



Informatica® Cloud Mass Ingestion January 2023

What's New

© Copyright Informatica LLC 2019, 2023

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: 2023-01-16

Table of Contents

Preface	4
Chapter 1: January 2023.....	5
New features and enhancements.	5
Common.	5
Mass Ingestion Applications.	6
Mass Ingestion Databases.	7
Changed behavior.	8
Fetch Size field available for initial load tasks with Google Analytics sources.	8
Log record sorting in the Secure Agent instead of in Db2 for z/OS.	8
Oracle Key Vault certified for storing master encryption keys for TDE-encrypted Oracle source tables.	8
Chapter 2: November 2022.....	9
New features and enhancements.	9
Mass Ingestion Databases.	9
Mass Ingestion Streaming.	10
Changed behavior.	10
Enable Persistent Storage check box is not displayed for sources where persistent storage is always required.	10
MySQL driver must be downloaded for MySQL sources.	10
Users with Secure Agent Group read permission can view runtime environments in the task wizard.	10
Chapter 3: October 2022.....	11
New features and enhancements.	11
Mass Ingestion Applications	11
Mass Ingestion Databases.	12
Mass Ingestion Streaming.	13
Changed behavior.	13
Capture progress markers for Db2 for i and Microsoft SQL Server sources in database ingestion jobs.	13
Databricks Delta sources and targets in file ingestion tasks.	13
Databricks Delta as targets in streaming ingestion tasks.	14
Index.....	15

Preface

Read *Mass Ingestion What's New* for a brief overview of new features in the Mass Ingestion service, which includes Mass Ingestion Applications, Mass Ingestion Databases, Mass Ingestion Files, and Mass Ingestion Streaming.

CHAPTER 1

January 2023

This section provides information about new features, enhancements, and behavior changes in the January 2023 release of Informatica Intelligent Cloud ServicesSM Mass Ingestion service.

New features and enhancements

The January 2023 release of Informatica Intelligent Cloud Services Mass Ingestion service includes the following new features and enhancements.

Common

The December 2022 release of Informatica Intelligent Cloud Services Mass Ingestion service includes the following new feature that is common to application ingestion and database ingestion tasks.

Undeploy jobs from the Mass Ingestion Command-Line Interface

You can now undeploy an application ingestion or database ingestion job from the Cloud Mass Ingestion Command-Line interface.

Use the following command:

```
job undeploy <job_name>
```

Additional parameters for the Mass Ingestion Command-Line Interface

Additional parameters have been added to the Mass Ingestion Command-Line Interface (CLI) for application ingestion and database ingestion to provide parity with task wizard options:

The Targets section contains the following new parameters under Advanced for Amazon S3, Flat File, Google Cloud Storage, Kafka, Microsoft Azure Data Lake Storage targets:

- addOperationType
- addOperationTime
- addOperationOwner
- addOperationTransactionId
- addBeforeImages

The Runtime options section contains the following new parameters under Advanced for Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage Gen2 targets::

- applyCycleChangeLimit
- applyCycleIntervalDays
- applyCycleIntervalHours
- applyCycleIntervalMins
- applyCycleIntervalSecs
- lowActivityFlushHours
- lowActivityFlushMins

And for Apache Kafka targets:

- checkpointAllRows
- checkpointEveryCommit
- checkpointRowCount

Ability to view source control logs in Monitor

The ability to view source control logs on the **Source Control Logs** page in Monitor has been reinstated.

If your organization is enabled for source control, you can monitor the actions users in your organization perform on source controlled objects such as projects, folders, and ingestion tasks of any type. Each time a user performs a pull, check in, check out, or other action on an object that is source controlled, Informatica Intelligent Cloud Services logs the action on the Source Control Logs page.

Mass Ingestion Applications

The January 2023 release of Mass Ingestion Applications includes the following new features and enhancements:

Amazon Redshift targets supported for Workday sources

Application ingestion jobs that use the initial, incremental, or combined initial and incremental load type can now replicate data from Workday sources to Amazon Redshift targets.

To connect to an Amazon Redshift target, use the Amazon Redshift V2 connector.

Databricks Delta targets supported for additional sources

Application ingestion initial load jobs can now replicate data to Databricks Delta targets from the following sources:

- Marketo
- Salesforce Marketing Cloud
- Workday

To connect to a Databricks Delta target, use the Databricks Delta connector.

Google BigQuery targets supported for Adobe Analytics sources

Application ingestion jobs that use the initial, incremental, or combined initial and incremental load type can now replicate data from Adobe Analytics sources to Google BigQuery targets.

