Informatica® Data Ingestion and Replication
July 2024

Data Ingestion and Replication What's New

Informatica Data Ingestion and Replication Data Ingestion and Replication What's New
July 2024

© Copyright Informatica LLC 2019, 2024

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

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

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

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

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

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

Publication Date: 2024-07-19

Table of Contents

Preface 4

Chapter 1: July 2024 5

New features and enhancements 5

Common 5

Application Ingestion and Replication 6

Database Ingestion and Replication 6

Streaming Ingestion and Replication 8

Changed behavior 8

Post-upgrade tasks for the July 2024 release 9

Chapter 2: May 2024 10

New features and enhancements 10

Common 10

Mass Ingestion Databases 12

Changed behavior 13

Chapter 3: April 2024 14

New features and enhancements 14

Common 14

Mass Ingestion Applications 15

Mass Ingestion Databases 15

Mass Ingestion Streaming 16

Changed behavior 16

Chapter 4: February 2024 17

New features and enhancements 17

Important notices 17

Common 17

Mass Ingestion Applications 18

Mass Ingestion Databases 19

Changed behavior 19

Index 21

Preface

Read *Data Ingestion and Replication What's New* for a brief overview of new features in the Data Ingestion and Replication service, which includes Application Ingestion and Replication, Database Ingestion and Replication, File Ingestion and Replication, and Streaming Ingestion and Replication.

CHAPTER 1

July 2024

This section provides information about new features, enhancements, and behavior changes in the July 2024 release of Data Ingestion and Replication.

New features and enhancements

Read about new features and enhancements in the July 2024 Data Ingestion and Replication release.

Common

The July 2024 release of Data Ingestion and Replication includes the following new features that are common to multiple types of ingestion and replication tasks.

Mass Ingestion has a new name!

The Mass Ingestion service is now called *Data Ingestion and Replication* to reflect its full range of data replication, synchronization, and CDC capabilities and to better align with market perceptions of its functionality. Correspondingly, the task names and solution names have also changed:

New solution names

- Application Ingestion and Replication
- Database Ingestion and Replication
- File Ingestion and Replication
- Streaming Ingestion and Replication

New task names

- Application ingestion and replication task
- Database ingestion and replication task
- File ingestion and replication task
- Streaming ingestion and replication task

These name changes have been made throughout most of the user interface and documentation set. For example, on the **My Services** page, you'll now see the **Data Ingestion and Replication** box instead of the **Mass Ingestion** box. In a few places, the original names have been maintained to avoid any disruption, such as for the Secure Agent's "Database Ingestion" service and in connector names that include "Mass Ingestion" or "Database Ingestion."

Unified Home page enhancement

The latest unified Home page is now available to *all* Data Ingestion and Replication users, including those who use the localized Japanese version of the user interface and new customers who onboarded to Data Ingestion and Replication after the May 2024 release.

Ability to add _OLD columns with before-images of Updates to Oracle targets when using Audit mode.

For application ingestion and replication jobs and database ingestion and replication jobs that have an Oracle target and use Audit apply mode, you can add _OLD metadata columns that contain before-image data for Updates to the target tables. You can use these columns to compare the old and new column values.

To add the _OLD columns to the target tables, select the **Add Before Images** check box in the **Advanced** section on the **Target** page of the task wizard.

Resync commands are now available in the Data Ingestion and Replication Command-Line Interface

The Data Ingestion and Replication Command-Line Interface (CLI) provides new job resycSubtask commands to resynchronize source and target objects for a subtask that is part of a running combined initial and incremental load job. This enhancement is available for application ingestion and replication jobs and database ingestion and replication jobs.

Previously, you had to use the Resync options in one of the monitoring user interfaces.

Schema drift support for SQL Server targets

Application Ingestion and Replication and Database Ingestion and Replication add support for automatic schema drift detection and handling for incremental load and combined initial and incremental load jobs that have SQL Server targets. When you create application ingestion and replication tasks or database ingestion and replication tasks, you can now set schema drift options on the **Schedule and Runtime Options** page. For pre-existing tasks, the schema options are set to Ignore.

Application Ingestion and Replication

The July 2024 release of Application Ingestion and Replication includes the following new features and enhancements:

Support for LOOKUP data type for Microsoft Dynamics 365 sources

Application Ingestion and Replication supports the LOOKUP data type for Microsoft Dynamics 365 sources.

Schema drift support for PostgreSQL targets

Application Ingestion and Replication adds support for automatic schema drift detection and handling for incremental load and combined initial and incremental load jobs that have a Salesforce source and a PostgreSQL target.

Database Ingestion and Replication

The July 2024 release of Database Ingestion and Replication includes the following new features and enhancements:

Support for Oracle source LOB columns in incremental load and combined load jobs

When you create a database ingestion and replication incremental load or combined initial and incremental load task that has an Oracle source with LOB columns and a target type other than Kafka, you can now select the **Include LOBs** option. This option causes change data to be captured from the LOB columns selected for replication. The supported LOB types are BLOB, CLOB, NCLOB, LONG, LONG RAW, and XML.

