Informatica® Mass Ingestion
January 2023

Mass Ingestion Applications

Informatica Mass Ingestion Mass Ingestion Applications
January 2023

Publication Date: 2023-01-16

Table of Contents

Preface

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

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 8 and "Supported targets" on page 18.

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.

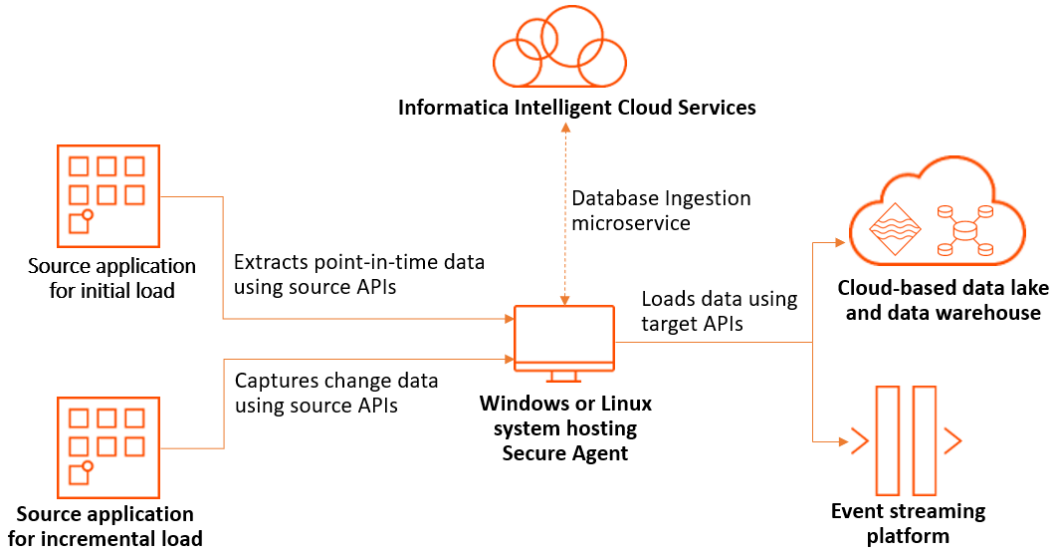
System requirements

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.

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
NetSuite	Initial load, incremental load, and combined initial and incremental load
Oracle Fusion Cloud Applications	Initial 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
SAP S4/HANA	Initial load, incremental load, and combined initial and incremental load
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 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 do not capture the change data for objects without the `modifiedon` field.
- Incremental load jobs and combined initial and incremental load job for Microsoft Dynamics 365 sources do not detect and replicate the changes that are made to the source schema.
- 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
Sales	Loyalty
	Partner Relationship Management
	Customer Data Management
	Sales Automation
	Subscription Management
Service	B2B Service

Module	Application
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

- 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.

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_lattitude`, Mass Ingestion Applications separately loads the data stored in the `geolocation_longitude` and `geolocation_lattitude` fields to the target instead of propagating the compound field `geolocation`.
- The masking configuration of source fields is retained on the target.
- 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.
- The maximum size limit for Base64 fields that application ingestion jobs can propagate is 16 MB.
- 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`.

Guidelines for SAP ECC and SAP S4/HANA sources

Consider the following guidelines when you use SAP ECC or SAP S4/HANA sources:

- 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.
- 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.

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_Budgest_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 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****.
- Mass Ingestion Applications does not propagate fields of the following ServiceNow data types:
  - Audio
  - Basic Image
  - Collection
  - Encrypted Text
  - FX Currency
  - Image
  - Journal List
  - List
  - Name-Value Pairs
  - Video
  - Wiki

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® Invitations
  - NPS® Recipients
  - NPS® 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, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, 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, Oracle, and Snowflake
Marketo	Amazon Redshift, Amazon S3, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, 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, 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, Oracle, and Snowflake
Oracle Fusion Cloud	Amazon Redshift, Amazon S3, Google BigQuery, Microsoft Azure Synapse Analytics, Oracle, and Snowflake
Salesforce	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Oracle, 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, Oracle, and Snowflake
SAP	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Oracle, and Snowflake

Source type	Supported target type
ServiceNow	Amazon Redshift, Apache Kafka, Databricks Delta, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Oracle, and Snowflake
Workday	Amazon Redshift, Amazon S3, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, Oracle, and Snowflake **Note:** Mass Ingestion Application supports only Snowflake and Oracle as target types for Workday RaaS.
Zendesk	Amazon Redshift, Amazon S3, Apache Kafka, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, 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.
- If your Informatica Intelligent Cloud Services organization is on the United Kingdom, Ireland, or Sydney pods and you use an Amazon VPC, you can configure Informatica Intelligent Cloud Services to communicate with your Amazon Redshift endpoint using AWS PrivateLink

Guidelines for Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets

Consider the following guidelines when you use Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets:

- When you configure an application ingestion task for an Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage Gen2 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. **Notes:** - If you selected any of the **Add Operation...** properties under **Advanced** on the **Target** page of the task wizard, the list of columns includes metadata columns for the operation type, time, owner, or transaction ID. - If you selected the **Add Before Images** check box, for each source column, the job creates a *column_name*_OLD column for UNDO data and *column_name*_NEW column for REDO data.

  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 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.

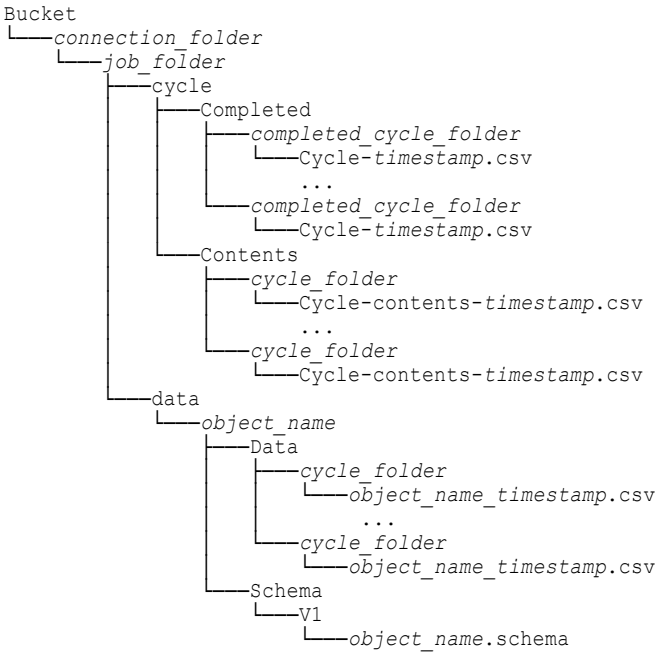
  For Update operations that do not change primary key values, application ingestion jobs process each Update operation as a single operation and writes only the after image to the target.

