary keys.
- Incremental load jobs and combined initial and incremental load jobs generate a recovery table named `INFORMATICA_CDC_RECOVERY` on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema of the target tables.
- After an application ingestion job loads data to a Microsoft Azure Synapse Analytics target by using external tables, the job does not drop the log tables and external tables created on the target, even though these tables might be re-created when the job starts again.
- Application ingestion jobs configured for Microsoft Azure Synapse Analytics targets do not replicate the renaming of source fields on the target.

## Guidelines for Microsoft Azure SQL Database targets

Consider the following guidelines when using SAP (using SAP Mass Ingestion connector for initial load only) and Salesforce sources and Microsoft Azure SQL Database target for all load types for instances such as, Microsoft Amazon RDS (Relational Database Service) for SQL Server, Microsoft Azure SQL Database, and Azure SQL Managed Instance:

- The SQL Server JDBC driver is delivered with Mass Ingestion Applications. You do not need to install it separately.
- The Mass Ingestion Applications user requires following database roles, at minimum, to create target tables and write data to the tables:
  - `db_datareader`
  - `db_datawriter`
  - `db_ddladmin`
- In Administrator, when you define a SQL Server connection in Administrator for connecting to a SQL Server target, complete the required properties only:
  - SQL Server Version. Select either **SQL Server 2017** or **SQL Server 2019**.
  - Authentication Mode. Select **SQL Server Authentication** or **Windows Authentication v2**.
  - User Name
  - Password
  - Host
  - Port
  - Database Name
  - Schema
  - Code Page

Other properties are not supported.

- Application ingestion incremental load and initial and incremental load jobs with a SQL Server target generate a LOG table based on the target table schema, together with a number of additional metadata columns. The LOG table is created right before the change data is flushed to the target. The incoming DML data is inserted to the LOG table by supplying a local CSV file to Bulk Copy API of the SQL Server driver. A merge apply statement is generated based on the information in the LOG table and the DML operations are applied to the actual target table. When this is complete, the LOG table is dropped. The LOG table might cause a temporary spike in additional space or size requirements in the customer database instance when running multiple jobs or a job with multiple tables. The space and size required by the LOG table depends on the number of rows received as part of a flush cycle.
- The number of fields in a source object that an application ingestion incremental load or an initial and incremental load job can propagate to a SQL Server target must not exceed 508 fields. If a source object contains more than 508 fields, the job fails while creating the LOG table.
- Schema drift options are not supported for SQL Server targets in application ingestion incremental load and combined initial and incremental load jobs.
- Application ingestion incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information that prevents jobs restarted after a failure from propagating previously processed data again. This recovery table is generated in the same schema as the target tables.

## Guidelines for Oracle targets

By default, Mass Ingestion Applications disables logging for the Oracle target tables to optimize performance. You can enable logging by setting the writerOracleNoLogging custom property to false on the **Target** page in the application ingestion task wizard.

## Guidelines for PostgreSQL targets

Consider the following guidelines when you use PostgreSQL targets:

- To deploy and run an application ingestion task that includes a PostgreSQL target, the target connection must specify a database user who has the required privileges.  
The user must have the CONNECT and TEMPORARY privileges for the database specified in the connection and the USAGE and CREATE privileges for the target schema specified in the target properties.  
Use the following SQL statements to grant these privileges to a user role and then assign the role to a user:

```
CREATE ROLE dbmi_role;
GRANT CONNECT ON DATABASE database TO dbmi_role;
GRANT TEMPORARY ON DATABASE database TO dbmi_role;
GRANT CREATE ON SCHEMA schema TO dbmi_role;
GRANT USAGE ON SCHEMA schema TO dbmi_role;
CREATE USER dbmi_user with PASSWORD 'password';
GRANT dbmi_role to dbmi_user;
```

**Note:** Privileges on the target tables that are generated when the job runs are granted to the user who runs the job.

- Application ingestion incremental load jobs with a PostgreSQL target generate a LOG table based on the target table schema, with some additional metadata columns. The LOG table is created right before change data is flushed to the target. The incoming DML data is inserted to the LOG table by supplying a local CSV file to the Bulk Copy API of the PostgreSQL driver. A set of merge apply statements is generated based on the information in the LOG table, and then the DML operations are applied to the actual target table. After the DML changes are applied, the LOG table is dropped.

The LOG table might cause a temporary spike in additional space or size requirements in the customer database instance if you run multiple jobs or a job with multiple tables. The space and size required by the LOG table depends on the number of rows received as part of a flush cycle.

- The number of fields in a source object that an application ingestion incremental load job can propagate to a PostgreSQL target must not exceed 796 fields. If a source object contains more than 796 fields, the job fails while creating the LOG table.
- PostgreSQL supports a maximum length of 63 characters for source object identifiers. The deployment of application ingestion tasks that have a PostgreSQL target will fail during validation if the length of any source table or column name exceeds 63 characters.
- Schema drift options are not supported for PostgreSQL targets in application ingestion jobs.
- Application ingestion incremental load jobs generate a recovery table named `INFORMATICA_CDC_RECOVERY` on the target to store internal service information that prevents jobs restarted after a failure from propagating previously processed data again. This recovery table is generated in the same schema as the target tables.

## Guidelines for Snowflake targets

### Target preparation

Target preparation varies depending on whether you use the Superpipe feature for high performance streaming of data to Snowflake target tables or write data to intermediate stage files.

#### With Superpipe

If you use the Superpipe feature, complete the following steps:

1. Create a Mass Ingestion user. Use the following SQL statement:

```
create user INFACMI_User password = 'Xxxx@xxx';
```

2. Create a new user role and grant it to the Mass Ingestion user. Use the following SQL statements:

```
create role INFACMI_superpipe;  
grant role INFACMI_superpipe to user INFACMI_User;
```

3. Grant usage on the Snowflake virtual warehouse to the new role. Use the following SQL statement:

```
grant usage on warehouse warehouse_name to role INFACMI_superpipe;
```

4. Grant usage on the Snowflake database to the new role. Use the following SQL statement:

```
grant usage on database INFACMI_DB1 to role INFACMI_superpipe;
```

5. Create a new schema. Use the following SQL statements:

```
use database INFACMI_DB1;  
create schema sh_superpipe;
```

6. Grant create stream, create view, and create table privileges on the new Snowflake schema to the new role. Use the following SQL statement:

```
grant create stream, create view, create table, usage on schema  
INFACMI_DB1.sh_superpipe to role INFACMI_superpipe;
```

7. Set the default role for the newly created user. Use the following SQL statement:

```
alter user INFACMI_User set default_role=INFACMI_superpipe;
```

8. Define a Snowflake Data Cloud connection to the target. You must use the **KeyPair** option as the authentication method. See Connectors and Connections > Snowflake Data Cloud connection properties.

9. Generate a private key with OpenSSL version 3.x.x and the PBE-SHA1-2DES or PBE-SHA1-3DES cipher. Use the following openssl commands to generate and format the private key:

```
openssl genrsa 2048 | openssl pkcs8 -topk8 -inform PEM -v1 PBE-SHA1-3DES -out  
rsa_key.p8
```

10. Generate the public key. Use the following openssl command, in which the -in option references the file (rsa\_key.p8) that contains the encrypted private key:

```
openssl rsa -in rsa_key.p8 -pubout -out rsa_key.pub
```

11. In Snowflake, assign the public key to the Snowflake user. Use the following SQL command:

```
alter user INFACMI_User set rsa_public_key='key_value';
```

**Next step:** When you create an ingestion task, select the **Superpipe** option on the **Target** page of the task wizard. You can also optionally specify a **Merge Frequency** value, which controls the frequency at which change data rows are merged and applied to the Snowflake target table.

### Without Superpipe

If you do NOT use the Superpipe feature for Snowflake targets, complete the following steps as the ACCOUNTADMIN user:

1. Create a Mass Ingestion user. Use one of the following SQL statements:

```
create user INFACMI_User password = 'Xxxx@xxx';
```

or

```
replace user INFACMI_User password = 'Xxxx@xxx';
```

2. Create a new role and grant the role to the Mass Ingestion user. Use the following SQL statements:

```
create role INFA_CMI_Role;  
grant role INFA_CMI_Role to user INFACMI_User;
```

3. Grant usage on the Snowflake virtual warehouse to the new role. Use the following SQL statement:

```
grant usage on warehouse CMIWH to role INFA_CMI_Role;
```

4. Grant usage on the Snowflake database to the new role. Use the following SQL statement:

```
grant usage, CREATE SCHEMA on database CMIDB to role INFA_CMI_Role;
```

5. Set the default role for the newly created user. Use the following SQL statement:

```
alter user INFACMI_User set default_role=INFA_CMI_Role;
```

Also, as the INFACMI\_User, create a new schema:

```
create schema CMISchema;
```

**Note:** If the user's default role is used for ingestion tasks and does not have the required privileges, the following error will be issued at runtime:

```
SQL compilation error: Object does not exist, or operation cannot be performed.
```

### Usage guidelines

- Mass Ingestion provides alternative methods of moving data to a Snowflake Data Cloud target:
  - If you select the **Superpipe** option when defining an ingestion task, the ingestion job uses the Snowpipe Streaming API to stream rows of data directly to the target tables with low latency. This method is available for all load types. You must use **KeyPair** authentication.
  - If you do not use Superpipe, ingestion jobs first write the data to data files in an internal stage, which has the name you specify in the task definition.
- If you do not use Superpipe and the internal stage specified in the target properties for an ingestion job does not exist, Mass Ingestion Databases automatically creates the stage by running the following SQL command:

```
Create stage if not exists "Schema"."Stage_Bucket"
```

For the command to run successfully, the following privilege must be granted to your user role:

```
GRANT CREATE STAGE ON SCHEMA "Schema" TO ROLE <your_role>;
```

- When you define a connection for a Snowflake target, you must set the **Additional JDBC URL Parameters** field to `database=target_database_name`. Otherwise, when you try to define the target in the application ingestion task wizard, an error message indicating that the list of schemas cannot be retrieved appears.
- When you define a connection for a Snowflake target using the **KeyPair** option as the authentication method and you generate the private key with OpenSSL 3.x.x version, use PBE-SHA1-2DES or PBE-SHA1-3DES cipher while generating the private key. Run one of the following commands:

```
openssl genrsa 2048 | openssl pkcs8 -topk8 -inform PEM -v1 PBE-SHA1-3DES -out
rsa_key.p8
```

or:

```
openssl genrsa 2048 | openssl pkcs8 -topk8 -inform PEM -v1 PBE-SHA1-2DES -out
rsa_key.p8
```

If you use a generic command without the PBE-SHA1-2DES or PBE-SHA1-3DES cipher, an error message about an invalid or unsupported private key might occur while fetching the target schema in the target definition step of the application ingestion task wizard.

The error message does not occur if you use OpenSSL 1.1.1 to generate the private key.

- Incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema that contains the target tables.
- The object must not contain any record that is greater than 16 MB.
- For Snowflake targets, you cannot alter the scale of NUMBER fields or change the data type of an existing field to a different data type because Snowflake does not support these actions.

## Configure private connectivity to Snowflake

You can access Snowflake using AWS or Azure Private Link endpoints.

The AWS or Azure Private Link setup ensures that the connection to Snowflake uses the AWS or Azure internal network and does not take place over the public Internet.

To connect to the Snowflake account over the private AWS network, see [AWS Private Link and Snowflake](#).

To connect to the Snowflake account over the private Azure network, see [Azure Private Link and Snowflake](#).

## Avro data types

Mass Ingestion Applications supports only some of the primitive and logical data types that Avro schemas provide.

A primitive data type is a type that allows you to represent a single data value. A logical type is an Avro primitive or complex type with extra attributes to represent a derived type. This topic applies to all target types that support Avro or Parquet output format.

The following table lists the primitive Avro data types that Mass Ingestion Applications supports:

Primitive data type	Description
INT	32-bit signed integer
LONG	64-bit signed integer
FLOAT	Single precision (32-bit) IEEE 754 floating-point number

Primitive data type	Description
DOUBLE	Double precision (64-bit) IEEE 754 floating-point number
BYTES	Sequence of 8-bit unsigned bytes
STRING	Unicode character sequence

The following table lists the logical Avro data types that Mass Ingestion Applications supports:

Logical data type	Description
DECIMAL	An arbitrary-precision signed decimal number of the form $\text{unscaled} \times 10^{-\text{scale}}$
DATE	A date, without reference to a time or time zone.
TIME	A time of day that has the precision of 1 millisecond or 1 microsecond, without reference to a time zone or date.
TIMESTAMP	A date and time value that has the precision of 1 millisecond or microsecond, without reference to a particular calendar or time zone.

## Handling source schema changes

You can configure Mass Ingestion Applications to automatically detect source schema changes, also called *schema drift*, and handle these changes on the target. This feature is available only for the incremental load and combined initial and incremental load tasks.

When you configure an application ingestion task, on the **Schedule and Runtime Options** page of the application ingestion task wizard, you can specify the types of source schema changes that Mass Ingestion Applications must propagate for the job associated with the task. You can also specify how the job must handle each type of source schema change. For example, you can configure the task to ignore the changes, replicate them, or stop the job when a particular type of schema change occurs on the source. For more information, see [“Configuring schedule and runtime options” on page 138](#).

**Note:** In combined initial and incremental load jobs, Mass Ingestion Applications starts detecting and replicating source schema changes only after the source object reaches the Normal state.

The following table describes the types of schema changes that Mass Ingestion Applications can detect for each source type:

Source type	Supported schema changes
Adobe Analytics	<ul style="list-style-type: none"> <li>- Add field</li> <li>- Modify field</li> <li>- Drop field</li> <li>- Rename field</li> </ul>
Google Analytics	<ul style="list-style-type: none"> <li>- Add column</li> <li>- Modify column</li> <li>- Drop column</li> <li>- Rename column</li> </ul>
Marketo	Not supported
Microsoft Dynamics 365	<ul style="list-style-type: none"> <li>- Add column</li> <li>- Modify column</li> <li>- Drop column</li> </ul>
NetSuite	<ul style="list-style-type: none"> <li>- Add column</li> <li>- Modify column</li> </ul>
Oracle Fusion Cloud	<ul style="list-style-type: none"> <li>- REST - Not supported</li> <li>- BICC - Not supported</li> </ul>
Salesforce	<ul style="list-style-type: none"> <li>- Add field</li> <li>- Modify field</li> <li>- Drop field</li> <li>- Rename field</li> </ul>
SAP	<ul style="list-style-type: none"> <li>- Using SAP ODP Extractor connector - Not supported</li> <li>- Using SAP Mass Ingestion connector - Not supported</li> </ul>
ServiceNow	<ul style="list-style-type: none"> <li>- Add column</li> <li>- Modify column</li> <li>- Drop column</li> </ul>
Workday	<ul style="list-style-type: none"> <li>- SOAP - Not supported</li> <li>- RaaS - Not supported</li> </ul>
Zendesk	Not supported

Mass Ingestion Applications detects a schema change in a source object only after Data Manipulation Language (DML) operations occur on the altered source object. If multiple schema changes occur without intervening DML operations, Mass Ingestion Applications detects all the schema changes together when a DML operation occurs.

**Note:**

- Application ingestion jobs do not replicate source changes that add, remove, or modify primary key or unique key constraints. If these types of changes occur on the source, you must re-synchronize the target tables.
- The application ingestion jobs configured for Microsoft Azure Synapse Analytics targets do not replicate the renaming of source fields on the target.

- The application ingestion jobs configured for Google BigQuery targets do not replicate the modification and renaming of source fields on the target.
- The application ingestion jobs configured for Snowflake targets support modify operations on source columns with the following limitations:
  - Snowflake targets cannot modify the scale of NUMBER columns.
  - Snowflake targets do not support changing the data type of an existing column to a different data type.
- If you try to replicate an unsupported schema change type on the target, the application ingestion jobs associated with the task will fail with an error.
- If you configured schema drift options to stop the job when Mass Ingestion Applications detects a schema change, you can use the **Resume With Options** command to resume the job with an override schema drift option.

## Ability to apply deletes as soft deletes on the target

For application ingestion incremental load and combined initial and incremental load jobs that have any supported source type and a Databricks Delta or Snowflake target, you can configure the task to process delete operations on the source as soft deletes on the target.

A soft delete marks a deleted row as deleted without actually removing it from the database. The row is applied to the target with the value of "D" in the generated INFA\_OPERATION\_TYPE metadata column.

**Important:** You can use Soft Deletes only if all of the source objects have primary keys and the source does not allow the primary key values to change in any row after the row is first created. If the primary key values change, duplicate rows might be written to the target, causing the target to become corrupted.

Example scenario: Your organization wants to use soft deletes in a data warehouse to mark the rows that were deleted at the source while still retaining the rows for audit purposes.

To enable soft deletes, set the **Apply Mode** field to **Soft Deletes on Target** page of the task wizard when you configure the ingestion task.

## Apply an audit history of all source table change operations to target tables

You can configure application ingestion incremental load and combined initial and incremental load tasks that have Snowflake targets to write an audit trail of every DML change operation made on the source tables to the target. A row for each DML change on a source table is written to the generated target table along with the audit columns you select. The audit columns contain metadata about the change, such as the DML operation type, time, owner, transaction ID, generated ascending sequence number, and before image.

When you define the task, select **Audit** in the **Apply Mode** field on the **Target** page. The **Apply Mode** field is available for new or undeployed tasks.

To specify the audit metadata columns to add, select one or more of the check boxes under **Advanced** on the **Target** page:

- **Add Last Replicated Time.** Adds a column that records the timestamp at which a record was inserted or last updated in the target table. For initial loads, all loaded records have the same timestamp, except for Snowflake targets that use the Superpipe option where the seconds or minutes might vary slightly. For incremental and combined initial and incremental loads, the column records the timestamp of the last DML operation that was applied to the target.
- **Add Operation <metadata\_type>.** Adds columns that contain metadata for change operations, such as the DML operation type, time, owner, transaction ID, and generated ascending sequence number. The columns are populated when data is loaded to the target tables.
- **Add Before Images.** Adds \_OLD columns that contain before-image data for Updates. You can compare the old and new column values in the tables.
- **Prefix for Metadata Columns.** Add a prefix to the names of the added audit columns to differentiate them from other table columns. Default is INFA\_.

These fields are optional. Only the **Add Operation Type** check box is selected by default to add a column that shows the DML operation type: D (delete), or U (update).

The first time you run a job associated with the task, the job generates the target tables with the selected audit metadata columns. Ensure that no constraints other than indexes exist on the target tables.

For targets that support **Soft Deletes** mode, including Snowflake targets, for each delete operation processed as a soft delete, the operation type of "D" appears in the INFA\_OPERATION\_TYPE column and values are written to any other metadata columns that you selected. However, for update and insert operations, the INFA\_OPERATION\_TYPE column and all other selected metadata columns are NULL.

#### Example for Audit mode

For example, assume the following DML change operations occur on a source table in the order shown:

```
Insert into tableA pkey = 1
Update tableA where pkey=1
Update tableA where pkey=1
Delete from tableA where pkey = 1
```

All of the following rows appear in the target table, providing an audit trail of all of the source DML changes:

```
opType=I, pkey=1...
opType=E, pkey=1...
opType=E, pkey=1...
opType=D, pkey=1...
```

In this example, the only audit column selected is opType.

Note that when the task's apply mode is Standard, none of these rows appears on the target table because the last DML operation is a Delete, which supersedes the prior changes.

**Note:** If a combined initial and incremental load job captures an incremental insert change record during the initial unload phase, the job manufactures a delete for the same row to remove any duplicate that might have been obtained from the initial unload. This manufactured activity will be reflected in audit apply mode.

# Configuring application ingestion tasks

In Data Integration, use the application ingestion task wizard to configure application ingestion tasks.

On the wizard pages, complete the following configuration tasks:

1. Define basic task information, such as the task name, project location, runtime environment, and load type.
2. Configure the source.
3. Configure the target.
4. Configure the task schedule and runtime options.

Click **Next** or **Back** to navigate from one page to another. At any point, you can click **Save** to save the information that you have entered until then.

After you complete all the wizard pages, save the information and then click **Deploy** to make the task available as an executable job to the Secure Agent.

## Before you begin

Before you configure an application ingestion task, complete the following prerequisite tasks in Administrator:

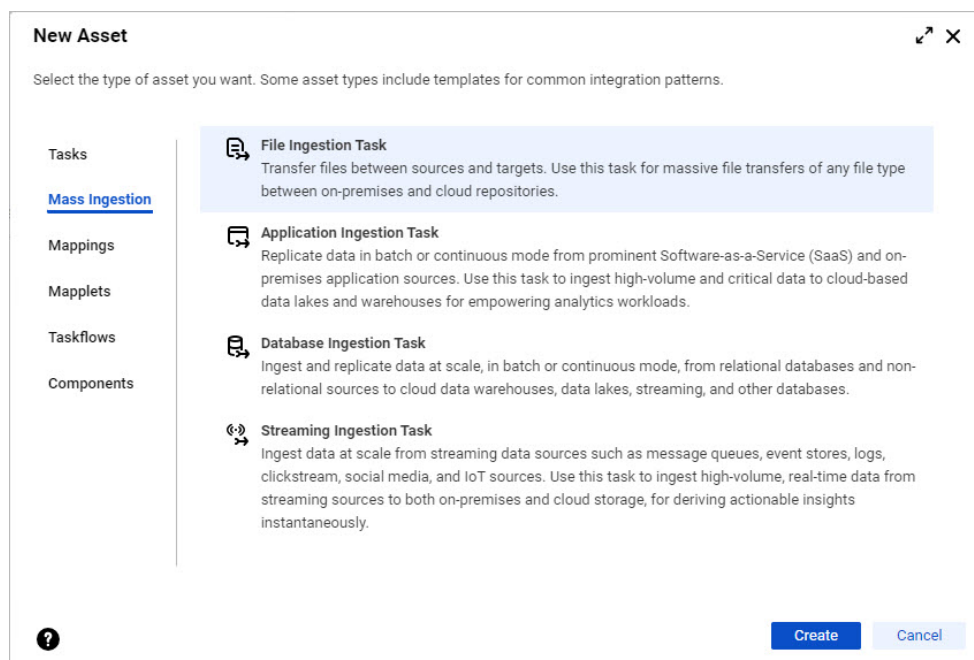
- Verify that the Secure Agent in your runtime environment is running and you can access the Mass Ingestion service.
- Define the source and target connections.

