

JDBC to S3/ADLS

Date published: 2021-04-06

Date modified: 2024-01-09



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

ReadyFlow overview: JDBC to JDBC.....	4
Prerequisites.....	4
List of required configuration parameters for the JDBC to JDBC ReadyFlow.....	5

ReadyFlow overview: JDBC to JDBC

You can use the JDBC to JDBC ReadyFlow to move data between a source database table and destination database table.

This ReadyFlow consumes data from a source database table. You can filter records by specifying a SQL query in the Filter Rule parameter. The filtered records are then inserted into the target database table. Failed database operations are retried automatically to handle transient issues. Define a KPI on the failure_PutDatabaseRecord connection to monitor failed insert operations. Define a KPI on the retry_PutDatabaseRecord connection to monitor retryable insert operations.

JDBC to JDBC ReadyFlow details	
Source	Database Table
Source Format	Database Table
Destination	Database Table
Destination Format	Database Table

Prerequisites

Learn how to collect the information you need to deploy the JDBC to JDBC ReadyFlow, and meet other prerequisites.

For your data ingest source and target

- You have obtained the source and target table names
- You have identified the source table column which should be used to incrementally fetch records
- You have obtained the JDBC source and target database connection URLs
- You have obtained the appropriate JDBC driver and JDBC driver class names for your source and target database types
- You have obtained the usernames and passwords to access the source and target tables

For DataFlow

- You have enabled DataFlow for an environment.

For information on how to enable DataFlow for an environment, see [Enabling DataFlow for an Environment](#).

- You have created a Machine User to use as the CDP Workload User.
- You have given the CDP Workload User the EnvironmentUser role.

1. From the Management Console, go to the environment for which DataFlow is enabled.
2. From the Actions drop down, click Manage Access.
3. Identify the user you want to use as a Workload User.




Note:


The CDP Workload User can be a machine user or your own user name. It is best practice to create a dedicated Machine user for this.

4. Give that user EnvironmentUser role.
- You have synchronized your user to the CDP Public Cloud environment that you enabled for DataFlow.

For information on how to synchronize your user to FreeIPA, see [Performing User Sync](#).

- You have granted your CDP user the DFCatalogAdmin and DFFlowAdmin roles to enable your user to add the ReadyFlow to the Catalog and deploy the flow definition.
 1. Give a user permission to add the ReadyFlow to the Catalog.
 - a. From the Management Console, click User Management.
 - b. Enter the name of the user or group you wish to authorize in the Search field.
 - c. Select the user or group from the list that displays.
 - d. Click Roles Update Roles .
 - e. From Update Roles, select DFCatalogAdmin and click Update.



Note: If the ReadyFlow is already in the Catalog, then you can give your user just the DFCatalogViewer role.
 2. Give your user or group permission to deploy flow definitions.
 - a. From the Management Console, click Environments to display the Environment List page.
 - b. Select the environment to which you want your user or group to deploy flow definitions.
 - c. Click Actions Manage Access to display the Environment Access page.
 - d. Enter the name of your user or group you wish to authorize in the Search field.
 - e. Select your user or group and click Update Roles.
 - f. Select DFFlowAdmin from the list of roles.
 - g. Click Update Roles.
 3. Give your user or group access to the Project where the ReadyFlow will be deployed.
 - a. Go to DataFlow Projects .
 - b. Select the project where you want to manage access rights and click  More Manage Access .
 4. Start typing the name of the user or group you want to add and select them from the list.
 5. Select the Resource Roles you want to grant.
 6. Click Update Roles.
 7. Click Synchronize Users.

Related Concepts

[List of required configuration parameters for the JDBC to JDBC ReadyFlow](#)

List of required configuration parameters for the JDBC to JDBC ReadyFlow

When deploying the JDBC to JDBC ReadyFlow, you have to provide the following parameters. Use the information you collected in *Prerequisites*.

Table 1: JDBC to JDBC ReadyFlow configuration parameters

Parameter name	Description
Destination Database Connection URL	Specify the JDBC database connection URL used to connect to your destination database. May contain database system name, host, port, database name and some parameters. The exact syntax of a database connection URL is specified by your DBMS, for example, jdbc:mysql://hostname:port/database for MySQL databases.
Destination Database Driver	Upload the database driver jar file for your destination database.
Destination Database Driver Class Name	Specify the destination database driver class name. For example, com.mysql.jdbc.Driver for MySQL databases.
Destination Database Password	Specify the destination database password.

Parameter name	Description
Destination Database Table Name	Specify the destination database table name.
Destination Database User	Specify the destination database user.
Filter Rule	Specify the filter rule expressed in SQL to filter database records. Records matching the filter are written to the destination. The default value forwards all records.
Source Database Connection URL	Specify the JDBC database connection URL used to connect to your source database. May contain database system name, host, port, database name and some parameters. The exact syntax of a database connection URL is specified by your DBMS, for example, jdbc:mysql://hostname:port/database for MySQL databases.
Source Database Driver	Upload the database driver jar file for your source database.
Source Database Driver Class Name	Specify the source database driver class name. For example, com.mysql.jdbc.Driver for MySQL databases.
Source Database Incremental Fetch Column	Specify the column that should be used to track state as new data lands in the source table.
Source Database Password	Specify the source database password.
Source Database Table Name	Specify the source database table name.
Source Database User	Specify the source database user.

Related Concepts

[Prerequisites](#)

Related Information

[Deploying a ReadyFlow](#)