  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, and Microsoft Azure Data Lake Storage Gen2 targets

Application ingestion jobs create directories on Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 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
connection_folder	Contains the Mass Ingestion Applications objects. This folder is specified in the **Folder Path** field of the Amazon S3 connection properties or in the **Directory Path** field of the Microsoft Azure Data Lake Storage Gen2 connection properties. **Note:** This folder is not created for Google Cloud Storage targets.
job_folder	Contains job output files. This folder is specified in the **Directory** field on the **Target** page of the application ingestion task wizard.

Folder	Description
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. **Note:** 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
  - `E` 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, and ADLS Gen2 targets

You can configure a custom directory structure for the output files that initial load, incremental load, combined initial and incremental load jobs write to Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage (ADLS) Gen2 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.

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 yyyymmdd_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:

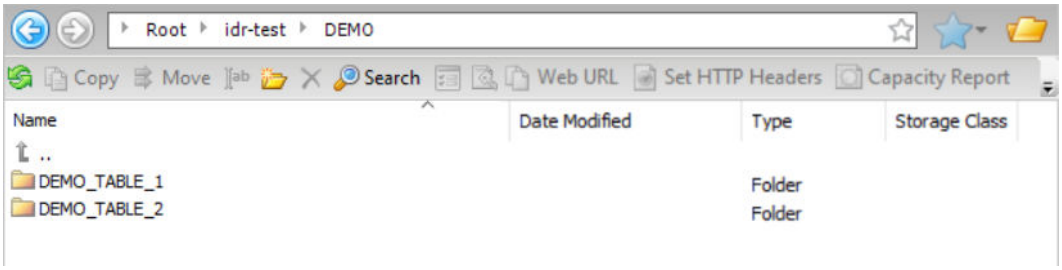
- toLower() to use lowercase for the values represented by the placeholder in parentheses
- toUpper() to use uppercase for the values represented by the placeholder in parentheses

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.

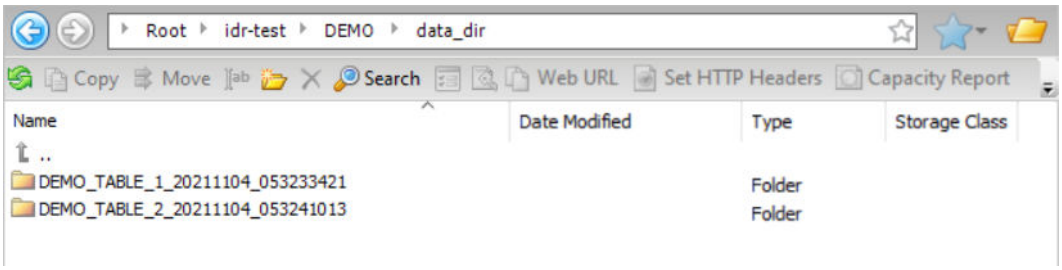
Based on this configuration, the resulting directory structure is:



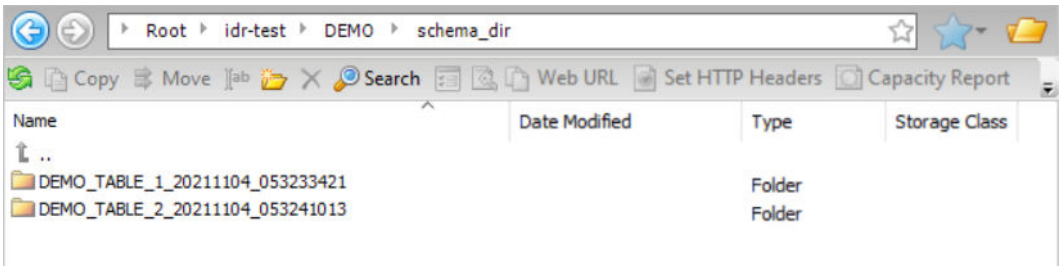
Example 2

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:



And the resulting schema directory structure is:



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, or Microsoft Azure Data Lake Storage (ADLS) Gen2 targets.

For all targets except Google Cloud Storage, 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.

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	Name of a root directory to use for storing output files for an incremental load task. If you select the **Connection Directory as Parent** 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. This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.	None
Connection Directory as Parent	Select this check box to use the parent directory specified in the connection properties.	Selected

Field	Description	Default
Data Directory	Path to the subdirectory that contains the data files. In the directory path, the {TableName} placeholder is required if data and schema files are *not* partitioned by CDC cycle.	{TaskTargetDirectory}/data/{TableName}/data
Schema Directory	Path to the subdirectory in which to store the schema file if you do not want to store it in the data directory. In the directory path, the {TableName} placeholder is required if data and schema files are not partitioned by CDC cycle.	{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	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.	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	Lists individual data files under the contents subdirectory. If **Use Cycle Partitioning for Summary Directories** 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 **Use Cycle Partitioning for Data Directory** is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.	Not selected if **Use Cycle Partitioning for Summary Directories** is selected. Selected if you cleared **Use Cycle Partitioning for Summary Directories**.

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

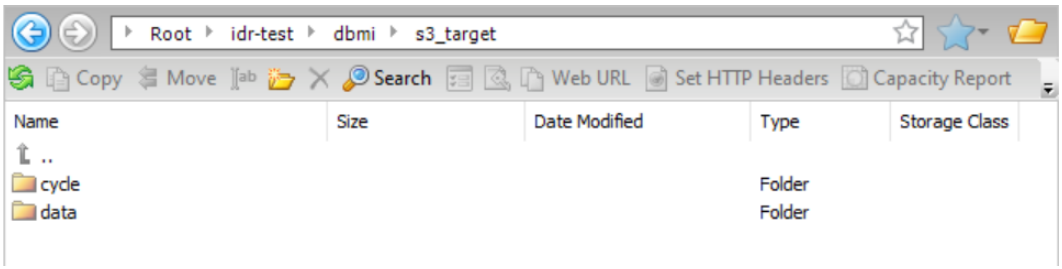
- {TaskTargetDirectory} for a task-specific base directory on the target to use instead of the directory the connection properties
- {TableName} for a target table name
- {Timestamp} for the date and time, in the format yyyymmdd_hhmissms
- {Schema} for the target schema name
- {YY} for a two-digit year
- {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.