## Defining basic task information

To define an application ingestion task, you must first enter some basic information about the task, such as the task name, project or project folder location, and load operation type.

1. Start the task wizard in one of the following ways:
  - On the Home page, click the **Ingest** panel and select **Application Ingestion Task**.

- In the navigation bar on the **Explore** page or the Home page, click **New** to open the **New Asset** dialog box. Then, select **Mass Ingestion > Application Ingestion Task** and click **Create**.



The **Definition** page of the application ingestion task wizard appears.

2. Configure the following properties:

Property	Description
Name	<p>Name of the application ingestion task.</p> <p>The name of the application ingestion task must be unique within the organization. The name can contain alphanumeric characters, spaces, periods (.), commas (,), underscores (_), plus signs (+), and hyphens (-).</p> <p>Task names are not case sensitive. The maximum length is 50 characters.</p> <p><b>Note:</b> If you include spaces in the name of an application ingestion task, the spaces do not appear in the name of the job associated with the task.</p>
Location	Project or folder in which you want to store the task.
Runtime Environment	<p>Runtime environment in which you want to run the task.</p> <p>You can run an application ingestion task only on a Secure Agent. The runtime environment can include a Secure Agent Group with only one agent.</p> <p><b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.</p>

Property	Description
Description	A brief description of the task. Maximum length is 4000 characters.
Load Type	Type of load operation that you want the application ingestion task to perform. You can select one of the following load types for the task: <ul style="list-style-type: none"> <li>- <b>Initial Load:</b> Loads data read at a specific point in time from the source application to the target in a batch operation. You can perform an initial load to materialize a target to which incremental change data will be sent.</li> <li>- <b>Incremental Load:</b> Propagates source data changes to a target continuously or until the job is stopped or ends. The job propagates the changes that have occurred since the last time the job ran or from a specific start point for the first job run.</li> <li>- <b>Initial and Incremental Load:</b> Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source objects on a continuous basis.</li> </ul>

3. Click **Next**.

## Configuring the source

You can configure the source on the **Source** page of the application ingestion task wizard.

Before you configure the source, ensure that the connection to the source is created in Administrator for the runtime environment that your organization uses.

1. From the **Connection** list, select the connection configured for the source application. The connection type appears in parentheses after the connection name.  
The list includes only the connections that are valid for the load type that you selected on the **Definition** page.

**Note:** After you deploy the ingestion task, you cannot change the connection without undeploying the associated ingestion job. After you change the connection, you must deploy the task again.

2. Based on the type of source that you want to configure, perform the steps described in the following topics:

- [“Configuring an Adobe Analytics source” on page 64](#)
- [“Configuring a Google Analytics source” on page 67](#)
- [“Configuring a Marketo source” on page 70](#)
- [“Configuring a Microsoft Dynamics 365 source” on page 73](#)
- [“Configuring a NetSuite source” on page 75](#)
- [“Configuring an Oracle Fusion Cloud source” on page 78](#)
- [“Configuring a Salesforce source” on page 81](#)
- [“Configuring a Salesforce Marketing Cloud source” on page 85](#)
- [“Configuring an SAP source with SAP ODP Extractor connector” on page 88](#)
- [“Configuring an SAP source with SAP Mass Ingestion connector” on page 91](#)
- [“Configuring a ServiceNow source” on page 96](#)
- [“Configuring a Workday source” on page 99](#)
- [“Configuring a Zendesk source” on page 102](#)

## Configuring an Adobe Analytics source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Adobe Analytics source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Path to Report Configuration File** field, enter the path to the JSON file that contains the report configurations.
2. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.

The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source objects to replicate.

3. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.

4. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

5. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.

- If you selected **Select All**, the list of objects is view only.
- If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Deselect any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.

**Note:** The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.

For each object, you can view a list of field names and data types. Click the highlighted number of fields in the **Fields** column to list the fields to the right.

To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

6. To download a list of source objects that match the selection rules, perform the following steps:
  - a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.
  - b. If you want to include the fields in the list, select **Include Fields**.
  - c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.

The information in the downloaded

is in the following format:

*status, object\_name, object\_type, field\_name, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The object is included for processing.</li> <li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li> </ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates an object.</li> <li>- <b>F</b>: Indicates a field.</li> </ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

7. Expand the **Advanced** section.
8. For initial load and combined initial and incremental load tasks, specify the date and time when the ingestion job should start replicating the source data.  
**Note:** The date and time must be in the time zone specified for ReportSuiteID in the JSON file with report configurations
9. For initial load tasks, specify the date and time when the ingestion job should stop replicating the source data.  
**Note:** The date and time must be in the time zone specified for ReportSuiteID in the JSON file with report configurations
10. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Coordinated Universal Time (UTC).
11. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 1 day.
12. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. The default value is 50000.
13. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
14. Click **Next**.

## Configuring a Google Analytics source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Google Analytics source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Account ID** field, enter the unique identifier of your Google Analytics service account.
2. In the **Property ID** field, enter the unique identifier of the property whose data you want to replicate.
3. In the **View ID** field, enter the unique identifier of the view whose data you want to replicate.
4. In the **Path to Report Configuration File** field, enter the path to the JSON file that contains the report configurations.
5. In the **Report Selection** section, select **Select All** only if you want to select all source reports for data replication. You cannot edit the selection in subsequent dimensions & metrics.

The **Reports Selected** field shows the count of all selected reports. If you have many source reports, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source reports to replicate.

6. If you selected **Rule-based Selection**, create the rules to select the source reports that you want to replicate on the target.

By default, an *Include* rule configured to select all source reports is defined in the task. If you do not want to replicate all the source reports, you can define additional *Include* rules and *Exclude* rules to select the specific reports that you want to replicate.

Perform the following steps to create a report selection rule:

- a. Select **Report Selection** as the rule type.
- b. From the adjacent list, select **Include** or **Exclude** as the action that you want the rule to perform.
- c. In the condition field, enter a report name or a report-name mask to specify the source reports that you want to include in or exclude from the list of selected reports.

### Notes:

- A mask can contain the asterisk (\*) wildcard character to represent one or more characters, the question mark (?) wildcard character to represent a single character, or both types of wildcard characters. You can use a wildcard character multiple times in a report-name mask.
- Report names are case sensitive. When you define the report selection rules, you must specify the report names or masks in the case in which they are defined on the source.
- If a report name contains delimiters, such as quotation marks or brackets, do not include them when you specify the report name in the rule.
- If a report name includes special characters, such as backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), replace each special character in the name with a backslash (\) when you specify the report name in the rule.

- d. Click **Add Rule**.

The rule appears in the **Rules** list.

To refine the selection, you can define additional *Include* rules and *Exclude* rules. The report selection rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed. For an example of using multiple rules, see [“Example of rules for selecting source objects” on page 104](#).

After you create the rules, you can click **Report Count** to display the number of source reports that match each rule in the **Objects Affected** column and the total number of reports selected based on all the selection rules in the **Total Reports Selected** field.

- e. To preview the reports to be selected based on all rules, click **Apply Rules**.

The reports are listed on the **Selected Reports** tab. The list shows the report names and column count.

To search for tables and columns, you can either browse the list of objects or enter a search string in the **Find** box.

**Tip:** Click the Refresh icon next to the **Updated** date to refresh the total reports count and the list of reports selected based on the current rules. You can check the results of new rules in this manner. Click the Settings icon to control the line spacing in the list of reports, from Comfortable (most spacing) to Compact (least spacing).

7. To perform trim actions on the columns of the source reports that were selected based on rules, create column action rules.

Perform the following steps to create a column action rule:

- a. Select **Column Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character column values.
  - **RTRIM**. Trims spaces to the right of character column values.
  - **TRIM**. Trims spaces to the left of and to the right of character column values.
- c. In the condition field, enter a column name or a column-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against columns of the selected source reports to identify the columns to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The column action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

8. Under **Report View**, view the selected reports, including the number of rules in each report.

- If you selected **Select All**, the list of reports is view only.
- If you applied rules, you can refine the set of selected reports by clicking the check box next to individual report. Deselect any report that you do not want to replicate, or select additional reports to replicate. Click the Refresh icon to update the selected reports count.

**Note:** The first time you change a check box setting for a report, the rules are no longer in effect. The selections under **Report View** take precedence. However, if you click the Add Rule (+) icon again, the reports that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Report View** list, click **Apply Rules** again.

For each report, you can view a list of columns. Click the highlighted number of columns in **Columns** to list the fields to the right.

To search for objects and fields, in the drop-down list above **Columns**, select **Report Name**, **Columns**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

9. To download a list of source reports that match the selection rules, perform the following steps:
  - a. From the **List Reports by Rule Type** list, select the type of selection rule for which you want to download the list of selected source reports.
  - b. If you want to include the columns in the list, select **Include Columns**.
  - c. Click the Download icon.

The list of source reports that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

*status,report\_name,report\_type,column\_name,comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source report from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The report is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The report is included for processing.</li><li>- <b>X</b>. The report is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the report is excluded.</li></ul>
report_name	Name of the source report.
report_type	Type of the source report. The possible values are: <ul style="list-style-type: none"><li>- <b>O</b>: Indicates a report.</li><li>- <b>F</b>: Indicates a column.</li></ul>
column_name	Name of the source column. This information appears only if you selected the <b>Include Columns</b> check box before downloading the list.
comment	Reason why a source report is excluded from processing even though it matches the selection rules.

10. Expand the **Advanced** section.
11. For initial load and combined initial and incremental load tasks, specify the date and time when the ingestion job should start replicating the source data.
12. For initial load tasks, specify the date and time when the ingestion job should stop replicating the source data.
13. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.

**Note:** You must specify the date in the time zone configured for the Google Analytics view.
14. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 1 day.
15. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. The default value is 50000.
16. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.

The custom properties are usually configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

17. Click **Next**.

## Configuring a Marketo source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Marketo source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.

The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source objects to replicate.

2. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.

3. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

4. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.
  - If you selected **Select All**, the list of objects is view only.
  - If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Deselect any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.

**Note:** The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.

For each object, you can view a list of field names and data types. Click the highlighted number of fields in the **Fields** column to list the fields to the right.

To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

5. To download a list of source objects that match the selection rules, perform the following steps:
  - a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.
  - b. If you want to include the fields in the list, select **Include Fields**.
  - c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

```
status,MARKETO,object_name,object_type,field_name,comment
```

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The object is included for processing.</li><li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li></ul>
MARKETO	Name of the source application.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"><li>- <b>O</b>: Indicates an object.</li><li>- <b>F</b>: Indicates a field.</li></ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

6. Expand the **Advanced** section.
7. For initial load tasks, in the **Start Date** field, specify the date on which the ingestion job associated with the application ingestion task starts reading records from lead and custom objects on the source.
8. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.

**Note:** You must specify the date and time in Coordinated Universal Time (UTC).

9. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
10. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.

The custom properties are usually configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

11. Click **Next**.

## Configuring a Microsoft Dynamics 365 source

On the **Source** page of the application ingestion task wizard, you can specify the tables that you want to ingest and configure the advanced properties for your Microsoft Dynamics 365 source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Table Selection** section, select **Select All** only if you want to select all source tables and columns for data replication. You cannot edit the selection in subsequent columns.

The **Tables Selected** field shows the count of all selected tables. If you have many source tables, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source tables to replicate.

2. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected tables by table under **Table View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source tables accessed with the selected connection. To see how many tables are selected by this rule, click the Refresh icon to display the table count in **Total Tables Selected** and click **Apply Rules** to see the table count in **Table View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Table Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Tables Affected** and **Total Tables Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the table counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Table View** list still lists tables. If you switch to **Select All**, the rules no longer appear.

3. To perform trim actions on the columns of the source tables that were selected based on rules, create column action rules.

Perform the following steps to create a column action rule:

- a. Select **Column Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character column values.
  - **RTRIM**. Trims spaces to the right of character column values.
  - **TRIM**. Trims spaces to the left of and to the right of character column values.
- c. In the condition field, enter a column name or a column-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against columns of the selected source tables to identify the columns to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The column action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

4. Under **Table View**, view the selected tables, including the number of columns in each table.
  - If you selected **Select All**, the list of tables is view only.
  - If you applied rules, you can refine the set of selected tables by clicking the check box next to individual table. Deselect any report that you do not want to replicate, or select additional tables to replicate. Click the Refresh icon to update the selected tables count.

**Note:** The first time you change a check box setting for a table, the rules are no longer in effect. The selections under **Table View** take precedence. However, if you click the Add Rule (+) icon again, the tables that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Table View** list, click **Apply Rules** again.

For each table, you can view a list of columns. Click the highlighted number of columns in **Columns** to list the fields to the right.

To search for tables and columns, in the drop-down list above **Columns**, select **Table Name**, **Columns**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

5. To download a list of source tables that match the selection rules, perform the following steps:
  - a. From the **List Tables by Rule Type** list, select the type of selection rule for which you want to download the list of selected source tables.
  - b. If you want to include the columns in the list, select **Include Columns**.
  - c. Click the Download icon.

The list of source tables that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

```
status, table_name, table_type, column_name, comment
```

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source table from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The table is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The table is included for processing.</li><li>- <b>X</b>. The table is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the table is excluded.</li></ul>
table_name	Name of the source table.
table_type	Type of the source object. The possible values are: <ul style="list-style-type: none"><li>- <b>O</b>: Indicates a table.</li><li>- <b>F</b>: Indicates a column.</li></ul>
column_name	Name of the source column. This information appears only if you selected the <b>Include Columns</b> check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

- For incremental load tasks and combined initial and incremental load tasks, expand the **Advanced** section.
- For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Coordinated Universal Time (UTC).
- For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
- In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
- Click **Next**.

## Configuring a NetSuite source

On the **Source** page of the application ingestion task wizard, you can specify the tables that you want to ingest and configure the advanced properties for your NetSuite source. You can also specify custom properties to address unique environments and special use cases.

- In the **Table Selection** section, select **Select All** only if you want to select all source tables and columns for data replication. You cannot edit the selection in subsequent columns.  
The **Tables Selected** field shows the count of all selected tables. If you have many source tables, the interface might take a long time to fetch them.  
Alternatively, you can use rules to define a subset of source tables to replicate.
- To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected tables by table under **Table View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source tables accessed with the selected connection. To see how many tables are selected by this rule, click the Refresh icon to display the table count in **Total Tables Selected** and click **Apply Rules** to see the table count in **Table View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Table Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.

- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Tables Affected** and **Total Tables Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the table counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Table View** list still lists tables. If you switch to **Select All**, the rules no longer appear.

3. To perform trim actions on the columns of the source tables that were selected based on rules, create column action rules.

Perform the following steps to create a column action rule:

- a. Select **Column Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character column values.
  - **RTRIM**. Trims spaces to the right of character column values.
  - **TRIM**. Trims spaces to the left of and to the right of character column values.
- c. In the condition field, enter a column name or a column-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against columns of the selected source tables to identify the columns to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The column action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

4. Under **Table View**, view the selected tables, including the number of columns in each table.

- If you selected **Select All**, the list of tables is view only.
- If you applied rules, you can refine the set of selected tables by clicking the check box next to individual table. Deselect any report that you do not want to replicate, or select additional tables to replicate. Click the Refresh icon to update the selected tables count.

**Note:** The first time you change a check box setting for a table, the rules are no longer in effect. The selections under **Table View** take precedence. However, if you click the Add Rule (+) icon again, the tables that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Table View** list, click **Apply Rules** again.

For each table, you can view a list of columns. Click the highlighted number of columns in **Columns** to list the fields to the right.

To search for tables and columns, in the drop-down list above **Columns**, select **Table Name**, **Columns**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

5. To download a list of source tables that match the selection rules, perform the following steps:

- From the **List Tables by Rule Type** list, select the type of selection rule for which you want to download the list of selected source tables.
- If you want to include the columns in the list, select **Include Columns**.
- Click the Download icon.

The list of source tables that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

*status, table\_name, table\_type, column\_name, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source table from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The table is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The table is included for processing.</li> <li>- <b>X</b>. The table is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the table is excluded.</li> </ul>
table_name	Name of the source table.
table_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates a table.</li> <li>- <b>F</b>: Indicates a column.</li> </ul>

Field	Description
column_name	Name of the source column. This information appears only if you selected the <b>Include Columns</b> check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

6. Expand the **Advanced** section.
7. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Greenwich Mean Time (GMT).
8. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
9. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. Default is 5000.
10. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
11. Click **Next**.

## Configuring an Oracle Fusion Cloud source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Oracle Fusion Cloud source. You can also specify custom properties to address unique environments and special use cases.

1. Select one of the following replication approaches:
  - Select **REST** to extract data from various applications of Oracle Fusion such as ERP, SCM, HCM, Sales, and Services, and transfer data to the target.
  - Select **BICC** (Business Intelligence Cloud Connector) to extract bulk data from the source to the target.
2. From the **Oracle Fusion Application** list, select the application from which you want to replicate data.
3. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.  
The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.  
Alternatively, you can use rules to define a subset of source objects to replicate.
4. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.  
When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.

- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.

5. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

6. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.
  - If you selected **Select All**, the list of objects is view only.
  - If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Deselect any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.

**Note:** The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.

For each object, you can view a list of field names and data types. Click the highlighted number of fields in the **Fields** column to list the fields to the right.

To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

7. To download a list of source objects that match the selection rules, perform the following steps:
  - a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.
  - b. If you want to include the fields in the list, select **Include Fields**.
  - c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

*status, object\_name, object\_type, field\_name, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The object is included for processing.</li> <li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li> </ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates an object.</li> <li>- <b>F</b>: Indicates a field.</li> </ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

To search for tables and columns, you can either browse the list of objects or enter a search string in the **Find** box.

8. Expand the **Advanced** section.
9. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in the time zone configured for the Oracle Fusion Cloud instance.
10. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
11. Select the **Include Child Objects** option to get the child object data of an object using an Oracle Fusion Cloud source. This applies only for the REST replication approach and for all load types only when the target is Google Big Query.
12. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. The default value is 50000.
13. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
14. Click **Next**.

## Configuring a Salesforce source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Salesforce source. You can also specify custom properties to address unique environments and special use cases.

1. For initial load tasks and combined initial and incremental load tasks, select the type of Salesforce API that you want to use to retrieve the source data.  
Options are:
  - **Standard (REST) API:** Replicates source fields of Base64 data type. Informatica recommends that you use the Bulk API 2.0 unless you want to ingest fields of Base64 data type or objects that are not supported by Bulk API 2.0 during initial loading of data. All incremental load activities use only the standard REST API.
  - **Bulk API 2.0:** Excludes replication of source fields of Base64 data type. Bulk API 2.0 is the default API for initial load tasks and the initial load of the combined initial and incremental load tasks.**Note:** By default, incremental load tasks can capture and replicate change data from source fields of Base64 data type.
2. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.  
The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.  
Alternatively, you can use rules to define a subset of source objects to replicate.
3. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.  
When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.

- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.

4. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:**

- You cannot use field action rules to trim the spaces in rich text area fields of Salesforce because Salesforce uses control characters instead of empty values to represent the spaces.
  - You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.
5. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.
- If you selected **Select All**, the list of objects are view only.
  - If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Clear any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.
  - You can also individually clear or reselect the fields in a selected source object. To view or change the fields from which data will be replicated for a selected object, click the highlighted number of fields in the **Fields** column. The field names and data types are displayed to the right. By default, all the fields are selected for a selected source object. To clear a column or reselect it, click the check box next to the field name. You can't clear a primary key column.

**Note:**

- To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.
  - The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.
  - If you rename a column while it is in an **Up and Running** state, the new name appears in the column selection, but it's not updated on the **View** tab.
  - If you select fields individually, the resulting set of fields is fixed and will not be updated by any schema change, regardless of the schema drift settings.  
  
For example, if a source field is added or renamed, that field is silently excluded from CDC processing because it's not in the list of selected fields. However, if a selected field is dropped on the source, the schema drift Drop Field option controls how it's handled. The dropped field operation is not reflected in the list of fields until you apply the rules again.
  - For combined initial and incremental load jobs, if you add to or modify a field selection, and redeploy the task, then a resync operation is triggered automatically. The resync is required to ensure that the target table has the same values as the source. For incremental load jobs, which do not have a resync option, modifying a field selection results in an error. If you modify the field selection of a deployed job and then redeploy it, the source and target will no longer match.  
  
If you remove a selected field, a resync operation is not triggered, and no error is reported. Removal of a selected field is treated the same as a drop field event, triggering your drop field schema drift settings for the task.
6. To download a list of source objects that match the selection rules, perform the following steps:
- a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.

- b. If you want to include the fields in the list, select **Include Fields**.
- c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.  
The information in the downloaded

is in the following format:

*status,object\_name,object\_type,field\_name,comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The object is included for processing.</li> <li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li> </ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates an object.</li> <li>- <b>F</b>: Indicates a field.</li> </ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

7. Expand the **Advanced** section.
8. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Greenwich Mean Time (GMT).
9. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
10. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. The default value for initial load operations is 50000 and the default value for incremental load operations is 2000.  
**Note:** For combined initial and incremental load tasks, you must specify the fetch size separately for initial load operations and incremental load operations.
11. For initial load and combined initial and incremental load tasks, select **Include Archived and Deleted Rows** to replicate the archived and soft-deleted rows from the source during the initial loading of data.
12. For initial load and combined initial and incremental load tasks, select **Enable Partitioning** to partition the source objects for initial loading. In the **Chunk Size** field, enter the number of records to be processed in a single partition. Based on the chunk size, bulk jobs are created in Salesforce. The default value is 50000 and the minimum value is 100.

When you partition an object, the application ingestion job processes the records for each partition in parallel. Mass Ingestion Applications determine the range of partitions by equal distribution of primary key values of an object.

**Note:** You can partition the objects only if you select **Bulk API 2.0** as the Salesforce API.

13. Select **Include Base64 Fields** to replicate the source fields of Base64 data type.

**Note:**

- You can replicate the Base64 fields only if you select **Standard (REST) API** as the Salesforce API.
- Replication of Base64 data might slow down the initial load operation of the application ingestion job.

14. If you selected the **Include Base64 Fields** check box, in the **Maximum Base64 Body Size** field, specify the body size for Base64 encoded data. The default body size for Base64 encoded data is 7 MB.

15. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.

The custom properties are usually configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

16. Click **Next**.

## Configuring a Salesforce Marketing Cloud source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Salesforce Marketing Cloud source. You can also specify custom properties to address unique environments and special use cases.