Note: Columns that have the LONG, LONG RAW, and XML data types are supported in incremental load and combined load jobs that use the **Query-based** CDC method. However, jobs that use the **Log-based** CDC method do not replicate data from these types of columns to the generated target table.

All of the LOB columns continue to be supported in initial load jobs.

Support for Oracle source LOB columns in jobs that have PostgreSQL targets

Database ingestion and replication jobs can now move data from Oracle source columns that have a LOB data type to Amazon Aurora PostgreSQL or RDS for PostgreSQL targets.

Support for Db2 for LUW source LOB columns

Database ingestion and replication initial load jobs and incremental load and combined initial and incremental load jobs that use the Query-based CDC method can replicate data from Db2 for LUW source columns that have a LOB data type to Microsoft Azure Data Lake Storage Gen 2, Microsoft Azure Synapse Analytics, or Snowflake targets. The supported LOB data types are: BLOB, CLOB, DBCLOB, LONG VARCHAR, LONG VARCHAR FOR BIT DATA, LONG VARGRAPHIC, and XML.

PostgreSQL sources with SQL Server targets

Database Ingestion and Replication now supports PostgreSQL sources with Microsoft SQL Server targets in initial load, incremental load, and combined load jobs.

Azure Database for MySQL sources

Database Ingestion and Replication now supports Azure Database for MySQL 8.0 sources with the following targets and load types:

- Snowflake targets in initial load, incremental load, and combined load jobs
- Confluent Kafka targets in incremental load jobs

Audit apply mode for Databricks targets

For database ingestion and replication incremental load and combined initial and incremental load jobs that have Databricks targets, you can configure tasks to use Audit apply mode to write a row for each DML operation on a source table to a generated audit table on the target. Optionally, add columns that contain metadata about the changes, such as SQL operation type, timestamp, owner, transaction ID, and sequence, and before image, to the audit table. This feature is useful when you need an audit trail of changes to perform downstream processing on the data before writing it to the target database or when you need to examine the metadata for changes.

Note: The audit tables cannot have constraints other than indexes.

To enable the use of audit tables, select **Audit** in the **Apply Mode** field on the **Target** page when defining a task. This field is available for new or undeployed tasks. Under **Advanced**, optionally select the check boxes for adding metadata columns to the audit table.

Custom data-type mappings for SQL Server sources and Snowflake targets

For database ingestion and replication tasks that have SQL Server sources and Snowflake targets, you can optionally define custom data-type mapping rules, which will override the default mappings for this source and target.

In the task wizard, you can create data-type mapping rules in the **Data Type Rules** section on the **Target** page.

Schema drift support for PostgreSQL targets

Database Ingestion and Replication adds support for automatic schema drift detection and handling for incremental load and combined initial and incremental load jobs that have a Db2 for i or Oracle source and an Amazon Aurora PostgreSQL or RDS for PostgreSQL target.

Support for the TIMESTAMP WITH LOCAL TIME ZONE data type in Oracle sources

Database Ingestion and Replication now supports Oracle source columns that have the TIMESTAMP WITH LOCAL TIME ZONE data type in jobs with any load type and any supported target type. To process columns that have the TIMESTAMP WITH LOCAL TIME ZONE data type, you must set the DBMI_ORACLE_SOURCE_ENABLE_TIMESTAMP_WITH_LOCAL_TZ environment variable for the Database Ingestion agent service to true. In Administrator, open your Secure Agent and click **Edit** . Under **Custom Configuration Details**, add the environment variable for the **Database Ingestion** service and the **DBMI_AGENT_ENV** type.

New Db2 for i journal receiver exit to prevent journal receiver deletion during CDC processing

Database Ingestion and Replication now provides a Db2 for i journal receiver exit to prevent the deletion of journal receivers while database ingestion and replication incremental load and combined load jobs are reading them for change data capture (CDC). The exit program causes the journal receivers to be locked while in use for CDC. To use the journal receiver exit, you must manually install the exit program and specify the pwx.cdcreader.iseries.option.useJournalReceiverExit and pwx.cdcreader.iseries.option.JournalReceiverExitJobToken custom properties on the **Source** page of the task wizard.

Streaming Ingestion and Replication

The July 2024 release of Streaming Ingestion and Replication includes the following new feature and enhancement:

Support for Business 360 Events Connector

Mass Ingestion Streaming now supports Business 360 Events connector as a source to transfer files.

Changed behavior

The July 2024 release of Data Ingestion and Replication includes the following changed behaviors.