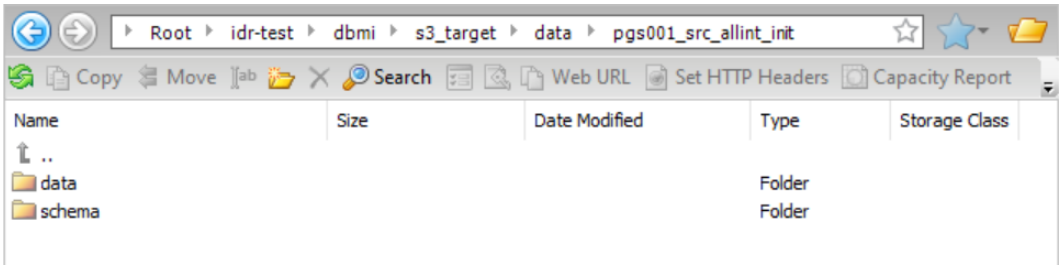
Example 1

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.

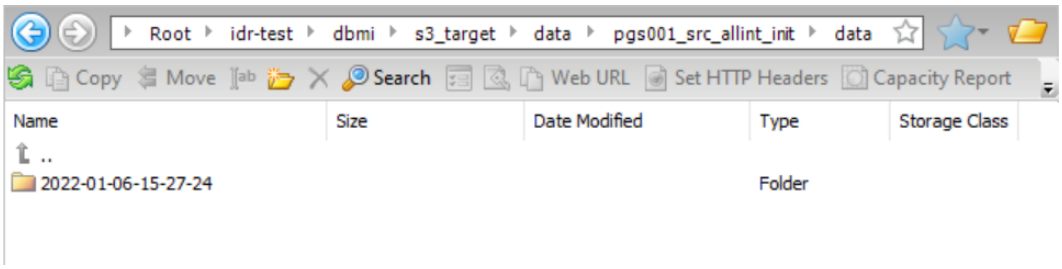
Based on this configuration, the resulting data directory structure is:



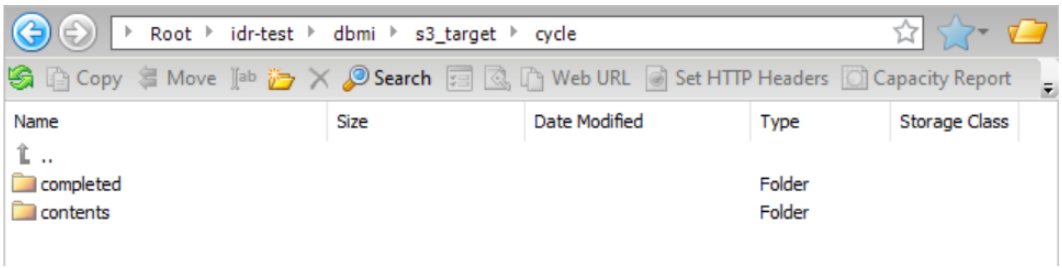
If you drill down on the data folder and then on a table in that folder (pgs001_src_allint_init), you can access the data and schema subdirectories:



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



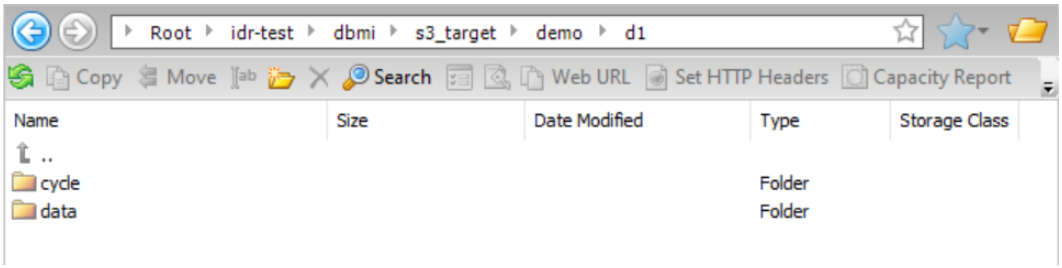
If you drill down on cycle, you can access the summary contents and completed subdirectories:



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:

  `schemaname_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:

  `schemaname_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 Simba Apache Spark JDBC 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.

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, private_key_id, client_id, and region ID values for the service account. You must enter these details when you create a Google BigQuery connection.

  Note: Specify the private_key_id and client_id values in the **Provide Optional Properties** field of the connection properties. Use the following format:

  ```
  "private_key_id": "<private_key_id_value>", "client_id": "<client_id_value>"
  ```

- 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

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 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 Oracle targets

Consider the following guidelines when you use Oracle targets:

- You can use Oracle targets in application ingestion jobs to replicate data from Salesforce sources.
- 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 Snowflake targets

Target preparation

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

- Before writing data to Snowflake target tables, application ingestion jobs write the data to an internal staging area. You must specify the staging directory when you configure the application ingestion task.
- 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 and choose the **KeyPair** option as the authentication method, it is recommended to use OpenSSL 1.1.1 to generate the private key. Otherwise, when you try to define the target in the application ingestion task wizard, an error message about an invalid or unsupported private key might occur while fetching the target schema.
- 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.
- 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.

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
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 unscaled × 10 $^{-scale}$
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 93.

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	- Add field - Modify field - Drop field - Rename field
Google Analytics	- Add column - Modify column - Drop column - Rename column
Marketo	Not supported
Microsoft Dynamics 365	- Add column - Modify column - Drop column
NetSuite	Add column

Source type	Supported schema changes
Oracle Fusion Cloud	Not supported
Salesforce	- Add field - Modify field - Drop field - Rename field
SAP	Not supported
ServiceNow	- Add column - Modify column - Drop column
Workday	Not supported
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 detect the schema changes that occur on the source after you deploy a job and before the first run of the job.
- 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.

Configuring application ingestion tasks

Use the application ingestion task wizard in Mass Ingestion to configure application ingestion tasks.

To configure an application ingestion task, perform the following tasks on the wizard:

1. "Defining basic task information" on page 37, such as the task name, project location, runtime environment, and load type.
2. "Configuring the source" on page 38.
3. "Configuring the target" on page 72.
4. "Configuring schedule and runtime options" on page 93.

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. In Mass Ingestion, click **New** > **Application Ingestion Task**.

   The **Definition** page of the application ingestion task wizard appears.
2. Configure the following properties:

Property	Description
Name	Name of the application ingestion task. 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 (-). Task names are not case sensitive. The maximum length is 50 characters. **Note:** 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.
Location	Project or folder in which you want to store the task.

Property	Description
Runtime Environment	Runtime environment in which you want to run the task. 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. **Note:** You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
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: - **Initial Load**: 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. - **Incremental Load**: 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. - **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.