1. In the **MID** field, enter the unique Member Identification code assigned to your Salesforce Marketing Cloud account.
2. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.

The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source objects to replicate.

3. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.

- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
    - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
    - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
    - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
    - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
  - d. Define additional rules as needed.
 

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.
  - e. When finished, click **Apply Rules**.
 

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.
4. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.
 

Perform the following steps to create a field action rule:

    - a. Select **Field Action** as the rule type.
    - b. From the adjacent list, select one of the following action types:
      - **LTRIM**. Trims spaces to the left of character field values.
      - **RTRIM**. Trims spaces to the right of character field values.
      - **TRIM**. Trims spaces to the left of and to the right of character field values.
    - c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
    - d. Click **Add Rule**.

**Note:**

    - You cannot use field action rules to trim the spaces in rich text area fields of Salesforce because Salesforce uses control characters instead of empty values to represent the spaces.
    - You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.
  5. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.
    - If you selected **Select All**, the list of objects is view only.

- If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Deselect any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.

**Note:** The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.

For each object, you can view a list of field names and data types. Click the highlighted number of fields in the **Fields** column to list the fields to the right.

To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

6. To download a list of source objects that match the selection rules, perform the following steps:
  - a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.
  - b. If you want to include the fields in the list, select **Include Fields**.
  - c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive. The information in the downloaded

is in the following format:

*status,object\_name,object\_type,field\_name,comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The object is included for processing.</li> <li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li> </ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates an object.</li> <li>- <b>F</b>: Indicates a field.</li> </ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

7. Expand the **Advanced** section.
8. In the **Batch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. Default is 2500.

9. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
10. Click **Next**.

## Configuring an SAP source with SAP ODP Extractor connector

On the **Source** page of the application ingestion task wizard, you can specify the data sources that you want to ingest, and configure the advanced properties for your SAP ECC or SAP S4/HANA source using the SAP ODP Extractor connector. You can also specify custom properties to address unique environments and special use cases.

1. From the **Context** list, select the context containing the source data sources that you want to replicate on the target.

SAP ODP Extractor Connector supports the following ODP providers or contexts for all load types:

Providers/Context	Source SAP System and ODPs
SAP Service Application Programming Interface (S-API)	SAP Data Sources/Extractors without Enterprise Search (ESH)
HANA	SAP HANA Information View
BW	SAP NetWeaver Business Warehouse
ABAP_CDS	ABAP Core Data Services
SAP SLT	SLT Queue

2. In the **Data Source Selection** section, select **Select All** only if you want to select all source reports for data replication. You cannot edit the selection in subsequent dimensions & metrics.

The **Data Source Selected** field shows the count of all selected reports. If you have many source reports, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source reports to replicate.

3. To use rules to select the data sources, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected data source by data source under **Data Source View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all data source accessed with the selected connection. To see how many reports are selected by this rule, click the Refresh icon to display the report count in **Total Data Source Selected** and click **Apply Rules** to see the data source count in **Data Source View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first data source under **Rules**. A row is added to define a new rule.
- b. In the **Data Source Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.

- c. In the condition field, enter a data source name or a data source-name mask to specify the data sources that you want to include in or exclude from the list of selected data sources.

**Notes:**

- A mask can contain the asterisk (\*) wildcard character to represent one or more characters, the question mark (?) wildcard character to represent a single character, or both types of wildcard characters. You can use a wildcard character multiple times in a data source-name mask.
- Data source names are case sensitive. When you define the data source selection rules, you must specify the data source names or masks in the case in which they are defined on the source.
- If a data source name contains delimiters, such as quotation marks or brackets, do not include them when you specify the data source name in the rule.
- If a data source name includes special characters, such as backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), replace each special character in the name with a backslash (\) when you specify the data source name in the rule.

- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Data Sources Affected** and **Total Data Sources Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the report counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Data Sources View** list still lists reports. If you switch to **Select All**, the rules no longer appear.

4. To perform trim actions on the fields of the data sources that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected data sources to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

5. Under **Data Source View**, view the selected data sources, including the number of fields in each data source.

- If you selected **Select All**, the list of data sources is view only.

- If you applied rules, you can refine the set of selected data sources by clicking the check box next to individual data source. Deselect any data source that you do not want to replicate, or select additional data sources to replicate. Click the Refresh icon to update the selected data sources count.

**Note:** The first time you change a check box setting for a data source, the rules are no longer in effect. The selections under **Data Source View** take precedence. However, if you click the Add Rule (+) icon again, the data sources that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Data Source View** list, click **Apply Rules** again.

For each data source, you can view a list of fields. Click the highlighted number of columns in the **Fields** column to list the fields to the right.

To search for data columns and fields, in the drop-down list above **Fields**, select **Data Source Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

- To download a list of data sources that match the selection rules, perform the following steps:
  - From the **List Data Sources by Rule Type** list, select the type of selection rule for which you want to download the list of selected data sources.
  - If you want to include the fields in the list, select **Include Fields**.
  - Click the Download icon.

The list of data sources that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

```
status,data_sources_name,data_sources_type,field_name,comment
```

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the data source from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The data source is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The data source is included for processing.</li> <li>- <b>X</b>. The data source is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the data source is excluded.</li> </ul>
data_source_name	Name of the data source.
data_source_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates a data source.</li> <li>- <b>F</b>: Indicates a field.</li> </ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a data source is excluded from processing even though it matches the selection rules.

- Expand the **Advanced** section.
- For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.

**Note:** By default, the ingestion job retrieves the change records from the latest available position in the data stream.

9. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.

**Note:** The CDC interval must be less than the data retention period configured in the SAP system for the Operational Delta Queue (ODQ).

10. In the **Fetch Size** field, enter the size of data that the application ingestion job associated with the task reads at a time from the source. The value must be in megabytes (MB). The default value for initial load tasks is 2 and the default value for combined initial and incremental load tasks is 8.
11. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
12. Click **Next**.

## Configuring an SAP source with SAP Mass Ingestion connector

On the **Source** page of the application ingestion task wizard, you can specify the tables that you want to ingest for your SAP ECC source using the SAP Mass Ingestion connector. You can also specify custom properties to address unique environments and special use cases.

1. In the **Table Selection** section, select **Select All** only if you want to select all source tables and columns for data replication. You cannot edit the selection in subsequent columns.

The **Tables Selected** field shows the count of all selected tables. If you have many source tables, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source tables to replicate.

2. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected tables by table under **Table View** and also set an option for trimming spaces in character data.

**Note:**

- The default "Include \*" rule selects all source tables accessed with the selected connection. To see how many tables are selected by this rule, click the Refresh icon to display the table count in **Total Tables Selected** and click **Apply Rules** to see the table count in **Table View**.
- The metadata fetched for each object has a dummy column with a **NULL** data type. To fetch metadata for the objects with the actual columns, contact Informatica Global Custom Support or download the list of source objects by enabling the **Include columns that match the selection rules** option.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Table Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.

- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
    - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
    - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
    - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
    - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
  - d. Define additional rules as needed.
 

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.
  - e. When finished, click **Apply Rules**.
 

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Tables Affected** and **Total Tables Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the table counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Table View** list still lists tables. If you switch to **Select All**, the rules no longer appear.
3. To perform trim actions on the columns of the source tables that were selected based on rules, create column action rules.
 

Perform the following steps to create a column action rule:

    - a. Select **Column Action** as the rule type.
    - b. From the adjacent list, select one of the following action types:
      - **LTRIM**. Trims spaces to the left of character column values.
      - **RTRIM**. Trims spaces to the right of character column values.
      - **TRIM**. Trims spaces to the left of and to the right of character column values.
    - c. In the condition field, enter a column name or a column-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against columns of the selected source tables to identify the columns to which the action applies.
    - d. Click **Add Rule**.
 

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The column action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.
  4. Under **Table View**, view the selected tables, including the number of columns in each table.
    - If you selected **Select All**, the list of tables is view only.

- If you applied rules, you can refine the set of selected tables by clicking the check box next to individual table. Deselect any report that you do not want to replicate, or select additional tables to replicate. Click the Refresh icon to update the selected tables count.

**Note:** The first time you change a check box setting for a table, the rules are no longer in effect. The selections under **Table View** take precedence. However, if you click the Add Rule (+) icon again, the tables that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Table View** list, click **Apply Rules** again.

For each table, you can view a list of columns. Click the highlighted number of columns in **Columns** to list the fields to the right.

To search for tables and columns, in the drop-down list above **Columns**, select **Table Name**, **Columns**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

- To download a list of source objects that match the selection rules, perform the following steps:
  - From the **List Tables by Rule Type** list, select the type of selection rule for which you want to download the list of selected tables.
  - Include the columns in the list, select **Include Columns**.
  - Click the download icon.

The list of source objects that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

```
status,object_name,object_type,field_name,comment
```

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an Exclude rule.</li> <li>- <b>I</b>. The object is included for processing.</li> <li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li> </ul>
table_name	Name of the table. <ul style="list-style-type: none"> <li>- <b>O</b>. Indicates a table.</li> <li>- <b>F</b>. Indicates a column.</li> <li>- <b>V</b>. Indicates a view.</li> </ul>
table_type	Type of the table. The possible values are:
field_name	Name of the source field. This information appears only if you selected the <b>Include Columns</b> check box before downloading the list.
comment	Reason why a table is excluded from processing even though it matches the selection rules.

- In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**. The custom properties are usually configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

7. For incremental load tasks, in the **Schema** field, enter the underlying database schema that includes the source tables. Perform the following steps to enter the schema value:
  - a. Log in to the SAP application.
  - b. Browse to **System > Status**.
  - c. Check the **Owner** value. Enter this value in the **Schema** field.
8. If you are defining an incremental load task that has an SAP source and one or more of the selected source tables are not enabled for change data capture, you can generate a script for enabling CDC, and then run or download the script.
  - a. In the **CDC Script** field, select **Enable CDC for all columns**. This enables CDC for all columns in the selected source tables.
  - b. To run the script, click **Execute**.  
 If you don't have a database role or privilege that allows you to run the script, click the download icon to download the script. The script file name uses the following format:  
`cdc_script_taskname_number.txt`. Then, ask your database administrator to run the script.  
 Make sure the script runs before you run the database ingestion task.
9. To create and download a list of the source tables that match the table selection criteria, perform the following substeps:
  - a. If you used rule-based table selection, in the **Tables Name** list, select the type of selection rules that you want to use. The options are:
    - **Include Rules Only**
    - **Exclude Rules Only**
    - **Include And Exclude Rules**
  - b. To list the columns, regardless of the table selection method that you used, select the **Include Columns** check box.
  - c. Click the download icon.

The downloaded list that includes columns uses the following format:

`status, schema_name, table_name, object_type, column_name, comment`

The following table describes the information that is displayed in the downloaded list:

Field	Description
status	Indicates whether Mass Ingestion Applications excludes the source table or column from processing because it has an unsupported type. Valid values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The object is excluded from processing by an Exclude rule.</li> <li>- <b>I</b>. The object is included in processing.</li> <li>- <b>X</b>. The object is excluded from processing because it is an unsupported type of object. For example, unsupported types of objects include columns with unsupported data types and tables that include only unsupported columns. The comment field provides information about the unsupported types.</li> </ul>
schema_name	Specifies the name of the source schema.
table_name	Specifies the name of the source table.

Field	Description
object_type	Specifies the type of the source object. Valid values are: - <b>C.</b> Column. - <b>T.</b> Table.
column_name	Specifies the name of the source column. This information appears only if you selected the <b>Columns</b> check box.
comment	Specifies the reason why a source object of an unsupported type is excluded from processing even though it matches the selection rules.

10. Under **Advanced**, set the advanced properties that are available for your source type and load type.

Property	Source and Load Type	Description
Initial Start Point for Incremental Load	SAP source with SAP Mass Ingestion connector - Incremental load	Set this field to customize the position in the source logs from which the application ingestion job starts reading change records the first time it runs. The <b>Latest Available</b> option provides the latest available position in the database log or structure.

11. Click **Next**.

#### Schema change handling

You can choose how the SAP Mass Ingestion connector handles schema changes that you make to some data object types.

##### Schema change handling for an SAP S4 HANA source

By default, if you make changes to the schema, the SAP Mass Ingestion connector detects the schema changes and generates an alert message.

**Note:** You can disable the schema detection using a custom property. For more information, contact Informatica Global Customer Support.

A schema change includes one or more of the following changes to the data object:

- Fields are added
- Fields are deleted
- Fields are renamed
- Field data type, precision, or scale is updated

You can handle the schema changes based on the following conditions:

- If you make a schema change to the data object while an incremental load job is in the **Up and Running** status, the SAP Mass Ingestion connector generates an alert message. When the alert appears, the processing of the data object stops, and you must create a new initial job with the changed data object. Run the new incremental job after the initial job completes successfully.
- If you make a schema change to the data object while an incremental load job is in the **Stopped** status, you must drop the trigger and the shadow table on the data object where the changes were made. Activate the data object through the SAP user interface and create a new initial job with the updated data object. After the initial job completes successfully, run the new incremental job.

**Note:** Run the incremental job immediately after the initial job completes successfully to prevent data loss.

### Schema change handling for an SAP ECC source

If you make schema changes to the data object, the processing of the data object stops, and the SAP Mass Ingestion connector generates an alert message. Further changes to the data object are not captured.

A schema change includes one or more of the following changes to the data object:

- Fields are added
- Fields are deleted
- Fields are renamed
- Field data type, precision, or scale is updated

You can handle the schema changes based on the following conditions:

- If you make a schema change to the data object for a combined load job, the SAP Mass Ingestion connector generates an alert message. You can then resynchronize the data object.
- If you make a schema change to the data object for an incremental load job, you must create and run an initial job with the changed data object. After the initial job completes successfully, run the new incremental job with the changed data object.

**Note:** Run the incremental job immediately after the initial job completes successfully to prevent data loss.

## Configuring a ServiceNow source

On the **Source** page of the application ingestion task wizard, you can specify the tables that you want to ingest and configure the advanced properties for your ServiceNow source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Table Selection** section, select **Select All** only if you want to select all source tables and columns for data replication. You cannot edit the selection in subsequent columns.

The **Tables Selected** field shows the count of all selected tables. If you have many source tables, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source tables to replicate.

2. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected tables by table under **Table View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source tables accessed with the selected connection. To see how many tables are selected by this rule, click the Refresh icon to display the table count in **Total Tables Selected** and click **Apply Rules** to see the table count in **Table View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Table Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.

- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
    - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
    - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
    - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
    - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
  - d. Define additional rules as needed.
 

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.
  - e. When finished, click **Apply Rules**.
 

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Tables Affected** and **Total Tables Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the table counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Table View** list still lists tables. If you switch to **Select All**, the rules no longer appear.
3. To perform trim actions on the columns of the source tables that were selected based on rules, create column action rules.
 

Perform the following steps to create a column action rule:

    - a. Select **Column Action** as the rule type.
    - b. From the adjacent list, select one of the following action types:
      - **LTRIM**. Trims spaces to the left of character column values.
      - **RTRIM**. Trims spaces to the right of character column values.
      - **TRIM**. Trims spaces to the left of and to the right of character column values.
    - c. In the condition field, enter a column name or a column-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against columns of the selected source tables to identify the columns to which the action applies.
    - d. Click **Add Rule**.
 

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The column action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.
  4. Under **Table View**, view the selected tables, including the number of columns in each table.
    - If you selected **Select All**, the list of tables is view only.

- If you applied rules, you can refine the set of selected tables by clicking the check box next to individual table. Deselect any report that you do not want to replicate, or select additional tables to replicate. Click the Refresh icon to update the selected tables count.

**Note:** The first time you change a check box setting for a table, the rules are no longer in effect. The selections under **Table View** take precedence. However, if you click the Add Rule (+) icon again, the tables that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Table View** list, click **Apply Rules** again.

For each table, you can view a list of columns. Click the highlighted number of columns in **Columns** to list the fields to the right.

To search for tables and columns, in the drop-down list above **Columns**, select **Table Name**, **Columns**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

- To download a list of source tables that match the selection rules, perform the following steps:
  - From the **List Tables by Rule Type** list, select the type of selection rule for which you want to download the list of selected source tables.
  - If you want to include the columns in the list, select **Include Columns**.
  - Click the Download icon.

The list of source tables that match the selection rules is downloaded to your local drive.

The information in the downloaded file is in the following format:

```
status,table_name,table_type,column_name,comment
```

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source table from processing. The possible values are: <ul style="list-style-type: none"> <li>- <b>E</b>. The table is excluded from processing by an <i>Exclude</i> rule.</li> <li>- <b>I</b>. The table is included for processing.</li> <li>- <b>X</b>. The table is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the table is excluded.</li> </ul>
table_name	Name of the source table.
table_type	Type of the source object. The possible values are: <ul style="list-style-type: none"> <li>- <b>O</b>: Indicates a table.</li> <li>- <b>F</b>: Indicates a column.</li> </ul>
column_name	Name of the source column. This information appears only if you selected the <b>Include Columns</b> check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

- For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.

**Note:** You must specify the date and time in Greenwich Mean Time (GMT).

- For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.

8. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. Default is 10000.
9. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
10. Click **Next**.

## Configuring a Workday source

On the **Source** page of the application ingestion task wizard, you can specify the operations that you want to ingest and configure the advanced properties for your Workday source. You can also specify custom properties to address unique environments and special use cases.

1. From the **Workday API** list, select the type of web service that you want to use to read source data.  
Options are:
  - **SOAP**: Uses SOAP APIs to extract Workday data.
  - **RaaS**: Uses Workday Report-as-a-Service (RaaS) to extract source data from custom objects and fields through custom reports. You can use Workday RaaS only in initial load tasks.
2. If you choose to use the SOAP API, perform the following steps:
  - a. From the **Product** list, select **Human Capital Management**.
  - b. From the **Services** list, select the Human Capital Management (HCM) services from which you want to ingest data to your target  
You can select multiple services from the **Services** list.
  - c. From the **Output Type** list, select the format in which you want the data to be stored on the target.  
The ingestion jobs extract the source data in an XML structure. Based on the format that you select, the job writes the extracted data to the target as a single object in either JSON or XML format.
3. If you choose to use the RaaS API, perform the following steps:
  - a. In the **Number of Reports** field, select the number of reports you want to extract from the source.
  - b. If you choose to extract a single report, in the **Report Name or URL** field, enter the name or URL of the custom report you want to read from the source.
  - c. If you choose to extract multiple reports, in the **Report Configuration File** field, enter the path to the CSV file that you created for the list of custom reports that you want to read from the source.
4. In the **Operation Selection** section, select **Select All** only if you want to select all operations for data replication. You cannot edit the selection in subsequent operations.  
The **Operation Selected** field shows the count of all selected operations. If you have many source operations, the interface might take a long time to fetch them.  
Alternatively, you can use rules to define a subset of source operations to replicate.
5. To use rules to select the source operations, make sure that the **Select All** check box is cleared and then add rules.  
When rule-based selection is used, you can refine the set of selected operation by operation under **Operation View** and also set an option for trimming spaces in character data.  
**Note:** The default "Include \*" rule selects all operation accessed with the selected connection. To see how many operations are selected by this rule, click the Refresh icon to display the operation count in **Total Operations Selected** and click **Apply Rules** to see the operation count in **Operation View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first operation under **Rules**. A row is added to define a new rule.
- b. In the **Operation Rule** column, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the condition field, enter an operation name to specify the operations that you want to include in or exclude from the list of selected operation sources.

**Notes:**

- A mask can contain the asterisk (\*) wildcard character to represent one or more characters, the question mark (?) wildcard character to represent a single character, or both types of wildcard characters. You can use a wildcard character multiple times in a data source-name mask.
  - Data source names are case sensitive. When you define the data source selection rules, you must specify the data source names or masks in the case in which they are defined on the source.
  - If a data source name contains delimiters, such as quotation marks or brackets, do not include them when you specify the data source name in the rule.
  - If a data source name includes special characters, such as backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), replace each special character in the name with a backslash (\) when you specify the data source name in the rule.
- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Operations Affected** and **Total Operations Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the operation counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Operations View** list still lists operations. If you switch to **Select All**, the rules no longer appear.

6. Under **Operation View**, view the selected operations, including the number of attributes in each operation.

- If you selected **Select All**, the list of operations is view only.
- If you applied rules, you can refine the set of selected operations by clicking the check box next to individual operation. Deselect any operation that you do not want to replicate, or select additional operations to replicate. Click the Refresh icon to update the selected operations count.

**Note:** The first time you change a check box setting for an operation, the rules are no longer in effect. The selections under **Operation View** take precedence. However, if you click the Add Rule (+) icon again, the operations that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Operation View** list, click **Apply Rules** again.

For each operation, you can view a list of attributes. Click the highlighted number of attributes in the **Attributes** column to list the attributes to the right.

To search for operations and attributes, in the drop-down list above **Attributes**, select **Operation Name**, **Attributes**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

7. To download a list of source operations that match the selection rules, perform the following steps:
  - a. From the **List Operations by Rule Type** list, select the type of selection rule for which you want to download the list of selected source operations.
  - b. Click the Download icon.

The list of source operations that match the selection rules is downloaded to your local drive.  
The information in the downloaded file is in the following format:

*status, operation\_name, operation\_type, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source operation from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The operation is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The operation is included for processing.</li><li>- <b>X</b>. The operation is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the operation is excluded.</li></ul>
operation_name	Name of the source operation.
operation_type	Type of the source object. The value 0 in this field indicates that the object is an operation.
comment	Reason why a source operation is excluded from processing even though it matches the selection rules.

8. For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Coordinated Universal Time (UTC).
9. For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
10. In the **Fetch Size** field, enter the number of records that the application ingestion job associated with the task reads at a time from the source. Default is 100.  
**Note:** The **Fetch Size** field appears only for the SOAP API.
11. Select the **Extract Non-default Fields** check box to replicate the source fields that do not contain any default value.  
**Note:** The **Extract Non-default Fields** check box appears only for the SOAP API.
12. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
13. Click **Next**.

## Configuring a Zendesk source

On the **Source** page of the application ingestion task wizard, you can specify the objects that you want to ingest and configure the advanced properties for your Zendesk source. You can also specify custom properties to address unique environments and special use cases.

1. In the **Object Selection** section, select **Select All** only if you want to select all source objects and fields for data replication. You cannot edit the selection in subsequent fields.