Monitor now shows the Rows Processed count for ingestion and replication jobs

On the **Running Jobs** and **All Jobs** pages in Monitor and on the **My Jobs** page accessed from the Data Integration unified Home page, the **Rows Processed** column now displays values for running and successfully completed application ingestion and replication, database ingestion and replication, and streaming ingestion and replication jobs. For incremental load and combined initial and incremental load jobs, the count is displayed after CDC processing ends. The count is the number of rows written to the target. Previously, only "N/A" was displayed.

Automatic switchover to another Secure Agent in a Secure Agent group

If the active Secure Agent on which application ingestion and replication jobs or database ingestion and replication jobs are running goes down unexpectedly, the jobs can automatically switch over to another available agent in the group after the 15-minute heartbeat interval elapses. You no longer have to Stop and Resume the job for the switchover to occur. Automatic switchovers are supported for:

- Application ingestion and replication and database ingestion and replication initial load jobs that have any source type and a target type other than Kafka

- Database ingestion and replication incremental load and combined load jobs that have an Oracle or SQL Server source, subject to the following limitations:
  - The job cannot have a Kafka target.
  - The job cannot have persistent storage enabled.
  - The job cannot use the Query-based CDC method to capture changes from the Oracle or SQL Server source.

Post-upgrade tasks for the July 2024 release

Perform the following task after Data Ingestion and Replication is upgraded to the July 2024 release.

Upgrade consideration for Google BigQuery using third-party jars

After the July 2024 upgrade, application ingestion and replication jobs and database ingestion and replication jobs might fail if you previously copied the Google BigQuery jar files, including jackson-databind-2.x.x.jar, jackson-core-2.x.x.jar, and jackson-annotations-2.x.x.jar, to the following location:

```
<INFA_HOME>/apps/Database_Ingestion/ext/
```

The jobs fail with the following error because the versions of the earlier jackson-databind-2.x.x.jar, jackson-core-2.x.x.jar, and jackson-annotations-2.x.x.jar files differ from the current 2.17.1 version of the .jar files that Application Ingestion and Replication and Database Ingestion and Replication use:

```
[DBMIP_20009] The process [CDC_COMBINED-SSLRCDCReader-CDC_COMBINED PWX Reader(63)-main]
encountered an unexpected error. Error: [PWX session start failed with
StartSession({host=localhost,port=44991,sl=true,sslTrustStore=}) failed. Caused by:
Failed to open a socket. Error:Connection refused (Connection refused)].]
```

You must delete the older jackson-databind-2.x.x.jar, jackson-core-2.x.x.jar, and jackson-annotations-2.x.x.jar files and replace them with the current jackson-databind-2.17.1.jar, jackson-core-2.17.1.jar, and jackson-annotations-2.17.1.jar files from the following location:

```
<INFA_HOME>/downloads/package-GoogleBigQueryV2.{latestversion}/package/bigqueryv2/
thirdparty/infa.bigqueryv2
```

Then restart the Secure Agent and resume your jobs.

Upgrade consideration for database ingestion and replication jobs that have Db2 for LUW sources

If a database ingestion and replication job that has a Db2 for LUW source tables with LOB columns was running during the upgrade to the July 2024 release and a Drop Column operation occurred on a source table, perform the following steps to ensure that the source and target definitions are still consistent in structure:

1. Stop the job.
2. Edit the associated task to select the **Include LOBs** option for the source.
3. Redeploy the job.

CHAPTER 2

May 2024

This section provides information about new features, enhancements, and behavior changes in the May 2024 release of Informatica Intelligent Cloud Services℠ Data Ingestion and Replication service.

New features and enhancements

Read about new features and enhancements in the May 2024 Mass Ingestion release.

Common

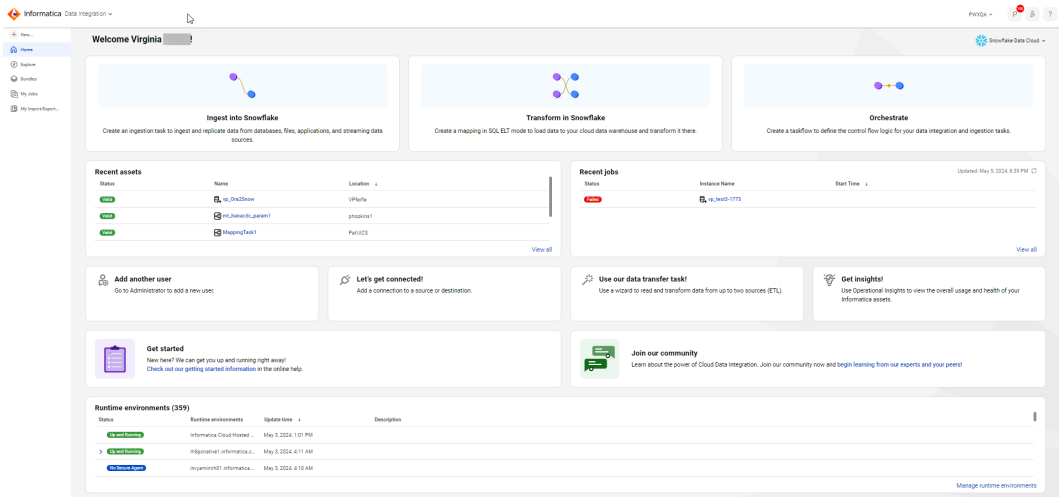
The May 2024 release of Informatica Intelligent Cloud Services Data Ingestion and Replication service includes the following new features that are common to multiple types of ingestion tasks.