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 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 39
   - "Configuring a Google Analytics source" on page 42
   - "Configuring a Marketo source" on page 45
   - "Configuring a Microsoft Dynamics 365 source" on page 47
   - "Configuring a NetSuite source" on page 50
   - "Configuring an Oracle Fusion Cloud source" on page 52
   - "Configuring a Salesforce source" on page 55
   - "Configuring a Salesforce Marketing Cloud source" on page 58
   - "Configuring a SAP source" on page 61
   - "Configuring a ServiceNow source" on page 64

- "Configuring a Workday source" on page 66
- "Configuring a Zendesk source" on page 69

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 one of the following options:
   - **Select All**: Selects all objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
3. If you selected **Rule-based Selection**, create the rules to select the source objects that you want to replicate on the target.

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

   Perform the following steps to create an object selection rule:

   a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

      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 an object-name mask.
      - Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
      - If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
      - If an object 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 object 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 object 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 71.

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

e. To preview the objects to be selected based on all rules, click **Preview Selection**.

The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** 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 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
8. 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
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 Coordinated Universal Time (UTC).
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 1 day.
11. 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.
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 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 one of the following options:
   - **Select All**: Selects all reports on the source for replication. The **Reports Selected** field displays the total number of reports selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific reports that you want to replicate. This option also enables you to define rules to perform trim actions on the columns of any selected reports.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
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 71.

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 Objects Selected** field.

e. To preview the reports to be selected based on all rules, click **Preview Selection**.

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

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

a. From the **List Objects** 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: - **E**. The report is excluded from processing by an *Exclude* rule. - **I**. The report is included for processing. - **X**. 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.
report_name	Name of the source report.
report_type	Type of the source report. The possible values are: - **O**: Indicates a report. - **F**: Indicates a column.
column_name	Name of the source column. This information appears only if you selected the **Include Columns** check box before downloading the list.
comment	Reason why a source report is excluded from processing even though it matches the selection rules.

9. Expand the **Advanced** section.
10. 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.
11. For initial load tasks, specify the date and time when the ingestion job should stop replicating the source data.
12. 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.
13. 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.
14. 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.
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 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 one of the following options:
   - **Select All**: Selects all objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
2. If you selected **Rule-based Selection**, create the rules to select the source objects that you want to replicate on the target.

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

   Perform the following steps to create an object selection rule:

   a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

      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 an object-name mask.
      - Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
      - If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
      - If an object 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 object 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 object 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 71.

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

e. To preview the objects to be selected based on all rules, click **Preview Selection**.

The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
MARKETO	Name of the source application.

Field	Description
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

5. Expand the **Advanced** section.
6. 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.
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 Coordinated Universal Time (UTC).
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 **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 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 one of the following options:
   - **Select All**: Selects all tables on the source for replication. The **Tables Selected** field displays the total number of tables selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific tables that you want to replicate. This option also enables you to define rules to perform trim actions on the columns of any selected table.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
2. If you selected **Rule-based Selection**, create the rules to select the source tables that you want to replicate on the target.

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

Perform the following steps to create a table selection rule:

a. Select **Table 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 table name or a table-name mask to specify the source tables that you want to include in or exclude from the list of selected tables.

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 table-name mask.
- Table names are case sensitive. When you define the table selection rules, you must specify the table names or masks in the case in which they are defined on the source.
- If a table name contains delimiters, such as quotation marks or brackets, do not include them when you specify the table name in the rule.
- If a table 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 table 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 table 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 71.

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

e. To preview the tables to be selected based on all rules, click **Preview Selection**.

The tables are listed on the **Selected Tables** tab. The list shows the table names and column count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total tables count and the list of tables 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 tables, from Comfortable (most spacing) to Compact (least spacing).

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

   a. From the **List Tables** 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: - **E**. The table is excluded from processing by an *Exclude* rule. - **I**. The table is included for processing. - **X**. 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.
table_name	Name of the source table.
table_type	Type of the source object. The possible values are: - **O**: Indicates a table. - **F**: Indicates a column.
column_name	Name of the source column. This information appears only if you selected the **Include Columns** check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

5. For incremental load tasks and combined initial and incremental load tasks, expand the **Advanced** section.

6. 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).

7. 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 **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.

9. 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.

1. In the **Table Selection** section, select one of the following options:
   - **Select All**: Selects all tables on the source for replication. The **Tables Selected** field displays the total number of tables selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific tables that you want to replicate. This option also enables you to define rules to perform trim actions on the columns of any selected table.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
2. If you selected **Rule-based Selection**, create the rules to select the source tables that you want to replicate on the target.

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

   Perform the following steps to create a table selection rule:

   a. Select **Table 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 table name or a table-name mask to specify the source tables that you want to include in or exclude from the list of selected tables.

      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 table-name mask.
      - Table names are case sensitive. When you define the table selection rules, you must specify the table names or masks in the case in which they are defined on the source.
      - If a table name contains delimiters, such as quotation marks or brackets, do not include them when you specify the table name in the rule.
      - If a table 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 table 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 table 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 71.

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

e. To preview the tables to be selected based on all rules, click **Preview Selection**.

The tables are listed on the **Selected Tables** tab. The list shows the table names and column count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total tables count and the list of tables 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 tables, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Tables** 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: - **E**. The table is excluded from processing by an *Exclude* rule. - **I**. The table is included for processing. - **X**. 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.
table_name	Name of the source table.
table_type	Type of the source object. The possible values are: - **O**: Indicates a table. - **F**: Indicates a column.
column_name	Name of the source column. This information appears only if you selected the **Include Columns** check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

5. Expand the **Advanced** section.
6. 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).
7. 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 5000.
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 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. From the **Oracle Fusion Application** list, select the application from which you want to replicate data.
2. In the **Object Selection** section, select one of the following options:
   - **Select All**: Selects all objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.

- **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

By default, the **Rule-based Selection** option is selected.

Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.

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

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

   Perform the following steps to create an object selection rule:

   a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

      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 an object-name mask.
      - Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
      - If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
      - If an object 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 object 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 object 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 71.

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

   e. To preview the objects to be selected based on all rules, click **Preview Selection**.

      The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

      Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** 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 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.

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. The default value is 50000.
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 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.
   - **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 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 one of the following options:

   - **Select All**: Selects all objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.

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

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

Perform the following steps to create an object selection rule:

a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

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 an object-name mask.
- Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
- If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
- If an object 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 object 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 object 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 71.

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

e. To preview the objects to be selected based on all rules, click **Preview Selection**.

The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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. To download a list of source objects that match the selection rules, perform the following steps:
   a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** 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 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. 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.