The **Objects Selected** field shows the count of all selected objects. If you have many source objects, the interface might take a long time to fetch them.

Alternatively, you can use rules to define a subset of source objects to replicate.

2. To use rules to select the source objects, make sure that the **Select All** check box is cleared and then add rules.

When rule-based selection is used, you can refine the set of selected objects by object under **Object View** and also set an option for trimming spaces in character data.

**Note:** The default "Include \*" rule selects all source objects accessed with the selected connection. To see how many objects are selected by this rule, click the Refresh icon to display the object count in **Total Objects Selected** and click **Apply Rules** to see the object count in **Object View**.

To add a rule:

- a. Click the Add Rule (+) icon above the first table under **Rules**. A row is added to define a new rule.
- b. In the **Object Rule** field, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the **Condition** column, enter an object name or an object-name mask that includes one or more wildcards to identify the source objects to include in or exclude from object selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the object names or masks in the case with which the objects were defined.
  - Do not include delimiters such as quotation marks or brackets, even if the source uses them.
  - If an object name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), escape each special character with a backslash (\) when you enter the rule.
- d. Define additional rules as needed.

The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

- e. When finished, click **Apply Rules**.

**Tip:** Click the Refresh icon to the right of the **Updated** timestamp to refresh the **Objects Affected** and **Total Objects Selected** counts.

After you apply rules, if you add, delete, or change rules, you must click **Apply Rules** again. Click the Refresh icon to update the object counts. If you delete all rules without clicking **Apply Rules**, a validation error occurs at deployment, even if the **Object View** list still lists objects. If you switch to **Select All**, the rules no longer appear.

3. To perform trim actions on the fields of the source objects that were selected based on rules, create field action rules.

Perform the following steps to create a field action rule:

- a. Select **Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character field values.
  - **RTRIM**. Trims spaces to the right of character field values.
  - **TRIM**. Trims spaces to the left of and to the right of character field values.
- c. In the condition field, enter a field name or a field-name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value that you enter is matched against fields of the selected source objects to identify the fields to which the action applies.
- d. Click **Add Rule**.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The field action rules are processed in the order in which they are listed in the **Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

4. Under **Object View**, view the selected objects, including the number of fields in each object and the field names and data types.
  - If you selected **Select All**, the list of objects is view only.
  - If you applied rules, you can refine the set of selected objects by clicking the check box next to individual objects. Deselect any objects that you do not want to replicate, or select additional objects to replicate. Click the Refresh icon to update the selected objects count.

**Note:** The first time you change a check box setting for an object, the rules are no longer in effect. The selections under **Object View** take precedence. However, if you click the Add Rule (+) icon again, the objects that you deselected or selected individually are reflected as new rules in the Rules list and the rules once again take precedence. If you want to return to the **Object View** list, click **Apply Rules** again.

For each object, you can view a list of field names and data types. Click the highlighted number of fields in the **Fields** column to list the fields to the right.

To search for objects and fields, in the drop-down list above **Fields**, select **Object Name**, **Fields**, or **All** and then enter a search string in the *Find* box and click **Search**. You can include a single asterisk (\*) wildcard at the beginning or end of the string.

5. To download a list of source objects that match the selection rules, perform the following steps:
  - a. From the **List Objects by Rule Type** list, select the type of selection rule for which you want to download the list of selected source objects.
  - b. If you want to include the fields in the list, select **Include Fields**.
  - c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.  
The information in the downloaded

is in the following format:

*status, object\_name, object\_type, field\_name, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The object is included for processing.</li><li>- <b>X</b>. The object is excluded from processing even though it matches the selection rules. The comment field in the file provides details on why the object is excluded.</li></ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"><li>- <b>O</b>: Indicates an object.</li><li>- <b>F</b>: Indicates a field.</li></ul>
field_name	Name of the source field. This information appears only if you selected the <b>Include Fields</b> check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

- For incremental load tasks and combined initial and incremental load tasks, expand the **Advanced** section.
- For incremental load tasks, in the **Initial Start Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting change records.  
**Note:** You must specify the date and time in Coordinated Universal Time (UTC).
- For incremental load tasks and combined initial and incremental load tasks, in the **CDC Interval** field, specify the time interval in which the application ingestion job runs to retrieve the change records for incremental load. The default interval is 5 minutes.
- In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, add the property name and value, and then click **Add Property**.  
The custom properties are usually configured to address unique environments and special use cases.  
**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.
- Click **Next**.

## Example of rules for selecting source objects

When you define a source for an application ingestion task, you can define object selection rules to select the source objects that you want to load to the target. The following example demonstrates how you can use selection rules to select the required objects.

### Example

A source has 1,000 objects with different prefixes. You want to select the objects that have the prefix "2021\_SALES" and all objects with other prefixes except "2021\_".

Define the following rules in the order in which they are listed:

- Rule 1: **Rule Action=Include** and **Condition=\*** includes all objects on the source. When only the asterisk (\*) wildcard is specified, all objects on the source are selected.

- Rule 2: **Rule Action=Exclude** and **Condition=2021\_\*** excludes the source objects that have names with the prefix "2021\_".
- Rule 3: **Rule Action=Include** and **Condition=2021\_SALES\*** includes the source objects that have names with the prefix "2021\_SALES".

The following image shows the rules in the **Object Selection Rules** section of the **Source** page:

▼ Object Selection Rules

---

Create Rule: Object Selection ▼ Include ▼ Enter the condition Add Rule

Object Rules ⓘ Object Count | ^ v

Action	Condition	Objects Affected
Include	*	
Exclude	2021_*	
Include	2021_SALES*	

## Configuring the target

You can configure the target on the **Target** page of the application ingestion task wizard.

Before you configure the target, ensure that the connection to the target is created in Administrator for the runtime environment that your organization uses.

1. From the **Connection** list, select the connection configured for the source application. The connection type appears in parentheses after the connection name.

The list includes only the connections that are valid for the load type that you selected on the **Definition** page.

**Note:** After you deploy the ingestion task, you cannot change the connection without undeploying the associated ingestion job. After you change the connection, you must deploy the task again.

2. Configure the target properties.

For descriptions of the target properties, see the following topics:

- [“Amazon Redshift target properties” on page 108](#)
- [“Amazon S3 target properties” on page 109](#)
- [“Databricks Delta target properties” on page 113](#)
- [“Google BigQuery target properties” on page 115](#)
- [“Google Cloud Storage target properties” on page 116](#)
- [“Kafka target properties” on page 120](#)
- [“Microsoft Azure Data Lake Storage Gen2 target properties” on page 122](#)
- [“Microsoft Azure Synapse Analytics target properties” on page 126](#)
- [“Microsoft SQL Server target properties” on page 130](#)
- [“Microsoft Fabric OneLake target properties” on page 127](#)
- [“Oracle Cloud Object Storage target properties” on page 131](#)
- [“Oracle target properties” on page 135](#)
- [“PostgreSQL target properties” on page 135](#)
- [“Snowflake target properties” on page 136](#)

3. If you want to rename the target objects that are associated with the selected source objects, define table renaming rules.

For more information about the table renaming rules, see [“Rules for renaming tables on the target” on page 106](#).

4. If you want to override the default mappings of source data types to target data types, perform the following steps in the **Data Type Rules** section to define data type rules:
  - a. In the **Create Rule** fields, enter the source data type for which you want to customize the mapping and then enter the target data type that you want to map to the source data type.

**Important:**

- Mass Ingestion Applications does not support the BYTE and CHAR semantics in data-type mappings rules.
- If a source data type has a default value, you must specify it in your rule.

- b. Click **Add Rule**.

The rule is created and appears in the rules list.

**Note:**

- You can define multiple data type mapping rules for a target. However, you can define only one rule for a source data type.
- After you deploy a task with custom mapping rules, you cannot edit the rules without undeploying the task.

5. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, click the **Add Property** icon, and then add the property name and value.

The custom properties are configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

6. Click **Next**.

## Rules for renaming tables on the target

When you configure a target with an existing schema, you can optionally define rules for renaming the target tables that correspond to the selected source objects.

To create a rule for renaming tables, perform the following steps in the **Table Renaming Rules** section:

1. In the **Create Rule** fields, enter the name of the source object that you want to rename and then enter the name that you want to assign to the target table corresponding to the object.

**Notes:**

- You can enter only the asterisk (\*) wildcard character to select all source objects that match the selection criteria defined on the **Source** page. Alternatively, you can enter the name of a specific source object or an object-name pattern that includes the asterisk (\*) wildcard character.
- If an object or table name includes special characters such as a backslash (\), asterisk(\*), dot (.), or question mark (?), replace each special character in the name with a backslash (\) when you create the rule.
- If you want to use an table-name pattern with a wildcard character for the target table, you must also use the wildcard character in the name of the corresponding source object.

- On Windows, if you enter target table renaming criteria that causes a target table name to exceed 232 characters in length, the name is truncated to 222 characters. Mass Ingestion Applications appends 14 characters to the name to add a date-time yyyyMMddHHmmss value, which causes the name to exceed the Windows maximum limit of 255. Ensure that the names of any renamed target tables will not exceed 232 characters.

2. Click **Add Rule**.

The rule is created and appears in the rules list.

You can define multiple table renaming rules. Unless a table matches multiple rules, the order in which the rules are processed does not depend on the order in which they are listed in the **Table Renaming Rules** section. If a table matches multiple rules, the last matching rule determines the name of the table.

To delete a rule, click the Delete icon on the row that contains the rule.

### Example

You want to add the prefix "PROD\_" to the names of target tables that are associated with all selected source objects. In the **Table Renaming Rules** section, enter the following values in the **Create Rule** fields:

- For the source, enter the asterisk (\*) wildcard character to specify all the source objects that match the object selection rules defined on the **Source** page.
- For the target, enter PROD\_\* to add this prefix to the names of all target tables corresponding to the source objects.

## Rules for customizing data-type mappings

When you configure a target for an application ingestion task, you can optionally define data-type mapping rules to override the default mappings of source data types to target data types.

For example, you can create a data-type rule that maps Salesforce ID fields that have no precision to Snowflake target NUMBER() columns that also have no precision, instead of using the default mapping to the Snowflake CHAR(72) data type.

To create a data-type mapping rule:

1. Expand **Data Type Rules**.
2. In the **Create Rule** fields, enter a source data type and the target data type that you want to map it to. In the **Source** field, you can include the percent (%) wildcard to represent the data type precision, scale, or size, for example, NUMBER(%,4), NUMBER(8,%), or NUMBER(%). Use the wildcard to cover all source fields that have the same data type but use different precision, scale, or size values, instead of specifying each one individually. For example, enter FLOAT(%) to cover FLOAT(16), FLOAT(32), and FLOAT(84). You cannot enter the % wildcard in the target data type. A source data type that uses the % wildcard must map to a target data type that uses specific precision, scale, or size value. For example, you could map the source data type FLOAT(%) to a target data type specification such as NUMBER(38,10).
3. Click **Add Rule**.  
The rule appears in the list of rules.

After you deploy a task with custom mapping rules, you cannot edit the rules until the task is undeployed.

### Notes:

- In general, binary data types cannot be mapped to character data types.
- If you define multiple data-type rules for the same source data type with the same length or same precision and scale values, you will not be able to save the application ingestion task.
- If you define multiple data-type rules for the same source data type but use the % wildcard to represent the length or precision and scale value in one rule and a specific length or precision and scale value in the

second rule, the rule that contains the specific value is processed first, before the rule with the % wildcard. For example, if you map the source data types FLOAT(84) and FLOAT(%), the FLOAT(84) rule is processed first and then the FLOAT(%) rule is processed to cover any other FLOAT source columns with different sizes.

- If a source data type requires a length or precision and scale value, make sure that you set the required attribute by using the % wildcard or a specific value, for example, VARCHAR(%) or VARCHAR(10).
- If you define an invalid mapping, an error message is written to the log.
- If a source data type has a default value, you must specify it in your rule. For example, you must use TIMESTAMP(6) instead of TIMESTAMP.

## Amazon Redshift target properties

When you define an application ingestion task, you must specify the properties for your Amazon Redshift target on the **Target** page of the task wizard.

The following table describes the Amazon Redshift target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	Select the target schema in which Mass Ingestion Applications creates the target tables.
Bucket	Specifies the name of the Amazon S3 bucket that stores, organizes, and controls access to the data objects that you load to Amazon Redshift.
Data Directory or Task Target Directory	Specifies the subdirectory where Mass Ingestion Applications stores output files for jobs associated with the task. This field is called <b>Data Directory</b> for an initial load job or <b>Task Target Directory</b> for an incremental load or combined initial and incremental load job.

The following table describes advanced target properties that appear under **Advanced**:

Property	Description
Enable Case Transformation	By default, target table names and column names are generated in the same case as the corresponding source names, unless cluster-level or session-level properties on the target override this case-sensitive behavior. If you want to control the case of letters in the target names, select this check box. Then select a <b>Case Transformation Strategy</b> option.
Case Transformation Strategy	<p>If you selected <b>Enable Case Transformation</b>, select one of the following options to specify how to handle the case of letters in generated target table (or object) names and column (or field) names:</p> <ul style="list-style-type: none"><li>- <b>Same as source</b>. Use the same case as the source table (or object) names and column (or field) names.</li><li>- <b>UPPERCASE</b>. Use all uppercase.</li><li>- <b>lowercase</b>. Use all lowercase.</li></ul> <p>The default value is <b>Same as source</b>.</p> <p><b>Note:</b> The selected strategy will override any cluster-level or session-level properties on the target for controlling case.</p>

## Amazon S3 target properties

When you define an application ingestion task, you must specify the properties for your Amazon S3 target on the **Target** page of the task wizard.

The following table describes the Amazon S3 target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"><li>- <b>CSV</b></li><li>- <b>AVRO</b></li><li>- <b>PARQUET</b></li></ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.
Parquet Compression Type	If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.
Avro Format	If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are: <ul style="list-style-type: none"><li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li><li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li><li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li></ul> The default value is <b>Avro-Flat</b> .
Avro Serialization Format	If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Binary</b></li><li>- <b>JSON</b></li></ul> The default value is <b>Binary</b> .
Avro Schema Directory	If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern: <i>schemaname_tablename.txt</i> <b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.
File Compression Type	Select a file compression type for output files in CSV or AVRO output format. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Deflate</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.

Property	Description
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load and combined initial and incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Task Target Directory	<p>For incremental load and combined initial and incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p>For incremental load and combined initial and incremental load tasks, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task. For Microsoft Fabric OneLake targets, the parent directory is the path specified in the <b>Lakehouse Path</b> field in the Microsoft Fabric OneLake connection properties.</p>
Connection Directory as Parent	<p>Select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load and combined initial and incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>

Property	Description
Schema Directory	<p>Specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads and combined initial and incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLowerCase</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:  <code>{toLowerCase(SchemaName)}</code></p> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle completed file. Default is <code>{TaskTargetDirectory}/cycle/completed</code> .
Cycle Contents Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle contents files. Default is <code>{TaskTargetDirectory}/cycle/contents</code> .
Use Cycle Partitioning for Data Directory	<p>For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>
Use Cycle Partitioning for Summary Directories	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	<p>For incremental load and combined initial and incremental load tasks, lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>

The following table describes the Amazon S3 advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	Select this check box to add a metadata column that records the source SQL operation type in the output that the job propagates to the target. For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert. By default, this check box is selected for incremental load and initial and incremental load jobs, and cleared for initial load jobs.
Add Operation Time	Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target. For initial loads, the job always writes the current date and time. By default, this check box is not selected.
Add Operation Owner	Select this check box to add a metadata column that records the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is not selected.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is not selected.
Add Before Images	Select this check box to include UNDO data in the output that a job writes to the target. For initial loads, the job writes nulls. By default, this check box is not selected.

## Databricks Delta target properties

When you define an application ingestion task, you must specify the properties for your Databricks Delta target on the **Target** page of the task wizard.

The following table describes the Databricks Delta target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.

Property	Description
Apply Mode	<p>For incremental load and combined initial and incremental load jobs, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Standard.</b> Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li> <li>- <b>Soft Deletes.</b> Apply source delete operations to the target as soft deletes. A soft delete marks the deleted row as deleted without actually removing it from the database. For example, a delete on the source results in a change record on the target with "D" displayed in the INFA_OPERATION_TYPE column. If an update followed by a delete occurs on the source, two records are written to the target both with "D" displayed in the INFA_OPERATION_TYPE column. Consider using soft deletes if you have a long-running business process that needs the soft-deleted data to finish processing, to restore data after an accidental delete operation, or to track deleted values for audit purposes.</li> </ul> <p>Default is Standard.</p>
Data Directory or Task Target Directory	<p>Specifies the subdirectory where Mass Ingestion Applications stores output files for jobs associated with the task. This field is called <b>Data Directory</b> for an initial load job or <b>Task Target Directory</b> for an incremental load or combined initial and incremental load job.</p>

Under **Advanced**, you can enter the following advanced target properties if you set **Apply Mode** to **Soft Deletes** to add metadata columns for each delete operation:

Field	Description
Add Operation Type	<p>Add a metadata column that includes the source SQL operation type in the output that the job propagates to the target tables. The column is named INFA_OPERATION_TYPE by default.</p> <p>This check box is displayed only when the <b>Apply Mode</b> option is set to <b>Soft Deletes</b>.</p> <p>In Soft Deletes mode, the job writes "D" for delete operations or NULL for inserts and updates to the INFA_OPERATION_TYPE column. When the operation type is NULL, the other "Add Operation..." metadata columns are also NULL. Only when the operation type is "D" will the other metadata columns contain non-null values.</p> <p>By default, this check box is selected. You cannot deselect it.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>By default, this check box is not selected.</p>
Add Operation Owner	<p>Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target.</p> <p>By default, this check box is not selected.</p> <p><b>Note:</b> This property is not available for jobs that have a PostgreSQL source.</p>
Add Operation Transaction Id	<p>Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations.</p> <p>By default, this check box is not selected.</p>
Prefix for Metadata Columns	<p>Add a prefix to the names of the added metadata columns to easily identify them and to prevent conflicts with the names of existing columns.</p> <p>The default value is INFA_.</p>

Field	Description
Create Unmanaged Tables	Select this check box if you want the task to create Databricks Delta target tables as unmanaged tables. After you deploy the task, you cannot edit this field to switch to managed tables. By default, this option is cleared and managed tables are created. For more information about Databricks Delta managed and unmanaged tables, see the Databricks Delta documentation.
Unmanaged Tables Parent Directory	If you choose to create Databricks Delta unmanaged tables, you must specify a parent directory in Amazon S3 or Microsoft Azure Data Lake Storage to hold the Parquet files that are generated for each target table when captured DML records are processed. <b>Note:</b> To use Unity Catalog, you must provide an existing external directory.

## Google BigQuery target properties

When you define an application ingestion task, you must specify the properties for your Google BigQuery target on the **Target** page of the task wizard.

The following table describes the Google BigQuery target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.
Apply Mode	For incremental load and combined initial and incremental load jobs, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are: <ul style="list-style-type: none"> <li>- <b>Standard.</b> Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li> <li>- <b>Audit.</b> Apply an audit trail of every DML operation made on the source tables to the target. A row for each DML change on a source table is written to the generated target table along with the audit columns you select under the <b>Advanced</b> section. The audit columns contain metadata about the change, such as the DML operation type, time, owner, transaction ID, generated ascending sequence number, and before image. Consider using Audit apply mode when you want to use the audit history to perform downstream computations or processing on the data before writing it to the target database or when you want to examine metadata about the captured changes.</li> </ul> The default value is <b>Standard</b> . <b>Note:</b> The <b>Audit</b> apply mode applies for an SAP source with SAP Mass Ingestion connector.
Bucket	Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.
Data Directory or Task Target Directory	Specifies the subdirectory where Mass Ingestion Applications stores output files for jobs associated with the task. This field is called <b>Data Directory</b> for an initial load job or <b>Task Target Directory</b> for an incremental load or combined initial and incremental load job.

The following table describes advanced Google BigQuery target properties under **Advanced** :

Field	Description
Add Last Replicated Time	Select this check box to add a metadata column that records the timestamp at which a record was inserted or last updated in the target table. For initial loads, all loaded records have the same timestamp. For incremental and combined initial and incremental loads, the column records the timestamp of the last DML operation that was applied to the target. By default, this check box is not selected.
Prefix for Metadata Columns	Add a prefix to the names of the added metadata columns to easily identify them and to prevent conflicts with the names of existing columns. Do not include special characters in the prefix. Otherwise, task deployment will fail. The default value is INFA_.
Enable Case Transformation	By default, target table names and column names are generated in the same case as the corresponding source names, unless cluster-level or session-level properties on the target override this case-sensitive behavior. If you want to control the case of letters in the target names, select this check box. Then select a <b>Case Transformation Strategy</b> option.
Case Transformation Strategy	If you selected <b>Enable Case Transformation</b> , select one of the following options to specify how to handle the case of letters in generated target table (or object) names and column (or field) names: <ul style="list-style-type: none"> <li>- <b>Same as source</b>. Use the same case as the source table (or object) names and column (or field) names.</li> <li>- <b>UPPERCASE</b>. Use all uppercase.</li> <li>- <b>lowercase</b>. Use all lowercase.</li> </ul> The default value is <b>Same as source</b> . <b>Note:</b> The selected strategy will override any cluster-level or session-level properties on the target for controlling case.

## Google Cloud Storage target properties

When you define an application ingestion task, you must specify the properties for your Google Cloud Storage target on the **Target** page of the task wizard.

The following table describes the Google Cloud Storage target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.
Parquet Compression Type	If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are: <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> The default value is <b>None</b> , which means no compression is used.

Property	Description
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load and combined initial and incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Bucket	<p>Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.</p>
Task Target Directory	<p>For incremental load and combined initial and incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p>For incremental load and combined initial and incremental load tasks, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task. For Microsoft Fabric OneLake targets, the parent directory is the path specified in the <b>Lakehouse Path</b> field in the Microsoft Fabric OneLake connection properties.</p>
Schema Directory	<p>Specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads and combined initial and incremental loads, the default directory for the schema file is {TaskTargetDirectory}/data/{TableName}/schema</p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets {}.</p> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:</p> <pre>{toLower(SchemaName)}</pre> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	<p>For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle completed file. Default is {TaskTargetDirectory}/cycle/completed.</p>

Property	Description
Cycle Contents Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle contents files. Default is <code>{TaskTargetDirectory}/cycle/contents</code> .
Use Cycle Partitioning for Data Directory	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory. If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.
Use Cycle Partitioning for Summary Directories	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	For incremental load and combined initial and incremental load tasks, lists individual data files under the contents subdirectory.  If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.  If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.