Unified Home page

Most Mass Ingestion organizations will see a new Home page that provides a guided and unified experience for Data Integration and Data Ingestion and Replication users.

The Home page provides a central place from which you can create tasks, mappings, taskflows, and other assets, depending on your licensing. You can also access your recently created or updated assets and check the statuses of your recent jobs.

The panels that appear on the **Home** page are customized based on criteria such as the organization's licenses and the user's roles. For example, the following Home page is displayed for an Admin user in an organization that uses both Data Ingestion and Replication and Data Integration:



The navigation bar on the left side still provides the traditional New, Explore, and My Jobs options for creating a new task, exploring all assets, and viewing all of your jobs.

Limitations in the May 2024 release include:

- If you use the localized Japanese version of the user interface, or if you onboard to Data Ingestion and Replication after the May 2024 release is rolled out, the pre-existing Data Integration Home page appears instead of the new unified Home page. On this Home page, you can use the New, Explore, and My Jobs options in the navigation bar to create ingestion tasks, access your assets, and view details for your jobs. The new unified Home page is planned to be available to all Data Ingestion and Replication users in the July 2024 release.
- If you use the **Do you use a cloud data warehouse as your primary destination?** panel on the Home page to define a cloud data warehouse target like Snowflake, the information will not appear on the **Target** page of the task wizard when you create a new ingestion task. However, you can define the target from the task wizard.

New UI look and feel

Along with the new unified home page, the Data Ingestion and Replication user interface has a fresh look and feel with a lighter palette, updated icons, and new fonts.

Monitoring ingestion jobs in the Monitor service and from the unified Home page

You can now view and monitor ingestion jobs, along with other job types, in the Monitor service. You can monitor the progress and status of all jobs and all running jobs on the **All Jobs** and **Running Jobs** pages, respectively. On either page, you can click any job name to drill down to detailed information about the job. You can also perform some actions on the listed jobs, depending on the job status and task type, by using the **Actions** menu at the right end of each job row. To help you find your jobs more easily, the **All Jobs** page provides Filter, Find, and Sort options.

Alternatively, you can monitor all of your jobs, including ingestion and integration jobs, on the **My Jobs** page that is accessed by clicking **My Jobs** on the navigation bar of the unified Home page.

Note: You can continue to monitor only ingestion jobs on the **Mass Ingestion** page in Operational Insights.

Microsoft Fabric OneLake targets in incremental load and combined load jobs

You can now use Microsoft Fabric OneLake as a target for application ingestion and database ingestion incremental load and combined initial and incremental load jobs.

Previously Microsoft Fabric OneLake was available as a target in initial load jobs only.

Secure Agent identified in the application ingestion and database ingestion job details.

The name of the Secure Agent on which an application ingestion or database ingestion job runs is now displayed in the job-specific details. To access the job details, drill down on a job from the **My Jobs** page in the Data Integration service, the **All Jobs** and **Running Jobs** pages in the Monitor service, or from the **Mass Ingestion** page in Operational Insights service.

For an initial load job, the Agent Name is displayed in the **Object Detail** pane. For an incremental load job, the Agent Name is displayed in the **Overview** section of the **Task Summary** pane. For combined initial and incremental load jobs, the Agent Name is displayed in both the **Object Detail** pane and the **Overview** section.

This enhancement was introduced to help identify the Secure Agent for which logs must be collected for troubleshooting when the job runs on a Secure Agent group and one of the tasks fails. Previously, if a task failed and the job was running in a Secure Agent group, you had to manually run the log collector on all of the Secure Agents in the Secure Agent group.

If an incremental load job fails over from one Secure Agent to another, the latest Secure Agent assigned to the job is displayed. If you need to collect logs for a previously assigned Secure Agent, download the complete log from each of the Secure Agents in the Secure Agent group.

Mass Ingestion Databases

The May 2024 release of Database Ingestion and Replication includes the following new features and enhancements:

Data validation for initial load jobs that have an Oracle or SQL Server source and a Snowflake target

Note: The data validation feature availability is controlled by means of an organization-level feature flag. If this functionality is not available for your organization but you want to use it, create a request for Informatica Global Customer Support.

For database ingestion initial load jobs that have an Oracle or SQL Server source and a Snowflake target, you can now use a new data validation mechanism to compare the target data with the corresponding source object. Database Ingestion and Replication checks data consistency to ensure that all data has been correctly replicated to the appropriate target columns without any data conversion or truncation.