10. 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.
11. For initial load and combined initial and incremental load tasks, select **Enable Partitioning** to partition the source objects for initial loading, and then enter the number of partitions you want to create. The default number of partitions is 5.

    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.
12. 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.
13. 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.
14. 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.
15. 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 one of the following options:
    - **Select All**: Selects all data extension objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.
    - **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

    By default, the **Rule-based Selection** option is selected.

    Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
3. If you selected **Rule-based Selection**, create the rules to select the source objects that you want to replicate on the target.

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

Perform the following steps to create an object selection rule:

a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

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 an object-name mask.
- Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
- If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
- If an object 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 object 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 object 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 71.

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

e. To preview the objects to be selected based on all rules, click **Preview Selection**.

The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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. To download a list of source objects that match the selection rules, perform the following steps:
   a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** 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. 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.
8. 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.
9. Click **Next**.

Configuring a SAP source

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. 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.
2. In the **Data Source Selection** section, select one of the following options:
   - **Select All**: Selects all data sources on the source for replication. The **Objects Selected** field displays the total number of data sources selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific data sources that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected data source.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
3. If you selected **Rule-based Selection**, create the rules to select the data sources that you want to replicate on the target.

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

   Perform the following steps to create a data sources selection rule:

   a. Select **Data Source 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 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. Click **Add Rule**.

      The rule appears in the **Rules** list.
      To refine the selection, you can define additional *Include* rules and *Exclude* rules. The data source 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 71.

After you create the rules, you can click **Data Source Count** to display the number of data sources that match each rule in the **Data Sources Affected** column and the total number of data sources selected based on all the selection rules in the **Total Data Sources Selected** field.

e. To preview the data sources to be selected based on all rules, click **Preview Selection**.

The data sources are listed on the **Selected Data Sources** tab. The list shows the data source names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total data source count and the list of data sources 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 data sources, from Comfortable (most spacing) to Compact (least spacing).

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. To download a list of data sources that match the selection rules, perform the following steps:

a. From the **List Data Sources** list, select the type of selection rule for which you want to download the list of selected data sources.

b. If you want to include the fields in the list, select **Include Fields**.

c. 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: - **E**. The data source is excluded from processing by an *Exclude* rule. - **I**. The data source is included for processing. - **X**. 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.
data source_name	Name of the data source.
data source_type	Type of the source object. The possible values are: - **O**: Indicates a data source. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** check box before downloading the list.
comment	Reason why a data source 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: By default, the ingestion job retrieves the change records from the latest available position in the data stream.
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.

   Note: The CDC interval must be less than the data retention period configured in the SAP system for the Operational Delta Queue (ODQ).
9. 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.
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 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 one of the following options:
   - **Select All**: Selects all tables on the source for replication. The **Tables Selected** field displays the total number of tables selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific tables that you want to replicate. This option also enables you to define rules to perform trim actions on the columns of any selected table.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
2. If you selected **Rule-based Selection**, create the rules to select the source tables that you want to replicate on the target.

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

   Perform the following steps to create a table selection rule:

   a. Select **Table 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 table name or a table-name mask to specify the source tables that you want to include in or exclude from the list of selected tables.

      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 table-name mask.
      - Table names are case sensitive. When you define the table selection rules, you must specify the table names or masks in the case in which they are defined on the source.
      - If a table name contains delimiters, such as quotation marks or brackets, do not include them when you specify the table name in the rule.
      - If a table 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 table 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 table 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 71.

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

e. To preview the tables to be selected based on all rules, click **Preview Selection**.

The tables are listed on the **Selected Tables** tab. The list shows the table names and column count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total tables count and the list of tables 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 tables, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Tables** 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: - **E**. The table is excluded from processing by an *Exclude* rule. - **I**. The table is included for processing. - **X**. 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.
table_name	Name of the source table.

Field	Description
table_type	Type of the source object. The possible values are: - **O**: Indicates a table. - **F**: Indicates a column.
column_name	Name of the source column. This information appears only if you selected the **Include Columns** check box before downloading the list.
comment	Reason why a source table is excluded from processing even though it matches the selection rules.

5. 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).

6. 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.
7. 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.
8. 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.
9. 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 one of the following options:
   - **Select All**: Selects all operations on the source for replication. The **Operations Selected** field displays the total number of operations selected for replication.
   - **Rule-based Selection**: Enables you to define rules to select only the specific operations that you want to replicate.

   By default, the **Rule-based Selection** option is selected.

   Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.
5. If you selected **Rule-based Selection**, create the rules to select the source operations that you want to replicate on the target.

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

   Perform the following steps to create an operation selection rule:
   a. Select **Operation 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 an operation name or an operation-name mask to specify the source operations that you want to include in or exclude from the list of selected operations.

      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 an operation-name mask.
      - Operation names are case sensitive. When you define the operation selection rules, you must specify the operation names or masks in the case in which they are defined on the source.
      - If an operation name contains delimiters, such as quotation marks or brackets, do not include them when you specify the operation name in the rule.
      - If an operation 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 operation 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 operation 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 71.

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

e. To preview the operations to be selected based on all rules, click **Preview Selection**.

The operations are listed on the **Selected Operations** tab. The list shows the operation names and column count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total operations count and the list of operations 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 operations, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Operations** 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: - **E**. The operation is excluded from processing by an *Exclude* rule. - **I**. The operation is included for processing. - **X**. 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.
operation_name	Name of the source operation.
operation_type	Type of the source object. The value O 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.

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 Coordinated Universal Time (UTC).

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 100.

Note: The **Fetch Size** field appears only for the SOAP API.

10. 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.

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 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 one of the following options:

    - **Select All**: Selects all objects on the source for replication. The **Objects Selected** field displays the total number of objects selected for replication.
    - **Rule-based Selection**: Enables you to define rules to select only the specific objects that you want to replicate. This option also enables you to define rules to perform trim actions on the fields of any selected object.

    By default, the **Rule-based Selection** option is selected.

    Note: If you select **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are retained.

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

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

    Perform the following steps to create an object selection rule:

    a. Select **Object 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 an object name or an object-name mask to specify the source objects that you want to include in or exclude from the list of selected objects.

    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 an object-name mask.
    - Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
    - If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name in the rule.
    - If an object 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 object 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 object 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 71.

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

e. To preview the objects to be selected based on all rules, click **Preview Selection**.

The objects are listed on the **Selected Objects** tab. The list shows the object names and field count.

Tip: Click the Refresh icon next to the **Updated** date to refresh the total objects count and the list of objects 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 objects, from Comfortable (most spacing) to Compact (least spacing).