The following table describes the Google Cloud Storage advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	Select this check box to add a metadata column that records the source SQL operation type in the output that the job propagates to the target. For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert. By default, this check box is selected for incremental load and initial and incremental load jobs, and cleared for initial load jobs.
Add Operation Time	Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target. For initial loads, the job always writes the current date and time. By default, this check box is not selected.
Add Operation Owner	Select this check box to add a metadata column that records the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is not selected.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is not selected.
Add Before Images	Select this check box to include UNDO data in the output that a job writes to the target. For initial loads, the job writes nulls. By default, this check box is not selected.

## Kafka target properties

When you define an application ingestion task, you must specify the properties for your Kafka target on the **Target** page of the task wizard.

These properties apply to incremental load operations only.

The following table describes the Kafka target properties that appear in **Target** section:

Property	Description
Use Table Name as Topic Name	<p>Indicates whether Mass Ingestion Applications writes messages that contain source data to separate topics, one for each source object, or writes all messages to a single topic.</p> <p>Select this check box to write messages to separate table-specific topics. The topic names match the source table names, unless you add the source schema name, a prefix, or a suffix in the <b>Include Schema Name</b>, <b>Table Prefix</b>, or <b>Table Suffix</b> properties.</p> <p>By default, this check box is cleared.</p>
Include Schema Name	<p>When <b>Use Table Name as Topic Name</b> is selected, this check box appears and is selected by default. This setting adds the source schema name in the table-specific topic names. The topic names then have the format <i>schemaname_tablename</i>.</p> <p>If you do <i>not</i> want to include the schema name, clear this check box.</p>
Table Prefix	<p>When <b>Use Table Name as Topic Name</b> is selected, this property appears so that you can optionally enter a prefix to add to the table-specific topic names. For example, if you specify <i>myprefix_</i>, the topic names have the format <i>myprefix_tablename</i>. If you omit the underscore ( <i>_</i> ) after the prefix, the prefix is prepended to the table name.</p>
Table Suffix	<p>When <b>Use Table Name as Topic Name</b> is selected, this property appears so that you can optionally enter a suffix to add to the table-specific topic names. For example, if you specify <i>_mysuffix</i>, the topic names have the format <i>tablename_mysuffix</i>. If you omit the underscore ( <i>_</i> ) before the suffix, the suffix is appended to the table name.</p>
Output Format	<p>Select the format of the output file. Options are:</p> <ul style="list-style-type: none"><li>- <b>CSV</b></li><li>- <b>AVRO</b></li><li>- <b>JSON</b></li></ul> <p>The default value is <b>CSV</b>.</p> <p><b>Note:</b> Output files in CSV format use double-quotation marks ( <i>"</i> ) as the delimiter for each field.</p> <p>If your Kafka target uses Confluent Schema Registry to store schemas for incremental load jobs, you must select <b>AVRO</b> as the format.</p>
JSON Format	<p>If <b>JSON</b> is selected as the output format, select the level of detail of the output. Options are:</p> <ul style="list-style-type: none"><li>- <b>Concise</b>. This format records only the most relevant data in the output, such as the operation type and the column names and values.</li><li>- <b>Verbose</b>. This format records detailed information, such as the table name and column types.</li></ul>
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"><li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li><li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li><li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li></ul> <p>The default value is <b>Avro-Flat</b>.</p>

Property	Description
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> <li>- <b>None</b></li> </ul> <p>The default value is <b>Binary</b>.</p> <p>If you have a Confluent Kafka target that uses Confluent Schema Registry to store schemas, select <b>None</b>. Otherwise, Confluent Schema Registry does not register the schema. Do not select <b>None</b> if you are not using Confluent Scheme Registry.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p> <p>If a source schema change is expected to alter the target, the Avro schema definition file is regenerated with a unique name that includes a timestamp, in the following format:</p> <p><i>schemaname_tablename_YYYYMMDDhhmmss.txt</i></p> <p>This unique naming pattern ensures that older schema definition files are preserved for audit purposes.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>

The following table describes the advanced Kafka target properties that appear under **Advanced**:

Property	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is selected.</p>
Add Operation Time	<p>Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is not selected.</p>
Add Before Images	<p>Select this check box to include UNDO data in the output that a job writes to the target.</p> <p>For initial loads, the job writes nulls.</p> <p>By default, this check box is not selected.</p>

Property	Description
Async Write	<p>Controls whether to use synchronous delivery of messages to Kafka.</p> <ul style="list-style-type: none"> <li>- Clear this check box to use synchronous delivery. Kafka must acknowledge each message as received before Mass Ingestion Applications sends the next message. In this mode, Kafka is unlikely to receive duplicate messages. However, performance might be slower.</li> <li>- Select this check box to use asynchronous delivery. Mass Ingestion Applications sends messages as soon as possible, without regard for the order in which the changes were retrieved from the source.</li> </ul> <p>By default, this check box is selected.</p>
Producer Configuration Properties	<p>Specify a comma-separated list of <i>key=value</i> pairs to enter Kafka producer properties for Apache Kafkatargets.</p> <p>You can specify Kafka producer properties in either this field or in the <b>Additional Connection Properties</b> field in the Kafka connection.</p> <p>If you enter the producer properties in this field, the properties pertain to the application ingestion jobs associated with this task only. If you enter the producer properties for the connection, the properties pertain to jobs for all tasks that use the connection definition, unless you override the connection-level properties for specific tasks by also specifying properties in the <b>Producer Configuration Properties</b> field.</p> <p>For information about Kafka producer properties, see the Apache Kafka documentation.</p>

## Microsoft Azure Data Lake Storage Gen2 target properties

When you define an application ingestion task, you must specify the properties for your Microsoft Azure Data Lake Storage Gen2 target on the **Target** page of the task wizard.

The following table describes the Microsoft Azure Data Lake Storage Gen2 target properties that appear in **Target** section:

Property	Description
Output Format	<p>Select the format of the output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> <p>The default value is <b>CSV</b>.</p> <p><b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.</p>
Add Headers to CSV File	<p>If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.</p>
Parquet Compression Type	<p>If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>

Property	Description
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load and combined initial and incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Task Target Directory	<p>For incremental load and combined initial and incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>

Property	Description
Data Directory	<p>For <i>initial load tasks</i>, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p>For <i>incremental load and combined initial and incremental load tasks</i>, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task. For Microsoft Fabric OneLake targets, the parent directory is the path specified in the <b>Lakehouse Path</b> field in the Microsoft Fabric OneLake connection properties.</p>
Connection Directory as Parent	<p>Select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load and combined initial and incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>

Property	Description
Schema Directory	<p>Specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads and combined initial and incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLowerCase</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:  <code>{toLowerCase(SchemaName) }</code></p> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle completed file. Default is <code>{TaskTargetDirectory}/cycle/completed</code> .
Cycle Contents Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle contents files. Default is <code>{TaskTargetDirectory}/cycle/contents</code> .
Use Cycle Partitioning for Data Directory	<p>For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>
Use Cycle Partitioning for Summary Directories	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	<p>For incremental load and combined initial and incremental load tasks, lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>

The following table describes the Microsoft Azure Data Lake Storage Gen2 advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that records the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is selected for incremental load and initial and incremental load jobs, and cleared for initial load jobs.</p>
Add Operation Time	<p>Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is not selected.</p>
Add Operation Owner	<p>Select this check box to add a metadata column that records the owner of the source SQL operation in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes "INFA" as the owner.</p> <p>By default, this check box is not selected.</p>
Add Operation Transaction Id	<p>Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations.</p> <p>For initial loads, the job always writes "1" as the ID.</p> <p>By default, this check box is not selected.</p>
Add Before Images	<p>Select this check box to include UNDO data in the output that a job writes to the target.</p> <p>For initial loads, the job writes nulls.</p> <p>By default, this check box is not selected.</p>

## Microsoft Azure Synapse Analytics target properties

When you define an application ingestion task, you must specify the properties for your Microsoft Azure Synapse Analytics target on the **Target** page of the task wizard.

The following table describes the Microsoft Azure Synapse Analytics target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	<p>Select the target schema in which Mass Ingestion Applications creates the target tables. The schema name that is specified in the connection properties is displayed by default.</p> <p>This field is case sensitive. Therefore, ensure that you entered the schema name in the connection properties in the correct case.</p>

## Microsoft Fabric OneLake target properties

When you define an application ingestion task that has a Microsoft Fabric OneLake target, you must enter some target properties on the **Target** page of the task wizard.

The following table describes the Microsoft Fabric OneLake target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"><li>- <b>CSV</b></li><li>- <b>AVRO</b></li><li>- <b>PARQUET</b></li></ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.
Parquet Compression Type	If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.
Avro Format	If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are: <ul style="list-style-type: none"><li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li><li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li><li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li></ul> The default value is <b>Avro-Flat</b> .
Avro Serialization Format	If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Binary</b></li><li>- <b>JSON</b></li></ul> The default value is <b>Binary</b> .
Avro Schema Directory	If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern: <i>schemaname_tablename.txt</i> <b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.
File Compression Type	Select a file compression type for output files in CSV or AVRO output format. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Deflate</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.

Property	Description
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load and combined initial and incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Task Target Directory	<p>For incremental load and combined initial and incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>

Property	Description
Data Directory	<p>For <i>initial load tasks</i>, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p>For <i>incremental load and combined initial and incremental load tasks</i>, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task. For Microsoft Fabric OneLake targets, the parent directory is the path specified in the <b>Lakehouse Path</b> field in the Microsoft Fabric OneLake connection properties.</p>
Schema Directory	<p>Specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads and combined initial and incremental loads, the default directory for the schema file is {TaskTargetDirectory}/data/{TableName}/schema</p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets {}.</p> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:</p> <pre>{toLower(SchemaName)}</pre> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	<p>For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle completed file. Default is {TaskTargetDirectory}/cycle/completed.</p>

Property	Description
Cycle Contents Directory	For incremental load and combined initial and incremental load tasks, the path to the directory that contains the cycle contents files. Default is {TaskTargetDirectory}/cycle/contents.
Use Cycle Partitioning for Data Directory	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory. If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.
Use Cycle Partitioning for Summary Directories	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	For incremental load and combined initial and incremental load tasks, lists individual data files under the contents subdirectory.  If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.  If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.

## Microsoft SQL Server target properties

When you define an application ingestion task, you must specify the properties for Microsoft SQL Server target on the **Target** page of the task wizard.

These properties apply to initial load operations only.

The following table describes the Microsoft SQL Server target properties that appear in **Target** section:

Property	Description
Target Creation	The Create Target Tables option generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Applications intelligently handles the target tables on subsequent job runs. Mass Ingestion Applications might truncate or re-create the target tables depending on specific circumstances.
Schema	Select the target schema in which Mass Ingestion Applications creates the target tables. The schema name that is specified in the connection properties is displayed by default. This field is case sensitive. Therefore, ensure that you entered the schema name in the connection properties in the correct case.

## Oracle Cloud Object Storage target properties

When you define a database ingestion task that has an Oracle Cloud Object Storage target, you must enter some target properties on the **Target** tab of the task wizard.

Under **Target**, you can enter the following Oracle Cloud Object Storage target properties:

Property	Description
Output Format	Select one of the following formats of the output file: <ul style="list-style-type: none"><li>- CSV</li><li>- AVRO</li><li>- PARQUET</li></ul> Default is <b>CSV</b> . <b>Note:</b> Output files in the CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If you selected <b>CSV</b> as the output format, select this check box to add a header with source column names to the output CSV file.
Avro Format	If you selected <b>AVRO</b> as the output format, select one of the following formats of the Avro schema that will be created for each source table: <ul style="list-style-type: none"><li>- Avro-Flat. This Avro schema format lists all Avro fields in one record.</li><li>- Avro-Generic. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li><li>- Avro-Nested. This Avro schema format organizes each type of information in a separate record.</li></ul> Default is <b>Avro-Flat</b> .
Avro Serialization Format	If you selected <b>AVRO</b> as the output format, select one of the following serialization formats of the Avro output file: <ul style="list-style-type: none"><li>- None</li><li>- Binary</li><li>- JSON</li></ul> Default is <b>Binary</b> .
Avro Schema Directory	If you selected <b>AVRO</b> as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern: <code>schemaname_tablename.txt</code> <b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.
File Compression Type	Select one of the following file compression type for output files in CSV or AVRO output formats: <ul style="list-style-type: none"><li>- None</li><li>- Deflate</li><li>- Gzip</li><li>- Snappy</li></ul> Default is <b>None</b> , which means no compression is used.
Avro Compression Type	If you selected <b>AVRO</b> as the output format, select one of the following Avro compression type: <ul style="list-style-type: none"><li>- None</li><li>- Bzip2</li><li>- Deflate</li><li>- Snappy</li></ul> Default is <b>None</b> , which means no compression is used.

Property	Description
Parquet Compression Type	<p>If you selected <b>PARQUET</b> as the output format, select one of the following Parquet supported compression type:</p> <ul style="list-style-type: none"> <li>- None</li> <li>- Gzip</li> <li>- Snappy</li> </ul> <p>Default is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If you selected <b>Deflate</b> as the <b>Avro Compression Type</b>, specify a compression level from 0 to 9. Default is 0.</p>
Task Target Directory	<p>For incremental load and combined initial and incremental load tasks, this field defines the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the Connection Directory as Parent option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the directory fields.</p>
Add Directory Tags	<p>For incremental load and combined initial and incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Connection Directory as Parent	<p>Select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b> fields. For incremental load and combined initial and incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b> fields.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the Task <b>Target Directory</b>.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (placeholder) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName}myDir1/ myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>Default is {TableName}_{Timestamp}.</p> <p>For incremental load and combined initial and incremental load tasks, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/data/{TableName}/data.</p> <p><b>Note:</b> For Amazon S3, Flat File, Microsoft Azure Data Lake Storage Gen2, and Oracle Cloud Object Store targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path when Connection Directory as Parent is selected. For Google Cloud Storage targets, Mass Ingestion Applications uses the Bucket name that you specify in the target properties for the ingestion task.</p>
Schema Directory	<p>Specify a custom directory to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads and combined initial and incremental loads, the default directory for the schema file is {TaskTargetDirectory}/data/{TableName}/schema.</p> <p>You can use the same placeholders as for the Data Directory field. Ensure that you enclose placeholders with curly brackets {}.</p> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose both the function and placeholder in curly brackets, for example: {toLower(SchemaName)}.</p> <p><b>Note:</b> Schema is written only to output data files in the CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	<p>For incremental load and combined initial and incremental load tasks, this field defines the path to the directory that contains the cycle completed file. Default is {TaskTargetDirectory}/cycle/completed.</p>
Cycle Contents Directory	<p>For incremental load and combined initial and incremental load tasks, this field defines the path to the directory that contains the cycle contents files. Default is {TaskTargetDirectory}/cycle/contents.</p>

Property	Description
Use Cycle Partitioning for Data Directory	For incremental load and combined initial and incremental load tasks, selecting this option causes a timestamp subdirectory to be created for each CDC cycle, under each data directory. If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.
Use Cycle Partitioning for Summary Directories	For incremental load and combined initial and incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	For incremental load and combined initial and incremental load tasks, lists individual data files under the contents subdirectory.  If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you configure custom subdirectories by using the placeholders, such as for timestamp or date.  If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.

Under **Advanced**, you can enter the following advanced target properties to add metadata columns for each delete operation or each DML change recorded in the audit table.

Field	Description
Add Operation Type	Select this check box to add a metadata column that records the source SQL operation type in the output that the job propagates to the target.  For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.  By default, this check box is selected for incremental load and initial and incremental load jobs, and cleared for initial load jobs.
Add Operation Time	Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target.  For initial loads, the job always writes the current date and time.  By default, this check box is not selected.
Add Before Images	Select this check box to include UNDO data in the output that a job writes to the target.  By default, this check box is not selected.

## Oracle target properties

When you define an application ingestion task, you must specify the properties for your Oracle target on the **Target** page of the task wizard.

The following table describes the Oracle target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	Select the target schema in which Mass Ingestion Applications creates the target tables.
Apply Mode	<p>For incremental load and combined initial and incremental load jobs, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are:</p> <ul style="list-style-type: none"><li>- <b>Standard</b>. Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li><li>- <b>Audit</b>. Apply an audit trail of every DML operation made on the source tables to the target. A row for each DML change on a source table is written to the generated target table along with the audit columns you select under the <b>Advanced</b> section. The audit columns contain metadata about the change, such as the DML operation type, time, owner, transaction ID, generated ascending sequence number, and before image. Consider using Audit apply mode when you want to use the audit history to perform downstream computations or processing on the data before writing it to the target database or when you want to examine metadata about the captured changes.</li></ul> <p>The default value is <b>Standard</b>. <b>Note:</b> The <b>Audit</b> apply mode applies for an SAP source with SAP Mass Ingestion connector.</p>

## PostgreSQL target properties

When you define an application ingestion task, you must specify the properties for PostgreSQL target on the **Target** page of the task wizard.

The following table describes the PostgreSQL target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.

## Snowflake target properties

When you define an application ingestion task, you must specify the properties for your Snowflake target on the **Target** page of the task wizard.

The following table describes the Snowflake target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.
Stage	The name of internal staging area that holds the data read from the source before the data is written to the target tables. The name must not include spaces. If the staging area does not exist, it will be automatically created.
Apply Mode	<p>For incremental load and combined initial and incremental load jobs with Snowflake targets, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are:</p> <ul style="list-style-type: none"><li>- <b>Standard</b>. Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li><li>- <b>Soft Deletes</b>. Apply source delete operations to the target as soft deletes. A soft delete marks the deleted row as deleted without actually removing it from the database. For example, a delete on the source results in a change record on the target with "D" displayed in the INFA_OPERATION_TYPE column.  After enabling Soft Deletes, any update in the source table during normal or backlog mode results in the deletion of the matching record, insertion of the updated record, and marking of the INFA_OPERATION_TYPE operation as NULL in the target table. Similarly, inserting a record in the source table during backlog mode results in marking the INFA_OPERATION_TYPE operation as E in the target table record.  Consider using soft deletes if you have a long-running business process that needs the soft-deleted data to finish processing, to restore data after an accidental delete operation, or to track deleted values for audit purposes.</li><li>- <b>Audit</b>. Apply an audit trail of every DML operation made on the source tables to the target. A row for each DML change on a source table is written to the generated target table along with the audit columns you select under the <b>Advanced</b> section. The audit columns contain metadata about the change, such as the DML operation type, transaction ID, and before image. Consider using <b>Audit</b> apply mode when you want to use the audit history to perform downstream computations or processing on the data before writing it to the target database or when you want to examine metadata about the captured changes.  After enabling the Audit apply mode, any update in the source table during backlog or normal mode results in marking the INFA_OPERATION_TYPE operation as E in the target table record. Similarly, inserting a record in the source table during backlog mode results in marking the INFA_OPERATION_TYPE operation as E in the target table record.</li></ul> <p><b>Note:</b> The <b>Audit</b> apply mode applies for SAP source with SAP Mass Ingestion connector. Default is <b>Standard</b>.</p>

Under **Advanced**, you can enter the following advanced target properties:

Field	Description
Add Last Replicated Time	Select this check box to add a metadata column that records the timestamp at which a record was inserted or last updated in the target table. For initial loads, all loaded records have the same timestamp, except for Snowflake targets that use the Superpipe option where minutes and seconds might vary slightly. For incremental and combined initial and incremental loads, the column records the timestamp of the last DML operation that was applied to the target. By default, this check box is not selected.
Add Operation Type	Add a metadata column that includes the source SQL operation type in the output that the job propagates to the target tables. The column is named INFA_OPERATION_TYPE by default. This field is displayed only when the <b>Apply Mode</b> option is set to <b>Audit</b> or <b>Soft Deletes</b> . In Audit mode, the job writes "I" for inserts, "E" for updates, or "D" for deletes to this metadata column. In Soft Deletes mode, the job writes "D" for deletes or NULL for inserts and updates. When the operation type is NULL, the other "Add Operation..." metadata columns are also NULL. Only when the operation type is "D" will the other metadata columns contain non-null values. By default, this check box is selected. You cannot deselect it.
Add Operation Time	Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target tables. This field is available only when <b>Apply Mode</b> is set to <b>Audit</b> or <b>Soft Deletes</b> . By default, this check box is not selected.
Add Operation Sequence	Select this check box to add a metadata column that records a generated, ascending sequence number for each change operation that the job inserts into the target tables. The sequence number reflects the change stream position of the operation. This field is available only when <b>Apply Mode</b> is set to <b>Audit</b> . By default, this check box is not selected.
Add Before Images	Select this check box to add _OLD columns with UNDO "before image" data in the output that the job inserts into the target tables. You can then compare the old and current values for each data column. For a delete operation, the current value will be null. This field is available only when <b>Apply Mode</b> is set to <b>Audit</b> . By default, this check box is not selected.
Prefix for Metadata Columns	Add a prefix to the names of the added metadata columns to easily identify them and to prevent conflicts with the names of existing columns. The default value is INFA_.
Superpipe	Select this check box to use the Snowpipe Streaming API to quickly stream rows of data directly to Snowflake Data Cloud target tables with low latency instead of first writing the data to stage files. This option is available for all load types. When you configure the target connection, select KeyPair authentication. By default, this check box is selected. Deselect it if you want to write data to intermediate stage files.
Merge Frequency	When <b>Superpipe</b> is selected, you can optionally set the frequency, in seconds, at which change data rows are merged and applied to the Snowflake target tables. This field applies to incremental load and combined initial and incremental load tasks. Valid values are 300 through 604800. Default is 3600 seconds.

Field	Description
Enable Case Transformation	By default, target table names and column names are generated in the same case as the corresponding source names, unless cluster-level or session-level properties on the target override this case-sensitive behavior. If you want to control the case of letters in the target names, select this check box. Then select a <b>Case Transformation Strategy</b> option. <b>Note:</b> This check box is not available if you selected the <b>Superpipe</b> option. You cannot enable case transformation if you are using the Superpipe option for Snowflake.
Case Transformation Strategy	If you selected <b>Enable Case Transformation</b> , select one of the following options to specify how to handle the case of letters in generated target table (or object) names and column (or field) names: <ul style="list-style-type: none"> <li>- <b>Same as source.</b> Use the same case as the source table (or object) names and column (or field) names.</li> <li>- <b>UPPERCASE.</b> Use all uppercase.</li> <li>- <b>lowercase.</b> Use all lowercase.</li> </ul> The default value is <b>Same as source</b> . <b>Note:</b> The selected strategy will override any cluster-level or session-level properties on the target for controlling case.