The data validation option is available only for tasks that have the status of **Completed**.

You can run data validation in the **Object Detail** pane of monitoring job-specific details. To access the **Object Detail** pane, drill down on a job from the **My Jobs** page in the Data Integration service, the **All Jobs** and **Running Jobs** pages in the Monitor service, or from the **Mass Ingestion** page in Operational Insights service.

When you run data validation in the Database Ingestion and Replication service, you will be charged based on the CPU consumption on the system where the Data Validation service runs.

Enhancements to Db2 for z/OS stored procedure processing

For database ingestion incremental load and combined initial and incremental load jobs that have Db2 for z/OS sources, the following enhancements have been made to Db2 for z/OS stored procedure capture processing:

- You can configure the stored procedure to skip empty URs when collecting log data from the z/OS system to help improve performance. Use the pwx.cdcreader.ZOS.strproc.skipEmptyUR custom property.
- During change data capture, Mass Ingestion Databases can now pass up to 16 KB of source table filtering information to the Db2 for z/OS stored procedure. Previously, the maximum was 4 KB. If the amount of filtered data exceeds 16 KB, Mass Ingestion Databases displays an error message.

Changed behavior

The May 2024 release of Data Ingestion and Replication includes the following changed behaviors.

Monitoring Mass Ingestion jobs in Data Integration and Monitor services

The Data Integration and Monitor services include the following changes for monitoring Mass Ingestion jobs:

- The **My Jobs** page in Data Integration and the **All Jobs** page in Monitor display all of your jobs or all jobs, respectively, but the retention time for undeployed jobs is based on the Job Log Service (JLS) purge policy.
  Both pages display only the new jobs created from the current release. The undeployed jobs from the previous release are excluded.

  Previously, the **My Jobs** page in Data Ingestion and Replication used to display all jobs without purging. Therefore, the number of jobs listed in the Data Ingestion and Replication service could be greater than in the Data Integration service.
- When you deploy an application ingestion task or database ingestion task, the newest job doesn't automatically appear at the top of the jobs list on the **My Jobs** page in Data Integration and the **All Jobs** page in the Monitor. You can sort jobs by using the **Start Time** property or use the find and filter features to search for a specific job.
  Previously, you could sort jobs from newest to oldest on the **My Jobs** page in the Data Ingestion and Replication service.
- When you click the instance name of a file ingestion job on the **My Jobs** page in Data Integration, the job details page appears. When you click the task name from the job details page, it opens in the edit mode.
  Previously, when you clicked the task name from the job details page, the details opened only in the view mode for file ingestion tasks in Mass Ingestion.
- When you undeploy a streaming ingestion and replication job, the history of the undeployed tasks appears on both the **My Jobs** page in Data Integration and the **All Jobs** page in Monitor.
  Previously, only the latest tasks were listed.

The generated name of the job instance contains a hyphen (-) instead of an underscore (_) before the job instance number

The generated name of the job instance now contains a hyphen (-) instead of an underscore (_) between the task name and job instance number on the **My Jobs** page, on the **Running Jobs** and **All Jobs** pages in Monitor, and in the **Recent jobs** panel of the unified Home page:

```
<taskname>-<job_instance_number>
```

In the Cloud Data Ingestion and Replication Command Line Interface (CLI) and in Operational Insights, the format of the job instance name continues to use the underscore (_) before the job instance number:

```
<taskname>_<job_instance_number>
```

In the CLI, you must use the *<taskname>_<job_instance_number* naming convention.

Default age-out period for incremental load processing has increased

Database ingestion incremental load and combined initial and incremental load jobs stop tracking transactions for restart purposes when those transactions have no DML for tables of interest after an age-out period. The age-out period has been increased from 60 minutes to 24 hours.

The restart point where Data Ingestion and Replication starts reading change data might include recapturing empty transactions that are up to 24 hours old.

CHAPTER 3

April 2024

This section provides information about new features, enhancements, and behavior changes in the April 2024 release of Informatica Intelligent Cloud Services℠ Data Ingestion and Replication service.

New features and enhancements

Read about new features and enhancements in the April 2024 Mass Ingestion release.

Watch the What's New video to learn about the new features and enhancements in the April 2024 release.

Common

The April 2024 release of Informatica Intelligent Cloud Services Data Ingestion and Replication service includes the following new features that are common to application ingestion and database ingestion tasks.

Amazon Aurora PostgreSQL targets with Oracle and Salesforce sources

You can use Amazon Aurora PostgreSQL targets in database ingestion jobs that have an Oracle source or in application ingestion jobs that have a Salesforce source. To connect to the Aurora PostgreSQL target, use the PostgreSQL connector.

Previously, you could use Amazon Aurora PostgreSQL as a target only in database ingestion jobs that had a Db2 for i source.