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

a. From the **List Objects** 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: - **E**. The object is excluded from processing by an *Exclude* rule. - **I**. The object is included for processing. - **X**. 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.
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: - **O**: Indicates an object. - **F**: Indicates a field.
field_name	Name of the source field. This information appears only if you selected the **Include Fields** check box before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

5. For incremental load tasks and combined initial and incremental load tasks, expand the **Advanced** section.
6. 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).
7. 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 **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.
9. 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

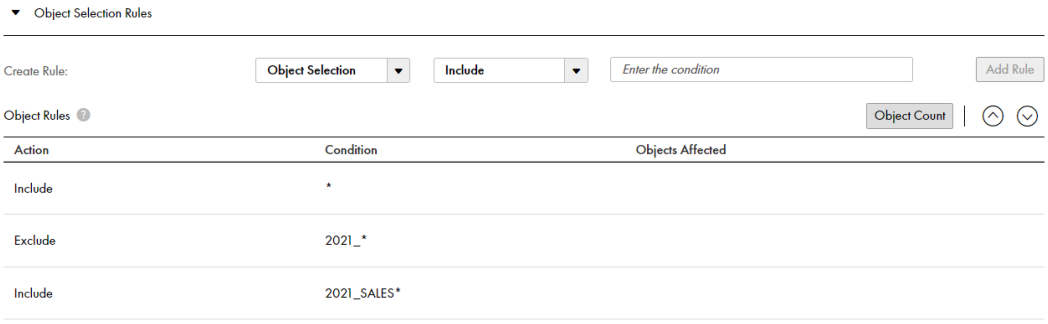
A source has 1,000 objects with different prefixes. You want to select the objects that have the prefix "2021_SALES" and all objects with other prefixes except "2021_".

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:



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 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 75
   - "Amazon S3 target properties" on page 75
   - "Google BigQuery target properties" on page 81
   - "Google Cloud Storage target properties" on page 81
   - "Kafka target properties" on page 85
   - "Microsoft Azure Data Lake Storage Gen2 target properties" on page 87
   - "Microsoft Azure Synapse Analytics target properties" on page 91
   - "Oracle target properties" on page 92
   - "Snowflake target properties" on page 92

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 73.

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.
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:

- 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 **Create Target Tables**, 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.
Directory	Specifies the subdirectory where Mass Ingestion Applications stores output files for this job.

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: - **CSV** - **AVRO** - **PARQUET** The default value is **CSV**. **Note:** Output files in CSV format use double-quotation marks ("") as the delimiter for each field.
Add Headers to CSV File	If **CSV** 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 **PARQUET** output format is selected, you can select a compression type that is supported by Parquet. Options are: - **None** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.
Avro Format	If you selected **AVRO** as the output format, select the format of the Avro schema that will be created for each source table. Options are: - **Avro-Flat**. This Avro schema format lists all Avro fields in one record. - **Avro-Generic**. This Avro schema format lists all columns from a source table in a single array of Avro fields. - **Avro-Nested**. This Avro schema format organizes each type of information in a separate record. The default value is **Avro-Flat**.

Property	Description
Avro Serialization Format	If **AVRO** is selected as the output format, select the serialization format of the Avro output file. Options are: - **None** - **Binary** - **JSON** The default value is **Binary**.
Avro Schema Directory	If **AVRO** 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: *schemaname_tablename*.txt **Note:** 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: - **None** - **Deflate** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.
Avro Compression Type	If **AVRO** is selected as the output format, select an Avro compression type. Options are: - **None** - **Bzip2** - **Deflate** - **Snappy** The default value is **None**, which means no compression is used.
Deflate Compression Level	If **Deflate** is selected in the **Avro Compression Type** field, specify a compression level from 0 to 9. The default value is 0.
Add Directory Tags	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.
Task Target Directory	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 **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. This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.

Property	Description

Data Directory	*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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.
	- Specific directory names.
	- The toUpper() and toLower() functions, which force the values for an associated (*placeholder*) to uppercase or lowercase.
	Note: Placeholder values are not case sensitive.
	Examples:
	`myDir1/{SchemaName}/{TableName}`
	`myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp}`
	`myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}`
	The default directory pattern is `{TableName)_{Timestamp}`.
	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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.
	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.
	- Specific directory names.
	The default directory pattern is `{TaskTargetDirectory}/data/{TableName}/data`
	Note: 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 **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.

Connection Directory as Parent	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 **Data Directory** and **Schema Directory**. For incremental load and combined initial and incremental load tasks, the parent directory is used in the **Data Directory**, **Schema Directory**, **Cycle Completion Directory**, and **Cycle Contents Directory**. This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the **Data Directory** field. For incremental loads, optionally specify a root directory for the task in the **Task Target Directory**.

Property	Description
Schema Directory	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. 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` You can use the same placeholders as for the **Data Directory** field. Ensure that you enclose placeholders with curly brackets { }. 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: `{toLower(SchemaName)}` **Note:** Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.
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 `{TaskTargetDirectory}/cycle/completed`.
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 **Use Cycle Partitioning for Summary DIrectories** 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 **Use Cycle Partitioning for Data Directory** 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 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 includes 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 includes the source SQL operation time 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 cleared.
Add Operation Owner	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. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared.
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 cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

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 **Create Target Tables**, 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	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: - **Standard**. 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. - **Soft Deletes**. 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. Default is Standard.
Directory	Specifies the subdirectory where Mass Ingestion Applications stores output files for this job.

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	Add a metadata column that includes the source SQL operation type in the output that the job propagates to the target. The job writes "I" for insert, "U" for update, or "D" for delete. By default, this check box is selected. You cannot deselect it.
Add Operation Time	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. By default, this check box is not selected.
Add Operation Owner	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. By default, this check box is not selected. **Note:** This property is not available for jobs that have a PostgreSQL source.
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. By default, this check box is not selected.
Add Columns Prefix	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_.

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 **Create Target Tables**, which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.
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.
Directory	Specifies the virtual directory for the Google Cloud Storage target objects that contain the data.

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: - **CSV** - **AVRO** - **PARQUET** The default value is **CSV**. **Note:** Output files in CSV format use double-quotation marks ("") as the delimiter for each field.
Add Headers to CSV File	If **CSV** 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 **PARQUET** output format is selected, you can select a compression type that is supported by Parquet. Options are: - **None** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.
Avro Format	If you selected **AVRO** as the output format, select the format of the Avro schema that will be created for each source table. Options are: - **Avro-Flat**. This Avro schema format lists all Avro fields in one record. - **Avro-Generic**. This Avro schema format lists all columns from a source table in a single array of Avro fields. - **Avro-Nested**. This Avro schema format organizes each type of information in a separate record. The default value is **Avro-Flat**.