## Configuring schedule and runtime options

On the **Schedule and Runtime Options** page in the application ingestion task wizard, you can specify a schedule for running the initial load jobs and configure the runtime options for jobs of all load types.

1. In the **Schema Drift Options** section, specify the schema drift option to use for each type of Data Definition Language (DDL) operation.

**Note:** The **Schema Drift Options** section appears only for incremental load and combined initial and incremental load tasks. Additionally, this section appears only for the sources that support automatic detection of schema changes.

The following table describes the schema drift options that you can specify for the DDL operations:

Option	Description
Ignore	Does not replicate DDL changes that occur on the source schema to the target.
Replicate	Allows the application ingestion job to replicate the DDL changes to the target. <b>Note:</b> <ul style="list-style-type: none"> <li>- Add Field operations that add a primary-key field are not supported and might cause unpredictable results.</li> <li>- Modify Field operations that change the NULL or NOT NULL constraint of a field are not replicated to the target.</li> </ul> The types of supported DDL operations are: <ul style="list-style-type: none"> <li>- Add Column</li> <li>- Modify Column</li> <li>- Drop Column</li> <li>- Rename Column</li> </ul> Application ingestion jobs doesn't support modifying or renaming columns for Google BigQuery target, and adding columns for Oracle targets.

Option	Description
Stop Job	Stops the application ingestion job.
Stop Object	<p>Stops processing the source object on which the DDL change occurred.</p> <p><b>Note:</b> When one or more objects are excluded from replication because of the Stop Object schema drift option, the status of the job changes to <b>Running with Warning</b>. The application ingestion job cannot retrieve the data changes that occurred on the source object after the job stops processing the changes. This action leads to data loss on the target. To avoid data loss, you must re-synchronize the source and target objects that the job stopped processing before you resume the application ingestion job.</p>

- Optionally, in the **Advanced** section, modify the value in the **Number of Rows in Output File** value to specify the maximum number of rows that the application ingestion task writes to an output file for an Amazon Redshift, Amazon S3, Google Big Query, Google Cloud Storage, Microsoft Azure Data Lake Storage, Microsoft Azure Synapse Analytics, Microsoft Fabric OneLake, Oracle, or Snowflake target.  
Valid values are 1 through 100000000 and the default value is 100000 rows.  
**Note:** For incremental load and combined initial and incremental load operations, change data is flushed to the target either when the specified number of rows is reached or when the flush latency period expires and the job is not in the middle of processing a transaction. The flush latency period is the time that the job waits for more change data before flushing data to the target. The latency period is set to 10 seconds and cannot be changed.
- For initial load jobs only, optionally clear the **File Extension Based on File Type** check box if you want the output data files for Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage, or Microsoft Fabric OneLake targets to have the .dat extension. This check box is selected by default, which causes the output files to have file-name extensions based on their file types.  
**Note:** For incremental load jobs with these target types, this option is not available. Mass Ingestion Applications always uses output file-name extensions based on file type.
- Optionally, configure an apply cycle. An apply cycle is a cycle of applying change data that starts with fetching the intermediate data from the source and ends with the commit of the data to the target. For continuous replication, the source processes the data in multiple low-latency apply cycles.

For application ingestion incremental load tasks that have Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, or Microsoft Fabric OneLake targets, you can configure the following apply cycle options:

Option	Description
Apply Cycle Interval	Specifies the amount of time that must elapse before an application ingestion job ends an apply cycle. You can specify days, hours, minutes, and seconds or specify values for a subset of these time fields leaving the other fields blank.  The default value is 15 minutes.
Apply Cycle Change Limit	Specifies the number of records that must be processed before an application ingestion job ends an apply cycle. When this record limit is reached, the ingestion job ends the apply cycle and writes the change data to the target.  The default value is 10000 records.  <b>Note:</b> During startup, jobs might reach this limit more frequently than the apply cycle interval if they need to catch up on processing a backlog of older data.
Low Activity Flush Interval	Specifies the amount of time, in hours, minutes, or both, that must elapse during a period of no change activity on the source before an application ingestion job ends an apply cycle. When this time limit is reached, the ingestion job ends the apply cycle and writes the change data to the target.  If you do not specify a value for this option, a database ingestion job ends apply cycles only after either the <b>Apply Cycle Change Limit</b> or <b>Apply Cycle Interval</b> limit is reached.  No default value is provided.

**Note:**

- Either the **Apply Cycle Interval** or **Apply Cycle Change Limit** field must have a non-zero value or use the default value.
  - An apply cycle ends when the job reaches any of the three limits, whichever limit is met first.
5. For incremental load jobs that have an Apache Kafka target, configure the following checkpointing options:

Option	Description
Checkpoint All Rows	Indicates whether a database ingestion job performs checkpoint processing for every message that is sent to the Kafka target.  <b>Note:</b> If this check box is selected, the <b>Checkpoint Every Commit</b> , <b>Checkpoint Row Count</b> , and <b>Checkpoint Frequency (secs)</b> options are ignored.
Checkpoint Every Commit	Indicates whether an application ingestion job performs checkpoint processing for every commit that occurs on the source.
Checkpoint Row Count	Specifies the maximum number of messages that an application ingestion job sends to the target before adding a checkpoint. If you set this option to 0, a database ingestion job does not perform checkpoint processing based on the number of messages. If you set this option to 1, a database ingestion jobs add a checkpoint for each message.
Checkpoint Frequency (secs)	Specifies the maximum number of seconds that must elapse before an application ingestion job adds a checkpoint. If you set this option to 0, a database ingestion job does not perform checkpoint processing based on elapsed time.

6. If you want the application ingestion job associated with the task to run in specific intervals based on a schedule, select **Run this task based on a schedule** in the **Schedule** section, and then select a predefined schedule for the job.

By default, **Do not run this task based on a schedule** is selected, which configures the job to run only when it is manually triggered.

**Note:** This field is available only for initial load tasks.

You can view and edit the job schedule options in Administrator. If you edit the schedule, the changes are automatically applied to all the jobs that are configured to run based on the schedule. If you change the schedule for a task that is already deployed, the updated schedule is automatically applied to the application ingestion job associated with the task.

If a job is about to be triggered based on its schedule when its previous run is still in progress, Mass Ingestion Applications does not run the job and allows the job run that is already in progress to complete.

7. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, click the **Add Property** icon, and then add the property name and value.

The custom properties are configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

## Deploying an application ingestion task

After you define an application ingestion task, deploy the task to create an executable job instance on the on-premises system that contains the Secure Agent and the Database Ingestion agent service. You can run an application ingestion job only after you deploy the associated task. When you deploy the task, Mass Ingestion Applications also validates the task definition.

If you undeploy a job and then want to run the job again, you must deploy the task again to create a new job instance. The new job instance name ends with an incremented number in the format *taskname-job\_instance\_number*. The job instance number is incremented each time you deploy the ingestion task by adding 1 to the maximum instance number across all ingestion jobs.

Before you deploy a task that is configured for a Snowflake target, you must drop the existing target tables that do not match the structure of source objects. The existing target tables might not match the structure of source objects because of newly added source fields, dropped source or target fields, or modified field null constraints or data types. When you deploy the task after dropping the existing target tables, a new set of target tables are generated based on the source object selection rules and target table renaming rules.

- To deploy a task, in the application ingestion task wizard, save the task and then click **Deploy**.

If you included spaces in the name of the application ingestion task, the spaces are omitted from the name of the corresponding application ingestion job.

After you deploy a task successfully, the associated job instance is in the Deployed state. You can run it from the **My Jobs** page that's accessed from the navigation bar of the Home page, from the **All Jobs** page in the Monitor service, or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights.

If the deployment fails, the status of the corresponding application ingestion job is Failed. To diagnose the error, you can download the error log from the **My Jobs** page in Data Integration, the **All Jobs** page in Monitor, or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights. To download the error log, on the **Actions** menu for the job, click **Error Log**. After you resolve the issue, deploy the task again from the application ingestion task wizard or from the **Actions** menu for the job.

**Note:**

- If the Secure Agent is restarted while the task is deploying, the job status switches to Failed. Avoid restarting the Secure Agent while tasks are being deployed.
- If a task appears to be hung in the Deploying state, restart the Secure Agent. The associated job instance acquires the status of Failed. You can then deploy it again.

## Running an application ingestion job

You can run the application ingestion jobs that are deployed and are in any state other than Undeployed.

You can run an application ingestion job from the **My Jobs** page in Data Integration, the **All Jobs** page in Monitor, or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights.

For initial load jobs, you can specify a schedule for running the job when you configure the corresponding application ingestion task.

- On the **Actions** menu for the job that you want to run, click **Run**.

A subtask is started for each source object.

**Notes:**

- If the initial load job fails to load data to a target table, the application ingestion job retries the subtask for the object up to three times. The minimum interval between the retries is 60 seconds. If all the initial load retries fail, Mass Ingestion Applications excludes the object from replication.
- If an initial load job detects inconsistencies between field definitions in the source and target objects, the job drops the target table and then re-creates it to be consistent with the source object before loading the source data to the target.
- Initial load jobs may take a long time to complete the ingestion if the source objects contain many records.

## Stopping an application ingestion job

You can stop an application ingestion job of any load type that is in the **Up and Running**, **Running with Warning**, or **On Hold** status.

When you stop an incremental load job, Mass Ingestion Applications records an identifier for the position in the change stream where it has stopped the incremental processing. The identifier is stored in a recovery table named `INFORMATICA_CDC_RECOVERY` on the target. If you restart the job, Mass Ingestion Applications uses this identifier to identify the last change record that was loaded to the target and starts loading the changes that were made after that point in the change stream.

For initial load jobs, the job stops only after its subtasks that are already running complete their operation. The subtasks that are not running remain in their current states.

1. Navigate to the row for the job that you want to stop in any of the following monitoring interfaces:
  - **My Jobs** page that's accessed from the navigation bar of the Home page
  - **All Jobs** page in Monitor

- **All Jobs** tab on the **Mass Ingestion** page in Operational Insights  
You can also use the Stop icon beside the **Actions** menu to stop the job.
2. From the Actions menu for the job, select **Stop**.  
The job state switches to **Stopping** and then to **Stopped**.  
**Tip:** If the Stop operation is taking too long, you can abort the job.

## Resuming an application ingestion job

You can resume an application ingestion job that is in the Stopped, Aborted, or Failed status.

You can resume an application ingestion job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights.

When you resume an initial load job that has multiple subtasks, Mass Ingestion Applications starts only the subtasks that are in the Failed, Stopped, Aborted, or Queued status.

When you resume an incremental load job or a combined initial and incremental load job, Mass Ingestion Applications resumes replicating source data change from the last position recorded in the checkpoint file or target recovery table. A checkpoint will not be available unless a change record was processed for at least one of the tables during the first job run after deployment. If a checkpoint is not available, the job resumes processing from the configured restart point, which is the latest available position in the change stream by default.

When you resume an incremental load job, Mass Ingestion Applications resumes propagating source data from where it last left off.

1. Navigate to the row for the job that you want to resume in any of the following monitoring interfaces:

- **My Jobs** page that's accessed from the navigation bar of the Home page
- **All Jobs** page in the Monitor service
- **All Jobs** tab on the **Mass Ingestion** page in Operational Insights

2. In the **Actions** menu for the row, click **Resume**.

**Note:** The **Resume** command is not available if the job is in the **Failed** state because the task deployment failed.

A subtask is started for each source table.

If an error occurs, an error message is displayed at the top of the page.

You can also use the Resume icon beside the **Actions** menu to resume the job.

## Restart and recovery for incremental load jobs

Mass Ingestion Applications can restart the incremental load jobs that stopped because of an error and the jobs that were stopped or aborted by users without any loss of change data.

After the first job run, Mass Ingestion Applications continually records an identifier for the processing position in the change stream as changes are applied to the target. The identifier is stored in a recovery table named INFORMATICA\_CDC\_RECOVERY on the target.

When you resume an incremental load job, the job uses the last position recorded in the recovery table to identify the change records that it must load to the target. This process ensures that all changes are ingested to the target.

# Overriding schema drift options when resuming an application ingestion job

You can override the schema drift options when you resume an application ingestion job that is in the Stopped, Aborted, or Failed state. The overrides affect only those objects that are currently in the Error state because of the **Stop Object**, **Stop Table**, **Stop Report**, or **Stop Job** schema drift option. Use the overrides to correct or resolve these errors.

You can override schema drift options and resume an incremental load job or a combined initial and incremental load job either from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights.

1. Navigate to the row for the job that you want to resume with an override.
2. In the Actions menu for the row, click **Resume With Options**.

**Note:** The **Resume With Options** command is not available if the job is in the **Failed** state because the task deployment failed.

The **Resume Options** dialog box appears.

Resume Options

Schema Drift Options:

Select a schema drift option

▼

Resume With Options

Cancel

3. In the **Schema Drift Options** list, select the schema drift option that will be used to process the DDL operation on the source that caused the application ingestion job to stop.

The following table describes the schema drift options:

Option	Description
Ignore	Do not replicate DDL changes that occur on the source to the target.
Stop Table	<p>Stop processing the source object on which the DDL change occurred.</p> <p><b>Important:</b> The application ingestion job cannot retrieve the data changes that occurred on the source object after the job stopped processing it. Consequently, data loss might occur on the target. To avoid data loss, you will need to resynchronize the source and target objects that the job stopped processing. Use the <b>Resume With Options &gt; Resync</b> option.</p>

Option	Description
Resync	Resynchronize the target table with the source object. Use this option for objects that the job stopped processing because of the <b>Stop Object</b> , <b>Stop Table</b> , or <b>Stop Report</b> setting for a <b>Schema Drift</b> option. <b>Important:</b> This option is available only for combined initial and incremental load jobs.
Replicate	Allow the application ingestion job to replicate the DDL change to the target. <b>Important:</b> If you specify the <b>Replicate</b> option for Rename Column operations on Microsoft Azure Synapse Analytics targets, the job will end with an error.

4. Click **Resume With Options**.

The resumed job will use the schema drift option that you specified in step 3 to process the schema change that caused the job to stop. Thereafter, the schema drift options that you specified when creating the task take effect again.

**Important:** Mass Ingestion Applications processes a schema change to a source object only after a DML operation occurs on the object. Therefore, after you resume a job, the object subtask state remains unchanged until the first DML operation occurs on the object.

## Redeploying an application ingestion job

Some fields in application ingestion tasks are editable even when you do not undeploy the associated application ingestion jobs. If you edit any available field in an application ingestion task without undeploying the associated job, you must redeploy the job so that the changes can take effect.

If the fields that you want to edit are uneditable, you must undeploy the associated job and then edit them. For more information about undeploying a job, see ["Undeploying an application ingestion job" on page 146](#).

The redeploy operation stops each job subtask for a source table, deploys the updated ingestion task, and then starts the subtasks. The subtasks that are started includes the subtasks that were previously stopped and the subtasks that are newly created due to the configuration changes in the task.

**Note:** For incremental load jobs and combined initial and incremental load jobs, the redeployment does not change the source objects that were selected for ingestion during the previous deployment. To update the list of objects, you must edit the object selection rules in the associated task, and then redeploy the job.

1. Navigate to the row for the job that you want to redeploy in any of the following monitoring interfaces:

- **My Jobs** page that's accessed from the navigation bar of the Home page
- **All Jobs** page in Monitor
- **All Jobs** tab on the **Mass Ingestion** page in Operational Insights

2. In the **Actions** menu for the row, select **Redeploy**.

The job instance automatically starts running.

If the job was running when you selected **Redeploy**, Mass Ingestion Applications stops the job and then redeploys the ingestion task and restarts the job.

# Undeploying an application ingestion job

You can undeploy an application ingestion job that you no longer want to run. You can also undeploy an application ingestion job if you want to edit the associated ingestion task to update a connection or property that cannot be updated without undeploying the job.

Before you attempt to undeploy a job, ensure that it is not running.

After you undeploy a job, you cannot run it again or redeploy it. If you want to run a job that is undeployed, you must deploy the associated task again from the application ingestion task wizard to create a new job instance. For example, if you want to change the target connection for a job, you must undeploy the job, edit the ingestion task to change the connection, deploy the task again, and then run the new job instance.

1. Navigate to the row for the job that you want to undeploy in any of the following monitoring interfaces:
  - **My Jobs** page that's accessed from the navigation bar of the Home page
  - **All Jobs** page in Monitor
  - **All Jobs** tab on the **Mass Ingestion** page in Operational Insights
2. In the **Actions** menu for the row, click **Undeploy**.

If the undeploy operation fails, the job status switches to Failed, even if it was in the Aborted state, or remains as Failed.

**Note:** After undeploying jobs, do not immediately shut down the Secure Agent. Mass Ingestion Applications requires some time to clean up files for tasks in the `/root/infaagent/apps/Database_Ingestion/data/tasks` directory.

# Aborting an application ingestion job

You can abort an application ingestion job of any load type that is in the Up and Running, Running with Warning, On Hold, or Stopping status.

You can abort an application ingestion job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights.

When you abort an incremental load job, Mass Ingestion Applications records an identifier for the position in the change stream where it has stopped the incremental processing. The identifier is stored in a recovery table named `INFORMATICA_CDC_RECOVERY` on the target. If you restart the job, Mass Ingestion Applications uses this identifier to identify the last change record that was loaded to the target and starts loading the changes that were made after that point in the change stream.

For initial load jobs, the subtasks that are already running stop immediately, and then the job stops. The subtasks that are not running remain in their current states.

- On the **Actions** menu for the job that you want to abort, select **Abort**.

The status of the job changes to Aborting and then changes to Aborted.

For initial load jobs, the status of the subtasks that were running change to Aborted. For incremental load jobs, the status of the subtasks change to Stopped.

# Resynchronizing source and target objects

You can resynchronize source and target objects for a subtask that is part of a running application ingestion job of combined initial and incremental load type. The subtask must be in a state other than Queued or Starting.

For example, you might want to resynchronize the target with the source if initial load or incremental load processing failed or if you want to start the job over again from a specific restart point.

**Important:** To resynchronize objects that stopped and are currently in the **Error** state because of the **Schema Drift** setting of **Stop Object**, **Stop Table**, or **Stop Report**, you must use the **Resume With Options > Resync** option in the Actions menu. For more information, see [“Overriding schema drift options when resuming an application ingestion job” on page 144](#).

1. Drill down on the ingestion job that you want to resynchronize from one of the following monitoring interfaces:

- **My Jobs** page that's accessed from the navigation bar of the Home page
- **All Jobs** page in Monitor
- **All Jobs** tab on the **Mass Ingestion** page in Operational Insights

The job must be in the **Up and Running** state and be for a combined initial and incremental load operation.

2. Click the **Object Detail** tab.
3. In the subtask row for the source and target objects that you want to resynchronize, click the Actions menu and select **Resync**. The resync operation retains the structure of the source and target tables that have been processed for CDC.

**Note:** For the Actions menu and **Resync** option to be available, the subtask must be in a state other than Queued or Starting.

If you resynchronize a subtask for an application ingestion combined load job that has a Salesforce source, use one of the following resync options instead of the **Resync** option:

- **Resync (refresh)**. Use this option to resynchronize the target table with the latest source object definition, including any DDL change that schema drift ignored. After the target table is refreshed, the target table structure matches the current source object structure. This option mimics the behavior of the **Resync** option.
- **Resync (retain)**. Use this option to resynchronize the same fields that have been processed for CDC, retaining the current structure of the source and target tables. No checks for changes in the source or target table definitions are performed. If source DDL changes affected the source object structure, those changes are not processed.

## Notes:

- If the source table contains many rows, the resynchronization might take a long time to perform.
- If the source table schema does not match the target table schema, the ingestion subtask drops the target table and creates a new table that matches the source schema. Regardless of whether the target tables are re-created, the subtask truncates the target tables and then reloads source data to the tables.
- When you resync an application ingestion subtask with a Snowflake target and use Audit apply mode, you can retain the audit information. Mass Ingestion re-creates the target table and renames the existing table that contains the audit information using an appended timestamp in the format `<target_table_name>_<current_UTC_timestamp>`. If you want the audit information in the actual target table, you need to load it, for example, with a join operation. If adding the timestamp to the existing table name causes the name to exceed the maximum number of characters, the subtask fails with an error. If you enable schema drift and a schema drift change, such as Add Column, occurs, the new column will be

in the re-created target table but not in the renamed table. To enable this behavior, set the backupTargetTableBeforeResync custom property to **true** on the **Target** page of the task wizard.

Consider the following limitations when you resync a combined load job that has an existing audit table:

- Storing the existing audit table on the target consumes extra database storage.
- To obtain a unified view of audit information, you need to join the multiple versions of the audit tables.

## Default Data Type Mappings

This reference provides default data-type mappings for source applications and Amazon Redshift, Databricks Delta, Google BigQuery, Microsoft Azure Synapse Analytics, Oracle, and Snowflake targets. When Mass Ingestion Applications generates the target data, it uses these mappings.

When you configure a target, you can optionally define data-type mapping rules to customize the default mappings of source data types to the target tables. For more information, see ["Rules for customizing data-type mappings" on page 107](#).

If a source data type is not listed, Mass Ingestion Applications either cannot extract data from the source columns with this data type or cannot apply the extracted data to any appropriate target data type.