Microsoft Fabric OneLake targets in initial load jobs

You can now use Microsoft Fabric OneLake as a target for application ingestion and database ingestion initial load jobs.

To connect to a Microsoft Fabric OneLake target, use the **Microsoft Fabric OneLake** connector.

Support for Unity Catalog and personal staging locations in Databricks Delta target connections

Mass Ingestion now supports Databricks Delta Unity Catalog and personal staging locations for application ingestion and database ingestion tasks of any load type. When you define a Databricks Delta target connection for a task, you can specify a catalog in the Unity Catalog metastore and indicate whether to use a personal staging location.

In the connection properties, specify the name of the catalog in the **Catalog Name** field. The catalog name is appended to the **SQL Warehouse JDBC URL** value for a data warehouse. The catalog, which contains schemas, is the first layer in the Unity Catalog hierarchy for organizing data assets. Databricks Delta recommends using Unity Catalog to administer data access policies and permissions, capture audit logs that record data access, capture lineage information on data assets, and query account data.

If you use Unity Catalog and want to stage data internally in Databricks Delta instead of using an Azure or AWS staging environment, you can select **Personal Staging Location** in the **Staging Environment** connection property. The Parquet data files for application ingestion or database ingestion jobs can then be staged to a local personal storage location, which has a data retention period of 7 days. If you use Unity Catalog, a personal storage location is automatically provisioned. Personal staging locations do not support Databricks Delta unmanaged tables, which are stored externally.

Mass Ingestion Applications

The April 2024 release of Mass Ingestion Applications includes the following new features and enhancements:

Audit apply mode for Google BigQuery and Oracle targets with SAP sources

For application ingestion incremental load and combined initial and incremental load jobs with Google BigQuery and Oracle targets, you can configure Audit apply mode, instead of using the Standard apply mode, for tasks to write a row for each DML operation on a source table to the generated target table. You can optionally add columns that contain metadata about the changes to the target table.

This feature is useful when you need an audit trail of changes to perform downstream processing on the data before writing it to the target database or when you need to examine the metadata for the changes. The target tables with the audit information can't have constraints other than indexes.

Ability to select Salesforce source fields for data replication

If you use rules to select Salesforce source objects when defining an application ingestion task, you can individually reselect or clear the fields in each of the selected objects from which to replicate data. Previously, all of the fields were selected and could not be cleared. This feature allows you to replicate only the data you need, thereby reducing the amount of data to be replicated and the replication cost and overhead.

Mass Ingestion Applications provides two new resync options, Resync (refresh) and Resync (retain), that you can use, instead of the existing Resync command, to resychronize the target with the Salesforce source. The options either refresh the target to match the current structure of the source or retain the existing source and table structure that has been used for CDC.

Mass Ingestion Databases

The April 2024 release of Database Ingestion and Replication includes the following new features and enhancements:

Query-based CDC support for database ingestion jobs with Db2 for LUW sources

You can now use the query-based CDC method for database ingestion incremental load and combined initial and incremental load tasks that have Db2 for LUW sources and Snowflake targets. This method captures inserts and updates from the source tables by querying a timestamp column that is updated when a change occurs. When you define an incremental load or combined load task on the **Source** page in the task wizard, the **CDC Method** field is automatically set to **Query-based**. You must enter the query column name and set the column type to Timestamp. The **Include LOBs** option is not supported.

Previously, for jobs with Db2 for LUW sources, only the initial load type was supported and the query-based CDC method was not available.

"Add Last Replicated Time" metadata column records the timestamp of the last DML operation applied to a Microsoft Azure Synapse Analytics or SQL Server target table

For application ingestion and database ingestion jobs that have a Microsoft Azure Synapse Analytics or SQL Server target and use any load type, you can add a metadata column to the target tables that records the date

and time at which the last DML operation was applied to the target table. To add the column, select the **Add Last Replicated Time** check box in the **Advanced** section on the **Target** page of the task wizard. You can optionally add a prefix to the name of the metadata column to easily identify it and to prevent conflicts with the names of existing columns.

Mass Ingestion Streaming

The April 2024 release of Mass Ingestion Streaming includes the following new feature and enhancement:

Cloud-hosted Bitbucket repositories

You can use the cloud-hosted Atlassian Bitbucket repository for source-controlled Mass Ingestion Streaming assets.

Changed behavior

The April 2024 release of Data Ingestion and Replication includes the following changed behaviors.

Data type mapping for an SAP source and an Oracle target

When you configure an application ingestion task with an SAP source and an Oracle target, the CHAR data type of an SAP source is mapped to the VARCHAR2 data type of an Oracle target.

Previously, the CHAR data type of an SAP source was mapped to the CHAR data type of an Oracle target.