Property	Description
Avro Serialization Format	If **AVRO** is selected as the output format, select the serialization format of the Avro output file. Options are: - **None** - **Binary** - **JSON** The default value is **Binary**.
Avro Schema Directory	If **AVRO** 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: *schemaname_tablename*.txt **Note:** 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: - **None** - **Deflate** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.
Avro Compression Type	If **AVRO** is selected as the output format, select an Avro compression type. Options are: - **None** - **Bzip2** - **Deflate** - **Snappy** The default value is **None**, which means no compression is used.
Deflate Compression Level	If **Deflate** is selected in the **Avro Compression Type** field, specify a compression level from 0 to 9. The default value is 0.
Add Directory Tags	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.
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.
Task Target Directory	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 **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. This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.

Property	Description

Data Directory	*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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.
	- Specific directory names.
	- The toUpper() and toLower() functions, which force the values for an associated (*placeholder*) to uppercase or lowercase.
	Note: Placeholder values are not case sensitive.
	Examples:
	`myDir1/{SchemaName}/{TableName}`
	`myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp}`
	`myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}`
	The default directory pattern is `{TableName)_{Timestamp}`.
	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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.
	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.
	- Specific directory names.
	The default directory pattern is `{TaskTargetDirectory}/data/{TableName}/data`
	Note: 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 **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.

Schema Directory	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. 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` You can use the same placeholders as for the **Data Directory** field. Ensure that you enclose placeholders with curly brackets { }. 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: `{toLower(SchemaName)}` **Note:** Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.
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 `{TaskTargetDirectory}/cycle/completed`.
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`.

Property	Description
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 **Use Cycle Partitioning for Summary DIrectories** 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 **Use Cycle Partitioning for Data Directory** 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 includes 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 includes the source SQL operation time 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 cleared.
Add Operation Owner	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. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared.
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 cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

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	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. 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 **Include Schema Name**, **Table Prefix**, or **Table Suffix** properties. By default, this check box is cleared.
Include Schema Name	When **Use Table Name as Topic Name** 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 *schemaname_tablename*. If you do *not* want to include the schema name, clear this check box.
Table Prefix	When **Use Table Name as Topic Name** 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 myprefix_, the topic names have the format myprefix_*tablename*. If you omit the underscore (_) after the prefix, the prefix is prepended to the table name.
Table Suffix	When **Use Table Name as Topic Name** 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 _mysuffix, the topic names have the format *tablename*_mysuffix. If you omit the underscore (_) before the suffix, the suffix is appended to the table name.
Output Format	Select the format of the output file. Options are: - **CSV** - **AVRO** - **JSON** The default value is **CSV**. **Note:** Output files in CSV format use double-quotation marks ("") as the delimiter for each field. If your Kafka target uses Confluent Schema Registry to store schemas for incremental load jobs, you must select **AVRO** as the format.
JSON Format	If **JSON** is selected as the output format, select the level of detail of the output. Options are: - **Concise**. This format records only the most relevant data in the output, such as the operation type and the column names and values. - **Verbose**. This format records detailed information, such as the table name and column types.
Avro Format	If you selected **AVRO** as the output format, select the format of the Avro schema that will be created for each source table. Options are: - **Avro-Flat**. This Avro schema format lists all Avro fields in one record. - **Avro-Generic**. This Avro schema format lists all columns from a source table in a single array of Avro fields. - **Avro-Nested**. This Avro schema format organizes each type of information in a separate record. The default value is **Avro-Flat**.

Property	Description
Avro Serialization Format	If **AVRO** is selected as the output format, select the serialization format of the Avro output file. Options are: - **Binary** - **JSON** - **None** The default value is **Binary**. If you have a Confluent Kafka target that uses Confluent Schema Registry to store schemas, select **None**. Otherwise, Confluent Schema Registry does not register the schema. Do not select **None** if you are not using Confluent Scheme Registry.

Avro Schema Directory	If **AVRO** 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:
	`schemaname_tablename.txt`
	Note: If this directory is not specified, no Avro schema definition file is produced.
	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:
	`schemaname_tablename_YYYYMMDDhhmmss.txt`
	This unique naming pattern ensures that older schema definition files are preserved for audit purposes.

Avro Compression Type	If **AVRO** is selected as the output format, select an Avro compression type. Options are: - **None** - **Bzip2** - **Deflate** - **Snappy** The default value is **None**, which means no compression is used.

The following table describes the advanced Kafka target properties that appear under **Advanced**:

Property	Description
Add Operation Type	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. 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.
Add Operation Time	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. For initial loads, the job always writes the current date and time. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

Property	Description
Async Write	Controls whether to use synchronous delivery of messages to Kafka. - 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. - 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. By default, this check box is selected.
Producer Configuration Properties	Specify a comma-separated list of *key=value* pairs to enter Kafka producer properties for Apache Kafkatargets. You can specify Kafka producer properties in either this field or in the **Additional Connection Properties** field in the Kafka connection. 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 **Producer Configuration Properties** field. For information about Kafka producer properties, see the Apache Kafka documentation.

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	Select the format of the output file. Options are: - **CSV** - **AVRO** - **PARQUET** The default value is **CSV**. **Note:** Output files in CSV format use double-quotation marks ("") as the delimiter for each field.
Add Headers to CSV File	If **CSV** 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 **PARQUET** output format is selected, you can select a compression type that is supported by Parquet. Options are: - **None** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.

Property	Description
Avro Format	If you selected **AVRO** as the output format, select the format of the Avro schema that will be created for each source table. Options are: - **Avro-Flat**. This Avro schema format lists all Avro fields in one record. - **Avro-Generic**. This Avro schema format lists all columns from a source table in a single array of Avro fields. - **Avro-Nested**. This Avro schema format organizes each type of information in a separate record. The default value is **Avro-Flat**.
Avro Serialization Format	If **AVRO** is selected as the output format, select the serialization format of the Avro output file. Options are: - **None** - **Binary** - **JSON** The default value is **Binary**.
Avro Schema Directory	If **AVRO** 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: *schemaname_tablename*.txt **Note:** 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: - **None** - **Deflate** - **Gzip** - **Snappy** The default value is **None**, which means no compression is used.
Avro Compression Type	If **AVRO** is selected as the output format, select an Avro compression type. Options are: - **None** - **Bzip2** - **Deflate** - **Snappy** The default value is **None**, which means no compression is used.
Deflate Compression Level	If **Deflate** is selected in the **Avro Compression Type** field, specify a compression level from 0 to 9. The default value is 0.
Add Directory Tags	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.
Task Target Directory	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 **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. This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.