### Adobe Analytics Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an Adobe Analytics source and an Amazon Redshift target:

Adobe Analytics Source Data Type	Amazon Redshift Target Data Type
currency(p,s), 1 <= p <= 38, 0 <= s <= 19	numeric(p,s), 1 <= p <= 38, 0 <= s <= 19
datetime(3)	timestamp without time zone
decimal(p,s), 1 <= p <= 38, 0 <= s <= 19	numeric(p,s), 1 <= p <= 38, 0 <= s <= 19
enum	binary varying(size), 264004 <= size <= 1020004
enum	character varying(65535)
enum	character varying(size), 4 <= size <= 65535
int(p,s), 1 <= p <= 38, 0 <= s <= 19	numeric(p,s), 1 <= p <= 38, 0 <= s <= 19
ordered-enum	binary varying(size), 264004 <= size <= 1020004
ordered-enum	character varying(65535)
ordered-enum	character varying(size), 4 <= size <= 65535
percent(p,s), 1 <= p <= 38, 0 <= s <= 19	numeric(p,s), 1 <= p <= 38, 0 <= s <= 19
string	binary varying(size), 264004 <= size <= 1020004

Adobe Analytics Source Data Type	Amazon Redshift Target Data Type
string	character varying(65535)
string	character varying(size), 4 <= size <= 65535

## Adobe Analytics Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an Adobe Analytics source and a Google BigQuery target:

Adobe Analytics Source Data Type	Google BigQuery Target Data Type
currency(p,s), 1 <= p <= 38, 0 <= s <= 19	bignumeric
datetime(3)	datetime
decimal(p,s), 1 <= p <= 38, 0 <= s <= 19	bignumeric
enum	string
int(p,s), 1 <= p <= 38, 0 <= s <= 19	bignumeric
ordered-enum	string
percent(p,s), 1 <= p <= 38, 0 <= s <= 19	bignumeric
string	string

## Adobe Analytics Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an Adobe Analytics source and a Microsoft Azure Synapse Analytics target:

Adobe Analytics Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
currency(p,s), 1 <= p <= 38, 0 <= s <= 19	decimal(p,s), 1 <= p <= 38, 1 <= s <= 19
datetime(3)	datetime
decimal(p,s), 1 <= p <= 38, 0 <= s <= 19	decimal(p,s), 1 <= p <= 38, 1 <= s <= 19
enum	varchar(size), 4 <= size <= max
int(p,s), 1 <= p <= 38, 0 <= s <= 19	decimal(p,s), 1 <= p <= 38, 1 <= s <= 19
ordered-enum	varchar(size), 4 <= size <= max

Adobe Analytics Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
percent(p,s), 1 <= p <= 38, 0 <= s <= 19	decimal(p,s), 1 <= p <= 38, 1 <= s <= 19
string	varchar(size), 4 <= size <= max

## Adobe Analytics Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an Adobe Analytics source and an Oracle target:

Adobe Analytics Source Data Type	Oracle Target Data Type
currency(p,s), 1 <= p <= 38, 0 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
datetime(3)	timestamp(3)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
enum	char(s char), 4 <= s <= 2000
enum	clob
int(p,s), 1 <= p <= 38, 0 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
ordered-enum	char(s char), 4 <= size <= 2000
ordered-enum	clob
percent(p,s), 1 <= p <= 38, 0 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
string	char(s char), 4 <= size <= 2000
string	clob

## Adobe Analytics Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an Adobe Analytics source and a Snowflake target:

Adobe Analytics Source Data Type	Snowflake Target Data Type
currency(p,0), 1 <= p <= 38	integer
currency(p,s), 1 <= p <= 38, 1 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
datetime(3)	datetime(3)
decimal(p,0), 1 <= p <= 38	integer
decimal(p,s), 1 <= p <= 38, 1 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19

Adobe Analytics Source Data Type	Snowflake Target Data Type
enum	char(size), 4 <= size <= 4192004
int(p,0), 1 <= p <= 38	integer
int(p,s), 1 <= p <= 38, 1 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
ordered-enum	char(size), 4 <= size <= 4192004
percent(p,0), 1 <= p <= 38	integer
percent(p,s), 1 <= p <= 38, 1 <= s <= 19	number(p,s), 1 <= p <= 38, 1 <= s <= 19
string	char(size), 4 <= size <= 16777216

## NetSuite Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and an Amazon Redshift target:

NetSuite Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
char	character varying(4)
clob	binary varying(size), 262144 <= size <= 400000
clob	character varying(size), 4 <= size <= 65535
double(p,s), 1 <= p <= 38, 0 <= s <= 18	numeric(p,s), 1 <= p <= 38, 0 <= s <= 18
number	character varying(size), 40 <= size <= 41
number	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
timestamp	timestamp without time zone
varchar2	character varying(size), 4 <= size <= 65535

## NetSuite Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and a Databricks Delta target:

NetSuite Source Data Type	Databricks Delta Target Data Type
bigint	byte
bigint	integer

NetSuite Source Data Type	Databricks Delta Target Data Type
bigint	long
char	string
clob	string
double(p,s), 1 <= p <= 38, 0 <= s <= 18	decimal(p,s), 1 <= p <= 38, 1 <= s <= 18
number	decimal(p,s), 1 <= p <= 38, 0 <= s <= 16
number	string
timestamp	timestamp
varchar2	string

## NetSuite Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and a Google BigQuery target:

NetSuite Source Data Type	Google BigQuery Target Data Type
bigint	int64
char	string
clob	string
double(p,s), 1 <= p <= 38, 0 <= s <= 18	bignumeric
number	bignumeric
number	bignumeric
number	numeric
timestamp	datetime
varchar2	string

## NetSuite Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and a Microsoft Azure Synapse Analytics target:

NetSuite Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
bigint	bigint
char	nchar
clob	char(8000)
clob	nchar(size), 1 <= size <= 4000
clob	nvarchar(max)
clob	nvarchar(max)
double(p,s), 1 <= p <= 38, 0 <= s <= 18	decimal(p,s), 1 <= p <= 38, 1 <= s <= 18
number	decimal(p,s), 1 <= p <= 38, 0 <= s <= 16
number	varchar(size), 40 <= size <= 41
timestamp	datetime
varchar2	varchar(size), 4 <= size <= max

## NetSuite Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and an Oracle target:

NetSuite Source Data Type	Oracle Target Data Type
bigint	char(s char), 2 <= s <= 13
char	char(4 char)
clob	char(s char), 4 <= s <= 2000
clob	clob
double(p,s), 1 <= p <= 38, 0 <= s <= 18	number(p,s), 1 <= p <= 38, 1 <= s <= 18
number	char(s char), 40 <= s <= 41
number	number(p,s), 1 <= p <= 38, 0 <= s <= 16
timestamp	timestamp(1)

NetSuite Source Data Type	Oracle Target Data Type
varchar2	char(s char), 4 <= s <= 2000
varchar2	clob

## NetSuite Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a NetSuite source and a Snowflake target:

NetSuite Source Data Type	Snowflake Target Data Type
bigint	char(size), 2 <= size <= 13
char	char(4)
clob	char(size), 4 <= size <= 400000
double(p,0), 1 <= p <= 38	integer
double(p,s), 1 <= p <= 38, 1 <= s <= 18	number(p,s), 1 <= p <= 38, 1 <= s <= 18
number	number(p,s), 1 <= p <= 38, 0 <= s <= 16
number	varchar(size), 40 <= size <= 41
timestamp	datetime(1)
varchar2	varchar(size), 4 <= size <= 262136

## Salesforce Source and Amazon Redshift Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and an Amazon Redshift target:

Salesforce Source Data Type	Amazon Redshift Target Data Type
anytype(precision), 1 <= p <= 255	character varying(size), 1 <= size <= 255
base64 <b>Note:</b> The precision depends on the base64 body size you select.	character varying(size), 1 <= size <= 65535
boolean	boolean
combobox(precision), 1 <= p <= 255	character varying(size), 1 <= size <= 255
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 0 <= s <= 17
date	date

Salesforce Source Data Type	Amazon Redshift Target Data Type
datetime	timestamp without time zone
double(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 0 <= s <= 17
email(precision), 1 <= p <= 80	character varying(size), 1 <= size <= 80
encryptedstring(precision), 1 <= p <= 175	character varying(size), 1 <= size <= 175
id(precision), 1 <= p <= 18	character varying(size), 1 <= size <= 18
int	integer
long	bigint
multipicklist(precision), 1 <= p <= 4099	character varying(size), 1 <= size <= 4099
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 0 <= s <= 17
phone(precision), 1 <= p <= 40	character varying(size), 1 <= size <= 40
picklist(precision), 1 <= p <= 255	character varying(size), 1 <= size <= 255
reference(precision), 1 <= p <= 18	character varying(size), 1 <= size <= 18
string(precision), 1 <= p <= 100	character varying(size), 1 <= size <= 100
textarea(precision), 1 <= p <= 131072	character varying(size), 1 <= size <= 65535
time	time without time zone
url(precision), 1 <= p <= 255	character varying(size), 1 <= size <= 255

## Salesforce Source and Databricks Delta Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and a Databricks Delta target:

Salesforce Source Data Type	Databricks Delta Target Data Type
anytype(precision), 1 <= p <= 255	string
base64 <b>Note:</b> The precision depends on the base64 body size you select.	string
boolean	boolean
combobox(precision), 1 <= p <= 255	string
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
date	string

Salesforce Source Data Type	Databricks Delta Target Data Type
datetime	timestamp
double(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
email(precision), 1 <= p <= 80	string
encryptedstring(precision), 1 <= p <= 175	string
id(precision), 1 <= p <= 18	string
int	integer
long	long
multipicklist(precision), 1 <= p <= 4099	string
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
phone(precision), 1 <= p <= 40	string
picklist(precision), 1 <= p <= 255	string
reference(precision), 1 <= p <= 18	string
string(precision), 1 <= p <= 100	string
textarea(precision), 1 <= p <= 131072	string
time	string
url(precision), 1 <= p <= 255	string

## Salesforce Source and Google BigQuery Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and a Google BigQuery target:

Salesforce Source Data Type	Google BigQuery Target Data Type
anytype(precision), 1 <= p <= 255	string
base64 <b>Note:</b> The precision depends on the base64 body size you select.	string
boolean	bool
combobox(precision), 1 <= p <= 255	string
currency(p,0), 1 <= p <= 18	numeric
currency(p,s), 1 <= p <= 18, 1 <= s <= 9	numeric

Salesforce Source Data Type	Google BigQuery Target Data Type
currency(p,s), 1 <= p <= 18, 10 <= s <= 17	bignumeric
date	date
datetime	datetime
double(p,0), 1 <= p <= 18	numeric
double(p,s), 1 <= p <= 18, 1 <= s <= 9	numeric
double(p,s), 1 <= p <= 18, 10 <= s <= 17	bignumeric
email(precision), 1 <= p <= 80	string
encryptedstring(precision), 1 <= p <= 175	string
id(precision), 1 <= p <= 18	string
int	int64
long	int64
multipicklist(precision), 1 <= p <= 4099	string
percent(p,0), 1 <= p <= 18	numeric
percent(p,s), 1 <= p <= 18, 1 <= s <= 9	numeric
percent(p,s), 1 <= p <= 18, 10 <= s <= 17	bignumeric
phone(precision), 1 <= p <= 40	string
picklist(precision), 1 <= p <= 255	string
reference(precision), 1 <= p <= 18	string
string(precision), 1 <= p <= 100	string
textarea(precision), 1 <= p <= 131072	string
time	time
url(precision), 1 <= p <= 255	string

## Salesforce Source and Microsoft Azure Synapse Analytics Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and a Microsoft Azure Synapse Analytics target:

Salesforce Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
anytype( <i>precision</i> ), 1 <= p <= 255	varchar( <i>size</i> ), 1 <= size <= 255
base64 <b>Note:</b> The precision depends on the base64 body size you select.	nchar( <i>size</i> ), 1 <= size <= 255
boolean	bit
combobox( <i>precision</i> ), 1 <= p <= 255	varchar( <i>size</i> ), 1 <= size <= 255
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
date	date
datetime	datetime2(3)
double(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
email( <i>precision</i> ), 1 <= p <= 80	varchar( <i>size</i> ), 1 <= size <= 80
encryptedstring( <i>precision</i> ), 1 <= p <= 175	varchar( <i>size</i> ), 1 <= size <= 175
id( <i>precision</i> ), 1 <= p <= 18	varchar( <i>size</i> ), 1 <= size <= 18
int	int
long	bigint
multipicklist( <i>precision</i> ), 1 <= p <= 4099	varchar( <i>size</i> ), 1 <= size <= 4099
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
phone( <i>precision</i> ), 1 <= p <= 40	varchar( <i>size</i> ), 1 <= size <= 40
picklist( <i>precision</i> ), 1 <= p <= 255	varchar( <i>size</i> ), 1 <= size <= 255
reference( <i>precision</i> ), 1 <= p <= 18	varchar( <i>size</i> ), 1 <= size <= 18
string( <i>precision</i> ), 1 <= p <= 100	varchar( <i>size</i> ), 1 <= size <= 100
textarea( <i>precision</i> ), 1 <= p <= 131072	varchar( <i>size</i> ), 1 <= size <= max
time	time(3)
url( <i>precision</i> ), 1 <= p <= 255	varchar( <i>size</i> ), 1 <= size <= 255

## Salesforce Source and Microsoft Azure SQL Database Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and Microsoft Azure SQL Database target:

Salesforce Source Data Type	PostgreSQL Target Data Type
anytype( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
base64 <b>Note:</b> The precision depends on the base64 body size you select.	character varying( <i>size</i> ), 1 <= size <= max
boolean	bit
combobox( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
date	date
datetime	datetime2 (3)
double(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
email( <i>precision</i> ), 1 <= p <= 80	character varying( <i>size</i> ), 1 <= size <= 80
encryptedstring( <i>precision</i> ), 1 <= p <= 175	character varying( <i>size</i> ), 1 <= size <= 175
id( <i>precision</i> ), 1 <= p <= 18	character varying( <i>size</i> ), 1 <= size <= 18
int	int
long	long
multipicklist( <i>precision</i> ), 1 <= p <= 4099	character varying( <i>size</i> ), 1 <= size <= 4099
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	decimal(p,s), 1 <= p <= 18, 1 <= s <= 17
phone( <i>precision</i> ), 1 <= p <= 40	character varying( <i>size</i> ), 1 <= size <= 40
picklist( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
reference( <i>precision</i> ), 1 <= p <= 18	character varying( <i>size</i> ), 1 <= size <= 18
string( <i>precision</i> ), 1 <= p <= 100	character varying( <i>size</i> ), 1 <= size <= 100
textarea( <i>precision</i> ), 1 <= p <= 131072	character varying( <i>size</i> ), 1 <= size <= 131072
time	time( <i>precision</i> ) without time zone, 3 <= p <= null
url( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255

## Salesforce Source and Oracle Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and an Oracle target:

Salesforce Source Data Type	Oracle Target Data Type
anytype( <i>precision</i> ), 1 <= p <= 255	char(s char), 1 <= s <= 255
base64 <b>Note:</b> The precision depends on the base64 body size you select.	clob
boolean	char(1 char)
combobox( <i>precision</i> ), 1 <= p <= 255	char(s char), 1 <= s <= 255
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
date	date
datetime	timestamp(3)
double(p,s), 1 <= p <= 18, 0 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
email( <i>precision</i> ), 1 <= p <= 80	char(s char), 1 <= s <= 80
encryptedstring( <i>precision</i> ), 1 <= p <= 175	char(s char), 1 <= s <= 175
id( <i>precision</i> ), 1 <= p <= 18	char(s char), 1 <= s <= 18
int	number(10)
long	number(19)
multipicklist( <i>precision</i> ), 1 <= p <= 4099	clob
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
phone( <i>precision</i> ), 1 <= p <= 40	char(s char), 1 <= s <= 40
picklist( <i>precision</i> ), 1 <= p <= 255	char(s char), 1 <= s <= 255
reference( <i>precision</i> ), 1 <= p <= 18	char(s char), 1 <= s <= 18
string( <i>precision</i> ), 1 <= p <= 100	char(s char), 1 <= s <= 100
textarea( <i>precision</i> ), 1 <= p <= 4000	char(s char), 1 <= s <= 4000
textarea( <i>precision</i> ), 4001 <= p <= 131072	clob
time	timestamp(3)
url( <i>precision</i> ), 1 <= p <= 255	char(s char), 1 <= s <= 255

## Salesforce Source and PostgreSQL Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and a PostgreSQL target:

Salesforce Source Data Type	PostgreSQL Target Data Type
anytype( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
base64 <b>Note:</b> The precision depends on the base64 body size you select.	character varying( <i>size</i> ), 1 <= size <= max
boolean	boolean
combobox( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
currency(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 1 <= s <= 17
date	date
datetime	timestamp( <i>precision</i> ) without time zone, 3 <= p <= null
double(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 1 <= s <= 17
email( <i>precision</i> ), 1 <= p <= 80	character varying( <i>size</i> ), 1 <= size <= 80
encryptedstring( <i>precision</i> ), 1 <= p <= 175	character varying( <i>size</i> ), 1 <= size <= 175
id( <i>precision</i> ), 1 <= p <= 18	character varying( <i>size</i> ), 1 <= size <= 18
int	integer
long	bigint
multipicklist( <i>precision</i> ), 1 <= p <= 4099	character varying( <i>size</i> ), 1 <= size <= 4099
percent(p,s), 1 <= p <= 18, 0 <= s <= 17	numeric(p,s), 1 <= p <= 18, 1 <= s <= 17
phone( <i>precision</i> ), 1 <= p <= 40	character varying( <i>size</i> ), 1 <= size <= 40
picklist( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255
reference( <i>precision</i> ), 1 <= p <= 18	character varying( <i>size</i> ), 1 <= size <= 18
string( <i>precision</i> ), 1 <= p <= 100	character varying( <i>size</i> ), 1 <= size <= 100
textarea( <i>precision</i> ), 1 <= p <= 131072	character varying( <i>size</i> ), 1 <= size <= 131072
time	time( <i>precision</i> ) without time zone, 3 <= p <= null
url( <i>precision</i> ), 1 <= p <= 255	character varying( <i>size</i> ), 1 <= size <= 255

## Salesforce Source and Snowflake Target

The following table identifies the default data-type mappings for Mass Ingestion Applications configurations with a Salesforce source and a Snowflake target:

Salesforce Source Data Type	Snowflake Target Data Type
anytype( <i>precision</i> ), 1 <= p <= 255	char( <i>size</i> ), 4 <= size <= 1020
base64 <b>Note:</b> The precision depends on the base64 body size you select.	char( <i>size</i> ), 4 <= size <= 1020
boolean	boolean
combobox( <i>precision</i> ), 1 <= p <= 255	char( <i>size</i> ), 4 <= size <= 1020
currency(p,0), 1 <= p <= 18	integer
currency(p,s), 1 <= p <= 18, 1 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
date	date
datetime	datetime(3)
double(p,0), 1 <= p <= 18	integer
double(p,s), 1 <= p <= 18, 1 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
email( <i>precision</i> ), 1 <= p <= 80	char( <i>size</i> ), 4 <= size <= 320
encryptedstring( <i>precision</i> ), 1 <= p <= 175	char( <i>size</i> ), 4 <= size <= 700
id( <i>precision</i> ), 1 <= p <= 18	char( <i>size</i> ), 4 <= size <= 72
int	number(10)
long	number(19)
multipicklist( <i>precision</i> ), 1 <= p <= 4099	char( <i>size</i> ), 4 <= size <= 16396
percent(p,0), 1 <= p <= 18	integer
percent(p,s), 1 <= p <= 18, 1 <= s <= 17	number(p,s), 1 <= p <= 18, 1 <= s <= 17
phone( <i>precision</i> ), 1 <= p <= 40	char( <i>size</i> ), 4 <= size <= 160
picklist( <i>precision</i> ), 1 <= p <= 255	char( <i>size</i> ), 4 <= size <= 1020
reference( <i>precision</i> ), 1 <= p <= 18	char( <i>size</i> ), 4 <= size <= 72
string( <i>precision</i> ), 1 <= p <= 100	char( <i>size</i> ), 4 <= size <= 400
textarea( <i>precision</i> ), 1 <= p <= 131072	char( <i>size</i> ), 4 <= size <= 131072

Salesforce Source Data Type	Snowflake Target Data Type
time	time(3)
url( <i>precision</i> ), 1 <= p <= 255	char( <i>size</i> ), 4 <= size <= 1020

## Salesforce Marketing Cloud Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and an Amazon Redshift target:

Salesforce Marketing Cloud Source Data Type	Amazon Redshift Target Data Type
boolean	boolean
date(0)	timestamp without time zone
decimal(p,s), 1 <= p <= 29, 0 <= s <= 8	numeric(p,s), 1 <= p <= 29, 0 <= s <= 8
emailaddress( <i>precision</i> ), 1 <= p <= 254	character varying( <i>size</i> ), 4 <= size <= 1016
locale( <i>precision</i> ), 1 <= p <= 5	character varying( <i>size</i> ), 4 <= size <= 20
number	integer
phone( <i>precision</i> ), 1 <= p <= 50	character varying( <i>size</i> ), 4 <= size <= 200
text( <i>precision</i> ), 1 <= p <= 4000	character varying( <i>size</i> ), 4 <= size <= 16000

## Salesforce Marketing Cloud Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and a Databricks Delta target:

Salesforce Marketing Cloud Source Data Type	Databricks Delta Target Data Type
boolean	boolean
date(0)	timestamp
decimal(p,s), 1 <= p <= 29, 0 <= s <= 8	decimal(p,s), 1 <= p <= 29, 1 <= s <= 8
emailaddress( <i>precision</i> ), 1 <= p <= 254	string
locale( <i>precision</i> ), 1 <= p <= 5	string
number	integer
phone( <i>precision</i> ), 1 <= p <= 50	string
text( <i>precision</i> ), 1 <= p <= 4000	string

## Salesforce Marketing Cloud Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and a Google BigQuery target:

Salesforce Marketing Cloud Source Data Type	Google BigQuery Target Data Type
boolean	bool
date(0)	datetime
decimal(p,s), 1 <= p <= 29, 0 <= s <= 8	bignumeric
emailaddress( <i>precision</i> ), 1 <= p <= 254	string
locale( <i>precision</i> ), 1 <= p <= 5	string
number	int64
phone( <i>precision</i> ), 1 <= p <= 50	string
text( <i>precision</i> ), 1 <= p <= 4000	string

## Salesforce Marketing Cloud Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and a Microsoft Azure Synapse Analytics target:

Salesforce Marketing Cloud Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
boolean	bit
date(0)	datetime
decimal(p,s), 1 <= p <= 29, 0 <= s <= 8	decimal(p,s), 1 <= p <= 29, 1 <= s <= 8
emailaddress( <i>precision</i> ), 1 <= p <= 254	nchar( <i>size</i> ), 1 <= size <= 254
locale( <i>precision</i> ), 1 <= p <= 5	nchar( <i>size</i> ), 1 <= size <= 5
number	int
phone( <i>precision</i> ), 1 <= p <= 50	nchar( <i>size</i> ), 1 <= size <= 50
text( <i>precision</i> ), 1 <= p <= 4000	nchar( <i>size</i> ), 1 <= size <= 4000