Default mapping of DATETIME and TIMESTAMP columns for a MySQL source and a SQL Server target

MySQL source DATETIME and TIMESTAMP columns that were previously mapped to a SQL Server DATETIME target column are now mapped to the SQL Server DATETIME2 column. This change was made to allow the whole range of source data to be correctly replicated to the target.

If you create and deploy a new task for an existing target table that contains a SQL Server DATETIME column, the table will be dropped and recreated. The table will not be dropped if an existing job is resumed or redeployed.

To avoid dropping an existing table, change the SQL Server DATETIME column to DATETIME2 column on the target side using the ALTER TABLE command.

Mass Ingestion sends a product tag to Databricks Delta when connecting to the target

When Mass Ingestion connects to a Databricks Delta target during an application ingestion or database ingestion job, it sends a special INFA_CMI_DATABRICKS tag as a key=value pair appended to the JDBC URL. Databricks Delta can use the tag to detect Informatica queries for data loading and to analyze Informatica consumption. On the Databricks Delta dashboard, the tag appears in the **Query Source** field.

CHAPTER 4

February 2024

This section provides information about new features, enhancements, and behavior changes in the February 2024 release of Informatica Intelligent Cloud Services℠ Data Ingestion and Replication service.

New features and enhancements

Read about new features and enhancements in the February 2024 Mass Ingestion release.

Important notices

The following notices identify preview features and any changes to support levels for existing features.

Previews initiated

Effective in the February 2024 release, you can preview the following new functionality in advance of its general release:

- Mass Ingestion Files now supports Microsoft Fabric OneLake connector as a source and target to transfer files.

Note: Preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Common

The February 2024 release of Informatica Intelligent Cloud Services Data Ingestion and Replication service includes the following new features that are common to application ingestion and database ingestion tasks.

"Add Last Replicated Time" metadata column records timestamp of last DML operation applied to a Google BigQuery or Snowflake target table

For application ingestion and database ingestion jobs that have a Google BigQuery or Snowflake target and use any load type and any apply mode, you can add a metadata column to the target tables that records the date and time at which the last DML operation was applied to the target table. To add the column, select the **Add Last Replicated Time** check box in the **Advanced** section on the **Target** page of the task wizard. You can optionally add a prefix to the name of the metadata column to easily identify it and to prevent conflicts with the names of existing columns.

Support for generating Databricks Delta unmanaged tables on the target

When you create an application ingestion or database ingestion task, you can optionally select the **Create Unmanaged Tables** check box on the **Target** page of the task wizard to generate Databricks Delta target tables as unmanaged tabes instead of managed tables. If you do so, you must also specify a parent directory that exists in Amazon S3 or Microsoft Azure Data Lake Storage to hold the Parquet files that are generated for each target table when captured DML records are processed.

Ability to control the case of letters in the names of generated target objects on Amazon Redshift, Google BigQuery, and Snowflake

When you create an application ingestion or database ingestion task that has an Amazon Redshift, Google BigQuery, or Snowflake target, you can set options on the **Target** page of the task wizard to control the case of letters in the names of generated target tables (or objects) and columns (or fields). Previously, the target names were always generated using the same case as the source names unless overridden by a cluster-level or session-level property on the target. Now, if you select **Enable Case Transformation**, you can select a **Case Transformation Strategy** option to use all uppercase, all lowercase, or the case of the source object names.

Note: For Snowflake targets, the **Enable Case Transformation** check box is unavailable if you select the **Superpipe** option.

Mass Ingestion Applications

The February 2024 release of Mass Ingestion Applications includes the following new features and enhancements:

Support for new targets in initial load jobs with Oracle Fusion sources

Mass Ingestion Applications supports initial load jobs for the following new targets with Oracle Fusion sources using the BICC replication approach:

- Amazon Redshift
- Amazon S3
- Databricks Delta
- Google BigQuery
- Google Cloud Storage
- Microsoft Azure Data Lake Storage Gen2
- Microsoft Azure Synapse Analytics

Child object support for Oracle Fusion sources in incremental load and combined load jobs

When you create an incremental load or combined initial and incremental load job with an Oracle Fusion source, you can choose to include the child object data. This feature applies only if you use the REST replication approach and Google Big Query targets.

Audit apply mode for Snowflake targets

For application ingestion incremental load and combined initial and incremental load jobs with Snowflake targets, you can configure Audit apply mode, instead of using the Standard apply mode, for tasks to write a row for each DML operation on a source table to the generated target table. You can optionally add columns that contain metadata about the changes to the target table, including Add Last Replicated Time, Add Operation Type (selected by default), Add Operation Time, Add Operation Sequence, Add Before Images, Prefix for Metadata Columns (INFA_ by default), Superpipe (selected by default), Merge Frequency, Enable Case Transformation (selected by default), and Case Transformation Strategy.

This feature is useful when you need an audit trail of changes to perform downstream processing on the data before writing it to the target database or when you need to examine the metadata for the changes. The target tables with the audit information cannot have constraints other than indexes.