To connect to a Google BigQuery target, use the Google BigQuery V2 connector.

Initial loads of Workday RaaS custom reports to Oracle and Snowflake

Application initial load jobs that have a Workday source can now use the Workday Report-as-a-Service (RaaS) API to extract source data from custom fields and objects into custom reports and then load the reports to Oracle or Snowflake targets. You can use a single custom report or multiple custom reports.

As an alternative, the SOAP API is still supported for reading source data.

Note: The **Fetch Size** field and **Extract Non-default Fields** field on the **Source** page of the task wizard are available only if you select the SOAP API.

Mass Ingestion Databases

The December 2022 release of Mass Ingestion Databases includes the following new features and enhancements:

Microsoft SQL Server targets in initial load jobs

You can use SQL Server targets in initial load jobs that have an Oracle or SQL Server source. Use version 2017 or 2019 of SQL Server Developer Edition, Enterprise Edition, or Standard Edition for the target. To connect to the target, use the SQL Server connector.

Note: For a SQL Server target, only the required connection properties marked with a red asterisk (*) are used.

Oracle source database partitioning to improve the performance of initial load jobs

For database ingestion initial load jobs that have an Oracle source, you can now enable partitioning of source tables and distribute the records read from the source across the partitions for parallel processing. Mass Ingestion Databases determines the range of partitions by using the ROWID as the partition key. Use this feature to improve the performance of jobs with large source tables. You can enable partitioning and configure the number of partitions to use from the **Source** page of the task wizard.

Query-based CDC technique for incremental load jobs with SQL Server sources and Snowflake targets

For incremental load jobs that have a SQL Server source and a Snowflake target, Mass Ingestion Databases adds Query-based CDC as a new technique to capture change data. Previously, only log-based CDC was available, which required some database configuration tasks.

You can enable query-based CDC on the **Source** page of the task wizard. When query-based change data capture is enabled, Mass Ingestion Databases uses a SQL WHERE clause that references a query column in the source tables to identify the rows with Insert and Update changes made during the CDC interval. Delete operations are not included.

Incremental load jobs that have other target types continue to use the log-based CDC technique by default.

JTOpen JDBC driver for Db2 for i SSL encryption

Mass Ingestion Databases can now use the JTOpen JDBC driver to support SSL data encryption for Db2 for i sources.

By default, Mass Ingestion Databases uses the DataDirect JDBC driver to connect to the Db2 for i database.

To use SSL data encryption, in the Db2 for i Database Ingestion connection properties, select **JTOpen** in the **JDBC Driver** field and select **SSL** in the **Encryption Method** field. Also add the required certificates to the Informatica Cloud Secure Agent JRE cacerts keystore.

Changed behavior

The January 2023 release of Mass Ingestion includes the following changed behaviors.

Fetch Size field available for initial load tasks with Google Analytics sources

The **Fetch Size** field is now available on the **Source** page of the task wizard for application ingestion initial load tasks that have Google Analytics sources. The **Fetch Size** field specifies the number of records that an application ingestion job associated with the task reads at a time from the source.

Previously, the **Fetch Size** field was available for incremental load and combined initial and incremental load jobs, but not for initial load jobs.

Log record sorting in the Secure Agent instead of in Db2 for z/OS

When slow log collection causes insufficient Db2 DSNDB07 workspace or sort pool allocation, you can set the Mass Ingestion Databases `pxw.cdcreader.ZOS.stored.procedure.sort` custom property to true to enable the sorting of log records in the Secure Agent instead of in Db2 for z/OS.

Oracle Key Vault certified for storing master encryption keys for TDE-encrypted Oracle source tables

If you use Oracle Transparent Data Encryption (TDE) to encrypt data in tablespaces that contain Oracle source tables in incremental load jobs, you can store the master encryption key in Oracle Key Value (OKV). This hardware security module has now been certified. You must specify the following custom property on the **Source** page of the task wizard:

```
pxw.cdcreader.oracle.database.additional TDEKEYSTORE=(TYPE=(HSM,HSMCLIENTLIB=/  
ENDPOINTDIR/lib/liborapks.dll,PWD="null")
```

For more information about this custom property, contact Informatica Global Custom Support.

CHAPTER 2

November 2022

This section provides information about new features, enhancements, and behavior changes in the November 2022 release of Informatica Intelligent Cloud ServicesSM Mass Ingestion service.

New features and enhancements

The November 2022 release of Informatica Intelligent Cloud Services Mass Ingestion service includes the following new features and enhancements.