## Salesforce Marketing Cloud Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and an Oracle target:

Salesforce Marketing Cloud Source Data Type	Oracle Target Data Type
boolean	char(1 char)
date(0)	date
decimal(p,s), 1 <= p <= 29, 0 <= s <= 8	number(p,s), 1 <= p <= 29, 1 <= s <= 8
emailaddress( <i>precision</i> ), 1 <= p <= 254	char(s char), 4 <= s <= 1016
locale( <i>precision</i> ), 1 <= p <= 5	char(s char), 4 <= s <= 20
number	number(10)
phone( <i>precision</i> ), 1 <= p <= 50	char(s char), 4 <= s <= 200
text( <i>precision</i> ), 1 <= p <= 2000	char(s char), 4 <= s <= 2000
text( <i>precision</i> ), 2001 <= p <= 4000	clob

## Salesforce Marketing Cloud Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a Salesforce Marketing Cloud source and a Snowflake target:

Salesforce Marketing Cloud Source Data Type	Snowflake Target Data Type
boolean	boolean
date(0)	datetime(0)
decimal(p,0), 1 <= p <= 29	integer
decimal(p,s), 1 <= p <= 29, 1 <= s <= 8	number(p,s), 1 <= p <= 29, 1 <= s <= 8
emailaddress( <i>precision</i> ), 1 <= p <= 254	char( <i>size</i> ), 4 <= size <= 1016
locale( <i>precision</i> ), 1 <= p <= 5	char( <i>size</i> ), 4 <= size <= 20
number	number(10)
phone( <i>precision</i> ), 1 <= p <= 50	char( <i>size</i> ), 4 <= size <= 200
text( <i>precision</i> ), 1 <= p <= 4000	char( <i>size</i> ), 4 <= size <= 16000

## SAP Source and Amazon Redshift Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP source and an Amazon Redshift target:

SAP Source Data Type	Amazon Redshift Target Data Type
d16r	double precision
d34r	double precision
fltp	double precision
int1	smallint
int2	smallint
int4	integer
int8	bigint

### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Source and Databricks Delta Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP source and a Databricks Delta target:

SAP Source Data Type	Databricks Delta Target Data Type
d16r	double
d34r	double
fltp	double
int1	integer
int2	integer
int4	integer
int8	long

### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Source and Google BigQuery Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP source and a Google BigQuery target:

SAP Source Data Type	Google BigQuery Target Data Type
d16r	float64
d34r	float64
fltp	float64
int1	int64
int2	int64
int4	int64
int8	int64

### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP source and a Microsoft Azure Synapse Analytics target:

SAP Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
d16r	float
d34r	float
fltp	float

SAP Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
int1	tinyint
int2	smallint
int4	int
int8	bigint

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with an SAP source and an Oracle target:

SAP Source Data Type	Oracle Target Data Type
CHAR	VARCHAR2(20)
ACCP	VARCHAR2(24)
CURR	NUMBER(31,8)
CUKY	VARCHAR2(20)
DF16_DEC	BINARY_DOUBLE
DF16_RAW	BINARY_DOUBLE
DF34_DEC	BINARY_DOUBLE
DF34_RAW	BINARY_DOUBLE
DEC	NUMBER(31,9)
DATS	DATE
CLNT	VARCHAR2(12)
INT2	NUMBER(5,0)
LANG	VARCHAR2(4)

SAP Source Data Type	Oracle Target Data Type
NUMC	VARCHAR2(40)
RAW	BLOB
SSTRING	CLOB
STRING	CLOB
TIMS	TIMESTAMP
QUAN	NUMBER(31,6)
UNIT	VARCHAR2(12)
INT1	NUMBER(3,0)
INT4	NUMBER(10,0)
LCHR	VARCHAR2(2000)

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Applications configurations with a SAP source and a Snowflake target:

SAP Source Data Type	Snowflake Target Data Type
d16r	float
d34r	float
fltp	float
int1	number(3)
int2	number(5)
int4	number(10)
int8	number(19)

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- D16S
- D34S
- PREC
- VARC

Additionally, application ingestion jobs do not support the RSTR data type.

## SAP Mass Ingestion connection Source and Amazon Redshift Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and an Amazon Redshift target:

SAP Mass Ingestion connection Source Data Type	Amazon Redshift Target Data Type
accp(6)	character varying(24)
char(p), 1 <= p <= 255	character varying(p), 4 <= p <= 1020
clnt(p), 1 <= p <= 3	character varying(p), 4 <= p <= 12
cuky(p), 1 <= p <= 5	character varying(p), 4 <= p <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
d16r	double precision
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
d34r	double precision
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
fltp	double precision
int1	smallint
int2	smallint
int4	integer
int8	bigint
lang(1)	character varying(4)
lchr *	character varying (65535)
lraw *	binary varying(1024000)

SAP Mass Ingestion connection Source Data Type	Amazon Redshift Target Data Type
numc(p), 1 <= p <= 255	character varying(p), 4 <= p <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
raw(p), 1 <= p <= 255	binary varying(p), 1 <= p <= 255
sstr(p), 1 <= p <= 1333	character varying(p), 4 <= p <= 5332
strg *	character varying (65535)
tims	time without time zone
unit(p), 1 <= p <= 5	character varying(p), 4 <= p <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP Mass Ingestion connection Source and Databricks Delta Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and a Microsoft Azure Synapse Analytics target:

SAP Mass Ingestion connection Source Data Type	Databricks Delta Target Data Type
accp(6)	string
char(p), 1 <= p <= 255	string
clnt(p), 1 <= p <= 3	string
cuky(p), 1 <= p <= 5	string
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	double
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16

SAP Mass Ingestion connection Source Data Type	Databricks Delta Target Data Type
d34r	double
dates	string
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	double
int1	integer
int2	integer
int4	integer
int8	long
lang(1)	string
lchr *	string
lraw *	binary
numc(p), 1 <= p <= 255	string
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	binary
sstr(p), 1 <= p <= 1333	string
strg *	string
tims	string
unit(p), 1 <= p <= 5	string
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP Mass Ingestion connection Source and Google BigQuery Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and a Google BigQuery target:

SAP Mass Ingestion connection Source Data Type	Google BigQuery Target Data Type
accp(6)	string
char(p), 1 <= p <= 255	string
clnt(p), 1 <= p <= 3	string
cuky(p), 1 <= p <= 5	string
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
d16r	float64
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
fltp	float64
int1	int64
int2	int64
int4	int64
int8	int64
lang(1)	string
lchr *	string
lraw *	bytes
numc(p), 1 <= p <= 255	string
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
raw(p), 1 <= p <= 255	bytes
sstr(p), 1 <= p <= 1333	string
strg *	string
tims	time

SAP Mass Ingestion connection Source Data Type	Google BigQuery Target Data Type
unit(p), 1 <= p <= 5	string
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP Mass Ingestion connection source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and an Microsoft Azure Synapse Analytics target:

SAP Mass Ingestion connection Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
accp(6)	varchar(24)
char(1)	nchar null
char(p), 2 <= p <= 255	nchar(p), 2 <= p <= 255
clnt(1)	nchar null
clnt(p), 2 <= p <= 3	nchar(p), 2 <= p <= 3
cuky(1)	nchar null
cuky(p), 2 <= p <= 5	nchar(p), 2 <= p <= 5
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	float
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	float
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16

SAP Mass Ingestion connection Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
flt	float
int1	bigint
int2	bigint
int4	bigint
int8	bigint
lang(1)	nchar null
lchr *	varchar(max)
lraw *	varbinary(max)
numc(1)	nchar null
numc(p), 2 <= p <= 255	nchar(p), 2 <= p <= 255
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(1)	binary null
raw(p), 2 <= p <= 255	binary(p), 2 <= p <= 255
sstr(1)	nchar null
sstr(p), 2 <= p <= 1333	nchar(p), 2 <= p <= 1333
strg *	char null
tims	time(0)
unit(1)	nchar null
unit(p), 2 <= p <= 5	nchar(p), 2 <= p <= 5
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP Mass Ingestion connection Source and Oracle Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion source and an Oracle target:

SAP Mass Ingestion Source Data Type	Oracle Target Data Type
accp(6)	varchar2(24 char)
char(p), 1 <= p <= 255	varchar2(s char), 4 <= s <= 1020
clnt(p), 1 <= p <= 3	varchar2(s char), 4 <= s <= 12
cuky(p), 1 <= p <= 5	varchar2(s char), 4 <= s <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	binary_double
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	binary_double
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	binary_double
int1	number(3)
int2	number(5)
int4	number(10)
int8	number(19)
lang(1)	varchar2(4 char)
lchr *	varchar2(1 char)
lraw *	blob null
numc(p), 1 <= p <= 255	varchar2(s char), 4 <= s <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	blob null
sstr(p), 1 <= p <= 1333	varchar2(s char), 4 <= s <= 4000
strg *	varchar2(1 char)
tims	timestamp(0)

SAP Mass Ingestion Source Data Type	Oracle Target Data Type
unit(p), 1 <= p <= 5	varchar2(s char), 4 <= s <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP Mass Ingestion connection Source and Snowflake Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and a Snowflake target:

SAP Mass Ingestion connection Source Data Type	Snowflake Target Data Type
accp(6)	varchar(24)
char(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
clnt(p), 1 <= p <= 3	varchar(p), 4 <= p <= 12
cuky(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	float
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	float
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	float
int1	number(3)
int2	number(5)
int4	number(10)

SAP Mass Ingestion connection Source Data Type	Snowflake Target Data Type
int8	number(19)
lang(1)	varchar(4)
lchr *	varchar(4)
lraw *	binary(1)
numc(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	binary(p), 1 <= p <= 255
sstr(p), 1 <= p <= 1333	varchar(p), 4 <= p <= 5332
strg *	varchar(4)
tims	time(0)
unit(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Amazon Redshift Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP ODP source and an Amazon Redshift target:

SAP ODP Source Data Type	Amazon Redshift Target Data Type
accp(6)	character varying(24)
char(p), 1 <= p <= 255	character varying(p), 4 <= p <= 1020
clnt(p), 1 <= p <= 3	character varying(p), 4 <= p <= 12
cuky(p), 1 <= p <= 5	character varying(p), 4 <= p <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16

SAP ODP Source Data Type	Amazon Redshift Target Data Type
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
d16r	double precision
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
d34r	double precision
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
fltp	double precision
int1	smallint
int2	smallint
int4	integer
int8	bigint
lang(1)	character varying(4)
lchr *	character varying (65535)
lraw *	binary varying (1024000)
numc(p), 1 <= p <= 255	character varying(p), 4 <= p <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	numeric(p,s), 1 <= p <= 38, 0 <= s <= 16
raw(p), 1 <= p <= 255	binary varying(p), 1 <= p <= 255
sstr(p), 1 <= p <= 1333	character varying(p), 4 <= p <= 5332
strg *	character varying (65535)
tims	time without time zone
unit(p), 1 <= p <= 5	character varying(p), 4 <= p <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Databricks Delta Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP ODP source and a Databricks Delta target:

SAP ODP Source Data Type	Databricks Delta Target Data Type
accp(6)	string
char(p), 1 <= p <= 255	string
clnt(p), 1 <= p <= 3	string
cuky(p), 1 <= p <= 5	string
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	double
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	double
dates	string
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	double
int1	integer
int2	integer
int4	integer
int8	long
lang(1)	string
lchr *	string
lraw *	binary
numc(p), 1 <= p <= 255	string
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	binary
sstr(p), 1 <= p <= 1333	string
strg *	string
tims	string

SAP ODP Source Data Type	Databricks Delta Target Data Type
unit(p), 1 <= p <= 5	string
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Google BigQuery Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP Mass Ingestion connection source and a Google BigQuery target:

SAP Mass Ingestion connection Source Data Type	Google BigQuery Target Data Type
accp(6)	string
char(p), 1 <= p <= 255	string
clnt(p), 1 <= p <= 3	string
cuky(p), 1 <= p <= 5	string
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
d16r	float64
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
fltp	float64
int1	int64
int2	int64
int4	int64
int8	int64
lang(1)	string

SAP Mass Ingestion connection Source Data Type	Google BigQuery Target Data Type
lchr *	string
lraw *	string
numc(p), 1 <= p <= 255	string
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	bignumeric
raw(p), 1 <= p <= 255	bytes
sstr(p), 1 <= p <= 1333	string
strg *	string
tims	time
unit(p), 1 <= p <= 5	string
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP ODP source and a Microsoft Azure Synapse Analytics target:

SAP ODP Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
accp(6)	varchar(24)
char(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
clnt(p), 1 <= p <= 3	varchar(p), 4 <= p <= 12
cuky(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	float

SAP ODP Source Data Type	Microsoft Azure Synapse Analytics Target Data Type
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	float
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	float
int1	tinyint
int2	smallint
int4	int
int8	bigint
lang(1)	varchar(4)
lchr *	varchar(max)
lraw *	varbinary(max)
numc(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	decimal(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(1)	binary null
raw(p), 2 <= p <= 255	binary(p), 2 <= p <= 255
sstr(p), 1 <= p <= 1333	varchar(p), 4 <= p <= 5332
strg *	varchar null
tims	time(0)
unit(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Oracle Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP ODP source and an Oracle target:

SAP ODP Source Data Type	Oracle Target Data Type
accp(6)	varchar2(24 char)
char(p), 1 <= p <= 255	varchar2(s char), 4 <= s <= 1020
clnt(p), 1 <= p <= 3	varchar2(s char), 4 <= s <= 12
cuky(p), 1 <= p <= 5	varchar2(s char), 4 <= s <= 20
curr(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	binary_double
d34d(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	binary_double
dates	date
dec(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
fltp	binary_double
int1	number(3)
int2	number(5)
int4	number(10)
int8	number(19)
lang(1)	varchar2(4 char)
lchr *	varchar2(1 char)
lraw *	blob null
numc(p), 1 <= p <= 255	varchar2(s char), 4 <= s <= 1020
quan(p,s), 1 <= p <= 38, 0 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	blob null
sstr(p), 1 <= p <= 1333	varchar2(s char), 4 <= s <= 4000
strg *	varchar2(1 char)
tims	timestamp(0)

SAP ODP Source Data Type	Oracle Target Data Type
unit(p), 1 <= p <= 5	varchar2(s char), 4 <= s <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

## SAP ODP Source and Snowflake Target

The following table identifies the recommended data type mappings for Mass Ingestion Applications configurations with an SAP ODP source and a Snowflake target:

SAP ODP Source Data Type	Snowflake Target Data Type
accp(6)	varchar(24)
char(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
clnt(p), 1 <= p <= 3	varchar(p), 4 <= p <= 12
cuky(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
curr(p,0), 1 <= p <= 38	integer
curr(p,s), 1 <= p <= 38, 1 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16d(p,0), 1 <= p <= 38	integer
d16d(p,s), 1 <= p <= 38, 1 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d16r	float
d34d(p,0), 1 <= p <= 38	integer
d34d(p,s), 1 <= p <= 38, 1 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
d34r	float
dates	date
dec(p,0), 1 <= p <= 38	integer
dec(p,s), 1 <= p <= 38, 1 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16

SAP ODP Source Data Type	Snowflake Target Data Type
flt	float
int1	number(3)
int2	number(5)
int4	number(10)
int8	number(19)
lang(1)	varchar(4)
lchr *	varchar(4)
lraw *	binary(1)
numc(p), 1 <= p <= 255	varchar(p), 4 <= p <= 1020
quan(p,0), 1 <= p <= 38	integer
quan(p,s), 1 <= p <= 38, 1 <= s <= 16	number(p,s), 1 <= p <= 38, 1 <= s <= 16
raw(p), 1 <= p <= 255	binary(p), 1 <= p <= 255
sstr(p), 1 <= p <= 1333	varchar(p), 4 <= p <= 5332
strg *	varchar(4)
tims	time(0)
unit(p), 1 <= p <= 5	varchar(p), 4 <= p <= 20
* SAP doesn't specify the maximum length for these data types.	

#### Unsupported source data types

Mass Ingestion Applications does not support the following obsolete SAP data types:

- d16s
- d34s
- prec
- varc

Additionally, application ingestion jobs do not support the rstr data type.

# INDEX

## A

- Adobe Analytics sources
  - mappings with Amazon Redshift targets [148](#)
  - mappings with Google BigQuery targets [149](#)
  - mappings with Microsoft Azure Synapse Analytics targets [149](#)
  - mappings with Oracle targets [150](#)
  - mappings with Snowflake targets [150](#)
- Amazon Redshift targets
  - mappings with Adobe Analytics sources [148](#)
  - mappings with NetSuite sources [151](#)
  - mappings with Salesforce Marketing Cloud sources [163](#)
  - mappings with Salesforce sources [154](#)
  - mappings with SAP ODP sources [178](#)
  - mappings with SAP sources [166](#), [170](#)
- application ingestion jobs
  - overriding schema drift options [144](#)
  - resynchronizing source and target objects [147](#)
- application ingestion tasks
  - configuring runtime options [138](#)
  - configuring the source [63](#)
  - configuring the target [105](#)
  - defining basic information for an application ingestion task [61](#)
  - limitations [10](#)
  - Oracle source privileges [24](#)
- apply modes
  - audit mode [59](#)
  - Soft Deletes [59](#)
- audit mode
  - apply mode [59](#)

## C

- Cloud Application Integration community
  - URL [6](#)
- Cloud Developer community
  - URL [6](#)
- connections
  - SAP configuration requirements [18](#)
  - SAP ODP Extractor configuration requirements [18](#)

## D

- Data Integration community
  - URL [6](#)
- data type mappings
  - customizing the default mappings [107](#)
- Databricks Delta targets
  - mappings with NetSuite sources [151](#)
  - mappings with Salesforce Marketing Cloud sources [163](#)
  - mappings with Salesforce sources [155](#)
  - mappings with SAP ODP sources [180](#)
  - mappings with SAP sources [166](#), [171](#)

## G

- Google BigQuery targets
  - mappings with Adobe Analytics sources [149](#)
  - mappings with NetSuite sources [152](#)
  - mappings with Salesforce Marketing Cloud sources [164](#)
  - mappings with Salesforce sources [156](#)
  - mappings with SAP sources [167](#), [173](#), [181](#)

## I

- Informatica Global Customer Support
  - contact information [7](#)
- Informatica Intelligent Cloud Services
  - web site [6](#)

## M

- maintenance outages [7](#)
- Microsoft Azure SQL Database targets
  - mappings with Salesforce sources [159](#)
- Microsoft Azure Synapse Analytics targets
  - mappings with Adobe Analytics sources [149](#)
  - mappings with NetSuite sources [153](#)
  - mappings with Salesforce Marketing Cloud sources [164](#)
  - mappings with Salesforce sources [158](#)
  - mappings with SAP ODP sources [182](#)
  - mappings with SAP sources [167](#), [174](#)

## N

- NetSuite sources
  - mappings with Amazon Redshift targets [151](#)
  - mappings with Databricks Delta targets [151](#)
  - mappings with Google BigQuery targets [152](#)
  - mappings with Microsoft Azure Synapse Analytics targets [153](#)
  - mappings with Oracle targets [153](#)
  - mappings with Snowflake targets [154](#)

## O

- Oracle
  - target guidelines for application ingestion [53](#)
- Oracle sources
  - archive log retention [31](#)
- Oracle targets
  - mappings with Adobe Analytics sources [150](#)
  - mappings with NetSuite sources [153](#)
  - mappings with Salesforce Marketing Cloud sources [165](#)
  - mappings with Salesforce sources [160](#)
  - mappings with SAP Mass Ingestion sources [176](#)
  - mappings with SAP ODP sources [184](#)

Oracle targets (*continued*)  
mappings with SAP sources [168](#)

## P

PostgreSQL targets  
mappings with Salesforce sources [161](#)

## S

Salesforce Marketing Cloud sources  
mappings with Amazon Redshift targets [163](#)  
mappings with Databricks Delta targets [163](#)  
mappings with Google BigQuery targets [164](#)  
mappings with Microsoft Azure Synapse Analytics targets [164](#)  
mappings with Oracle targets [165](#)  
mappings with Snowflake targets [165](#)  
Salesforce sources  
mappings with Amazon Redshift targets [154](#)  
mappings with Databricks Delta targets [155](#)  
mappings with Google BigQuery targets [156](#)  
mappings with Microsoft Azure SQL Database targets [159](#)  
mappings with Microsoft Azure Synapse Analytics targets [158](#)  
mappings with Oracle targets [160](#)  
mappings with PostgreSQL targets [161](#)  
mappings with Snowflake targets [162](#)  
SAP  
installing transport files for processing SAP ODPs [18](#)  
installing transport files for reading SAP ODPs [18](#)  
SAP Mass Ingestion connection sources  
mappings with Amazon Redshift targets [170](#)  
mappings with Databricks Delta targets [171](#)  
mappings with Google BigQuery targets [173](#), [181](#)  
mappings with Microsoft Azure Synapse Analytics targets [174](#)  
mappings with Oracle targets [176](#)  
mappings with Snowflake targets [177](#)  
SAP ODP connection sources  
mappings with Oracle targets [184](#)  
SAP ODP sources  
mappings with Amazon Redshift targets [178](#)

SAP ODP sources (*continued*)  
mappings with Databricks Delta targets [180](#)  
mappings with Microsoft Azure Synapse Analytics targets [182](#)  
mappings with Snowflake targets [185](#)  
SAP sources  
mappings with Amazon Redshift targets [166](#)  
mappings with Databricks Delta targets [166](#)  
mappings with Google BigQuery targets [167](#)  
mappings with Microsoft Azure Synapse Analytics targets [167](#)  
mappings with Oracle targets [168](#)  
mappings with Snowflake targets [169](#)  
SAP user authorization  
configuring to process table data [18](#)  
Snowflake targets  
mappings with Adobe Analytics sources [150](#)  
mappings with NetSuite sources [154](#)  
mappings with Salesforce Marketing Cloud sources [165](#)  
mappings with Salesforce sources [162](#)  
mappings with SAP ODP sources [185](#)  
mappings with SAP sources [169](#), [177](#)  
soft deletes  
apply mode [59](#)  
status  
Informatica Intelligent Cloud Services [7](#)  
system status [7](#)

## T

trust site  
description [7](#)

## U

upgrade notifications [7](#)

## W

web site [6](#)