To enable the use of audit tables, select **Audit** in the **Apply Mode** field on the **Target** page when defining a task. This field is available for new or undeployed tasks. Under **Advanced**, optionally select the check boxes for the metadata columns.

Mass Ingestion Databases

The February 2024 release of Database Ingestion and Replication includes the following new features and enhancements:

Amazon Aurora PostgreSQL targets with Db2 for i sources in initial load and combined load jobs

You can use Amazon Aurora PostgreSQL targets in database ingestion initial load and combined initial and incremental load jobs that have a Db2 for i source. To connect to the Aurora PostgreSQL target, use the PostgreSQL ODBC driver and the PostgreSQL connector.

Previously, you could use Amazon Aurora PostgreSQL as a target only in incremental load jobs that had a Db2 for i source.

Changed behavior

The February 2024 release of Data Ingestion and Replication includes the following changed behaviors.

Improved performance of application ingestion jobs with Salesforce sources

For application ingestion tasks that have a Salesforce source, you can use the **Chunk Size** field in the **Source** page to specify the number of records to be run in parallel. Chunk size can improve memory utilization, reduce the need for a larger JVM requirement, and run more parallel partitions.

Previously, the partitions were entered in the **Number of Partitions** field. This field is no longer available.

Soft Deletes apply mode for application ingestion Snowflake targets

When you configure an application ingestion task with any source and a Snowflake target with **Soft Deletes** apply mode, any update in the source table during normal or backlog mode now results in the deletion of the matching record, insertion of the updated record, and marking of the INFA_OPERATION_TYPE operation as NULL in the target table. Similarly, inserting a record in the source table during backlog mode results in deletion of the matching record, insertion of the updated record, and marking of the INFA_OPERATION_TYPE operation as NULL in the target table.

Previously, updating or inserting records in the source table resulted in updating the existing record and marking of the INFA_OPERATION_TYPE operation as D and NULL in the new record in the target table.

Schema change handling for application ingestion jobs with SAP S4 HANA sources

When you make a schema change to a data object in an application ingestion incremental load job that have an SAP S4 HANA source, the SAP Mass Ingestion connector now generates an alert message. When the alert appears, the processing of the data object stops, and you must create a new task to perform an initial load and then an incremental load with the changed data object to apply the schema changes, depending on the job status.

Previously, no alert messages were generated, and the schema changes were ignored.

Initial Start Point for Incremental Load field no longer available for database ingestion combined load tasks

When you define a combined initial and incremental load database ingestion task, the **Initial Start Point for Incremental Load** is no longer available under the Advanced properties on the **Source** page of the task wizard. This field is not displayed for the combined initial and incremental load jobs because for this load type, initial loading is not performed until the incremental processing of change data reaches the end of the current transaction log. Consequently, the user input for the **Initial Start Point for Incremental Load** field is irrelevant.

You can now customize the **Initial Start Point for Incremental Load** field for incremental load jobs only.

More consistent handling of DML changes in Soft Deletes apply mode for Databricks Delta targets

If you enable the **Soft Deletes** apply mode for application ingestion and database ingestion tasks that have Databricks Delta targets, any source updates or inserts are now applied to the target with NULL shown in the INFA_OPERATION_TYPE column and in any other "Add Operation..." metadata columns you selected on the **Target** page. For deletes, "D" is shown in the INFA_OPERATION_TYPE column and non-null values are written to the other metadata columns you selected. The soft deletes behavior for Databricks Delta is now consistent with the behavior for Snowflake.

Previously, for the unload phase of a combined job, all DML changes caused the INFA_OPERATION_TYPE column and all selected metadata columns to be populated with values. For the incremental CDC phase, inserts and updates caused all metadata fields except INFA_OPERATION_TYPE to be populated, and deletes caused all metadata fields including INFA_OPERATION_TYPE to be populated.

Windows Authentication v2 support for SQL Server connections

If you want to use Windows authentication to connect to SQL Server, select **Windows Authentication v2** in the **Authentication Mode** property when you define SQL Server connection properties. If you previously selected **Windows Authentication (Deprecated)** for SQL Server connections, switch to **Windows Authentication v2** to avoid potential errors.

INDEX

B

behavior changes
- Data Ingestion and Replication July 2024 8
- Mass Ingestion April 2024 16
- Mass Ingestion February 2024 19
- Mass Ingestion May 2024 13

D

Data Ingestion and Replication
- changed behavior in July 2024 8
- new features in July 2024 5

M

Mass Ingestion
- changed behavior in April 2024 16
- changed behavior in February 2024 19
- changed behavior in May 2024 13
- new features in April 2024 14
- new features in February 2024 17
- new features in May 2024 10

N

new features
- Data Ingestion and Replication July 2024 5
- Mass Ingestion April 2024 14
- Mass Ingestion February 2024 17
- Mass Ingestion May 2024 10