Property	Description

Data Directory	*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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.
	- Specific directory names.
	- The toUpper() and toLower() functions, which force the values for an associated (*placeholder*) to uppercase or lowercase.
	Note: Placeholder values are not case sensitive.
	Examples:
	`myDir1/{SchemaName}/{TableName}`
	`myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp}`
	`myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}`
	The default directory pattern is `{TableName)_{Timestamp}`.
	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:
	- 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 yyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.
	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.
	- Specific directory names.
	The default directory pattern is `{TaskTargetDirectory}/data/{TableName}/data`
	Note: 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 **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.

Connection Directory as Parent	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 **Data Directory** and **Schema Directory**. For incremental load and combined initial and incremental load tasks, the parent directory is used in the **Data Directory**, **Schema Directory**, **Cycle Completion Directory**, and **Cycle Contents Directory**. This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the **Data Directory** field. For incremental loads, optionally specify a root directory for the task in the **Task Target Directory**.

Property	Description
Schema Directory	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. 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` You can use the same placeholders as for the **Data Directory** field. Ensure that you enclose placeholders with curly brackets { }. 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: `{toLower(SchemaName)}` **Note:** Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.
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 `{TaskTargetDirectory}/cycle/completed`.
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 **Use Cycle Partitioning for Summary DIrectories** 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 **Use Cycle Partitioning for Data Directory** 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 Microsoft Azure Data Lake Storage Gen2 advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	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. 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 includes the source SQL operation time 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 cleared.
Add Operation Owner	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. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared.
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 cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

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 **Create Target Tables**, which generates the target tables based on the source objects.
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 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 **Create Target Tables**, which generates the target tables based on the source objects.
Schema	Select 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 **Create Target Tables**, 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	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: - **Standard**. 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. - **Soft Delete**. 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. Default is Standard.

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 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 database or inserts into the audit table on the target system. For incremental loads and combined initial and 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. You cannot deselect it if you are using soft deletes.
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 database or inserts into the audit table on the target system. By default, this check box is not selected.
Audit Columns Prefix	Select this check box to 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_.

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. **Note:** - Add Field operations that add a primary-key field are not supported and might cause unpredictable results. - Modify Field operations that change the NULL or NOT NULL constraint of a field are not replicated to the target.

Option	Description
Stop Job	Stops the application ingestion job.
Stop Object	Stops processing the source object on which the DDL change occurred. **Note:** When one or more objects are excluded from replication because of the Stop Object schema drift option, the status of the job changes to **Running with Warning**. 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.

2. 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 Redishift, Amazon S3, Google Big Query, Google Cloud Storage, Microsoft Azure Data Lake Storage, Microsoft Azure Synapse Analytics, 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.

3. 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, and Microsoft Azure Data Lake Storage 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.

4. For application ingestion incremental load tasks that have Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage Gen2 targets, 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.
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 **Apply Cycle Change Limit** or **Apply Cycle Interval** 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. **Note:** If this check box is selected, the **Checkpoint Every Commit**, **Checkpoint Row Count**, and **Checkpoint Frequency (secs)** 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 successfully deploy a task, an application ingestion job is created and the status of the job is Deployed. You can run the job from the **My Jobs** page in Mass Ingestion 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 Mass Ingestion 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 the Mass Ingestion service 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.

You can stop the 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 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.

▶ On the **Actions** menu for the job that you want to stop, select **Stop**.

The status of the job changes to Stopping and then changes to Stopped.

Tip: If the job takes too long to stop, you can abort the job.

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.

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, Mass Ingestion Applications resumes propagating source data from where it last left off.

- On the **Actions** menu for the job that you want to run, click **Resume**.

  A subtask is started for each source object.

  Note: The **Resume** option is not available if the job is in Failed state because the task deployment failed.

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

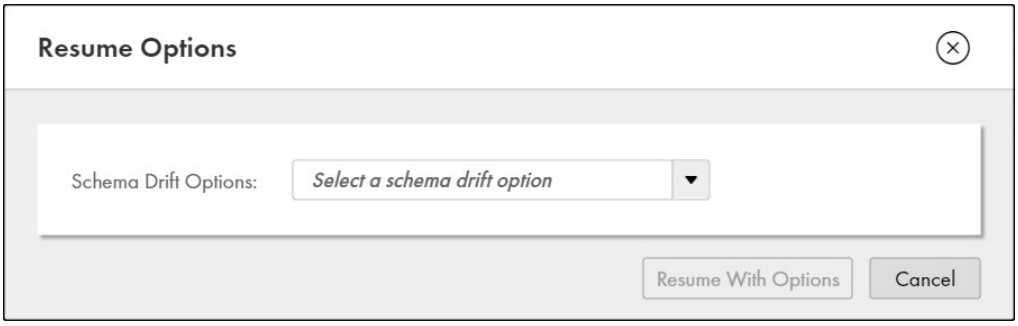
because of the **Stop Object**, **Stop Table**, **Stop Report**, or **Stop Job** schema drift option. Use the overrides to correct or resolve these errors.

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.



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	Stop processing the source object on which the DDL change occurred. **Important:** 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 **Resume With Options** > **Resync** option.
Resync	Resynchronize the target table with the source object. Use this option for objects that the job stopped processing because of the **Stop Object**, **Stop Table**, or **Stop Report** setting for a **Schema Drift** option. **Important:** 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. **Important:** If you specify the **Replicate** 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 100.

When you redeploy a job, Mass Ingestion Applications stops all subtasks for the source objects. After the subtasks are stopped, Mass Ingestion Applications 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. On the **My Jobs** page, navigate to the row for the job that you want to redeploy.
2. On the **Actions** menu for the row, click **Redeploy**.

   The job instance starts running with the updated configurations.

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.

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. Ensure that the job is not running.
2. On the **My Jobs** page in Mass Ingestion or on the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights, navigate to the row for the job that you want to undeploy.
3. On the **Actions** menu for the row, click **Undeploy**.

   The Undeploy option is not available for jobs that are in Failed status and for jobs that have not been previously deployed.

   If Mass Ingestion Applications fails to undeploy the job, the status of the job changes to Failed irrespective of its current status.

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 98.

1. On the **My Jobs** page in the Mass Ingestion service or on the **All Jobs** tab of the **Mass Ingestion** page in Operational Insights, drill down on an ingestion job to display job details.

   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**.

   For the Actions menu and **Resync** option to be available, the subtask must be in a state other than Queued or Starting.

If the source object 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.

Important: If the source object contains many rows, the resynchronization might take a long time to perform.

INDEX