Mass Ingestion Databases

The November 2022 release of Mass Ingestion Databases includes the following new features and enhancements:

Support for Amazon Aurora MySQL sources in incremental load jobs

Mass Ingestion Databases adds support for Amazon Aurora MySQL 2.x and 3.x versions for sources in incremental load jobs. Previously these versions were supported as sources in initial load jobs only.

Support for Azure Database for PostgreSQL - Flexible Server sources

Mass Ingestion Databases adds support for Azure Databases for PostgreSQL - Flexible Server sources in initial load and incremental load jobs. Supported versions are 12, 13, and 14.

Support for MySQL geometric data types

Mass Ingestion Databases now supports the following MySQL geometric data types:

- GEOMETRY
- GEOMETRYCOLLECTION
- LINESTRING
- MULTILINESTRING
- MULTIPOINT
- MULTIPOLYGON
- POINT
- POLYGON

Mass Ingestion Streaming

The November 2022 release of Mass Ingestion Streaming includes the following new features and enhancements:

Google BigQuery V2 target in streaming ingestion tasks

You can use Google BigQuery V2 as a target in a streaming ingestion task.

Cross account authentication for Amazon Kinesis

You can configure cross account authentication for the type Kinesis service for additional security.

Changed behavior

The November 2022 release of Mass Ingestion includes the following changed behaviors.

Enable Persistent Storage check box is not displayed for sources where persistent storage is always required

The **Enable Persistent Storage** check box is not displayed on the **Source** page in the database ingestion task wizard for all sources for which persistent storage is required and cannot be disabled. These sources include MongoDB, PostgreSQL, SAP HANA, and SAP HANA Cloud. Previously, the check box was displayed in an unavailable state for SAP HANA sources.

MySQL driver must be downloaded for MySQL sources

The MySQL driver is no longer provided with the MySQL connector.

You must download the MySQL driver file, `mysql-connector-java-<version>.jar`, if you use any of the following load type and source edition combinations:

- - Incremental load jobs that have MySQL Community Edition or MySQL Enterprise Edition sources
 - Initial load jobs that have MySQL Community Edition sources
 - Initial load jobs that have Amazon Relational Database Service (RDS) for MySQL sources

Note: You do not need to download the driver if you use MySQL Enterprise Edition sources in initial load jobs.

After you download the driver, copy it to the following directory:

```
<Secure_Agent_installation_directory>/ext/connectors/thirdparty/com.mysql/
```

Users with Secure Agent Group read permission can view runtime environments in the task wizard

When creating a Mass Ingestion ingestion task of any type, users with Secure Agent Group read permission can now view runtime environments in the **Runtime Environments** list.

Previously, users were required to have Secure Agent read permission to view them.

CHAPTER 3

October 2022

This section provides information about new features, enhancements, and behavior changes in the October 2022 release of Informatica Intelligent Cloud Services™ Mass Ingestion service.

You can access the What's New video on [YouTube](#).

New features and enhancements

The October 2022 release of Informatica Intelligent Cloud Services Mass Ingestion service includes the following new features and enhancements.

Mass Ingestion Applications

The October 2022 release of Mass Ingestion Applications includes the following new features and enhancements:

Ability to resynchronize source and target objects in jobs with Salesforce Marketing Cloud and Zendesk sources

You can now resynchronize the source and target objects for a subtask that is part of a running initial load job configured for a Salesforce Marketing Cloud source. Similarly, you can also resynchronize the source and target objects for a subtask that is part of a running combined initial and incremental load job configured for a Zendesk source.

Combined initial and incremental load jobs for cloud data lake targets

You can now configure combined initial and incremental load jobs to replicate data from NetSuite, SAP ECC, SAP S4/HANA, and ServiceNow sources to the following cloud data lake targets:

- Amazon S3 V2
- Google Cloud Storage V2
- Microsoft Azure Data Lake Storage Gen2

To connect to the targets, use the following connectors:

- **Amazon S3 V2:** Writes data to Amazon S3 targets.
- **Google Cloud Storage V2:** Writes data to Google Cloud Storage targets.
- **Microsoft Azure Data Lake Storage Gen2:** Writes data to Microsoft Azure Data Lake Storage Gen2 targets.

Oracle targets for additional sources

Application ingestion jobs of all load types can now replicate data to Oracle targets from the following sources:

- NetSuite
- SAP ECC
- SAP S4/HANA
- ServiceNow

To connect to an Oracle target, use the Oracle Database Ingestion connector.

Mass Ingestion Databases

The October 2022 release of Mass Ingestion Databases includes the following new features and enhancements:

Ability to read data from partitions in SAP HANA source tables

To support very large SAP HANA source tables that contain two billion records or more, Mass Ingestion Databases requires the tables to use table partitioning and row distribution. Database ingestion initial load jobs can then read data from the partitions in a partitioned SAP HANA source table.

Advance restart tokens written to the target for more efficient restart processing

For database ingestion incremental load or combined initial and incremental load jobs that have a Db2 for i, Db2 for z/OS, Microsoft SQL Server, or Oracle source, Mass Ingestion Databases now persists an advance restart token to the target during periods of no or low change activity. When a restart of a job is requested, the Writer can compare the last write to target with the log position in the advance restart marker that is saved to a persisted cache on the target. The Writer uses the later of the two positions. This feature prevents jobs from restarting from the beginning of the logs and re-reading previously captured changes or from reprocessing logs that have no changes of CDC interest.

You can use the source readerMinutesRestartAdvanceMarker custom property to set the frequency at which the restart markers are generated or to disable the generation of the restart markers. The default property value is 60 minutes. A value of 0 disables the generation of the restart markers.

MySQL sources in incremental load jobs with Snowflake targets

Mass Ingestion Databases incremental load jobs can now replicate change data from MySQL sources to a Snowflake target.

Mass Ingestion Databases reads change events from the MySQL binary log (binlog) files either by using a unique Global Transaction ID if you enable GTID mode in MySQL or by fetching the binlog file name and position.

To connect to a MySQL source, use the MySQL connector.

Additional target types supported for database ingestion jobs with MongoDB sources

Database ingestion initial load and incremental load jobs that have MongoDB 4.x or 5.x sources can now replicate data to Amazon Redshift, Microsoft Synapse Analytics, and Snowflake targets. Previously, jobs with

MongoDB sources replicated data only to Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage Gen2 targets.

SAP HANA Cloud sources in database ingestion initial and incremental load jobs

SAP HANA Cloud databases are now supported as sources in database ingestion initial load and incremental jobs.

Use the SAP HANA Database Ingestion connector to connect to the Cloud source. In the connection properties, enter the following required properties in the **Advanced Connection Properties** field:

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

Note: Some SAP HANA on-premises data types are not supported for SAP HANA Cloud sources, such as ALPHANUM, CHAR, BINTTEXT, and CLOB.

Mass Ingestion Streaming

The October 2022 release of Mass Ingestion Streaming includes the following new features and enhancements:

REST APIs

You can copy and update a streaming ingestion task using REST APIs.

Changed behavior

The October 2022 release of Mass Ingestion includes the following changed behaviors.

Capture progress markers for Db2 for i and Microsoft SQL Server sources in database ingestion jobs

By default, Mass Ingestion Databases now generates capture progress makers that indicate the log read position for database ingestion incremental load and combined initial and incremental load jobs that have Db2 for i or Microsoft SQL Server sources. The markers are generated every 60 seconds by default. You can use the following source custom properties to control whether the capture markers are used or ignored during database ingestion, the frequency at which the markers are generated, and whether the messages are included in messages: `readerProcessCaptureProgressMarker`, `readerSecondsCaptureProgressMarkers`, and `readerLogCaptureProgressMarker`.

Databricks Delta sources and targets in file ingestion tasks

This release includes the following changes to Databricks Delta in file ingestion tasks:

Databricks Delta sources

You can't use Databricks Delta as a source in a file ingestion task.

Previously, you could use Databricks Delta as a file ingestion task source.

Databricks Delta targets

When you use Databricks Delta as a target in a file ingestion task, the target table must exist before you create the task.

Previously, the Mass Ingestion Files created the target table if it didn't exist.

Databricks Delta as targets in streaming ingestion tasks

Databricks Delta uses SQL endpoints, not clusters, to write data to tables. Therefore, you don't need to configure the cluster when you use Databricks Delta as a target in streaming ingestion tasks.

Previously, Databricks Delta used clusters to write data to tables, and you had to configure the cluster when you used Databricks Delta as a target in streaming ingestion tasks.

INDEX

B

behavior changes

Mass Ingestion January 2023 [8](#)

Mass Ingestion November 2022 [10](#)

Mass Ingestion October 2022 [13](#)

M

Mass Ingestion

changed behavior in January 2023 [8](#)

changed behavior in November 2022 [10](#)

changed behavior in October 2022 [13](#)