Cloudera Streaming Analytics 1.5.0

Data Providers

Date published: 2019-12-17
Date modified: 2021-09-29



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. 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

Registering Data Providers in SSB.....4
Adding Kafka as Data Provider.....4
Adding Schema Registry as Catalog.....6
Adding Kudu as Catalog.....7
Adding Hive as Catalog.....8
Adding Custom Catalogs.....8

Managing registered Data Providers.....9

Registering Data Providers in SSB

Data Providers are a set of data endpoints to be used as sources, sinks and catalogs. Data Providers allow you to connect to an already installed component on your cluster, then use that provider for adding tables in SQL Stream Builder (SSB).

You can access the **Data Providers** page through the Streaming SQL Console:

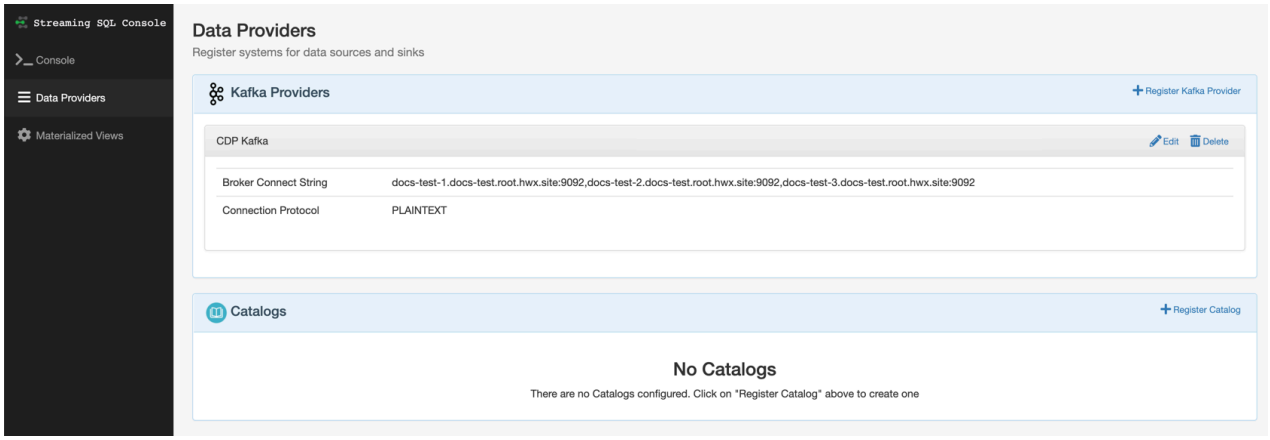
1. Go to your cluster in Cloudera Manager.
2. Select SQL Stream Builder from the list of services.
3. Click SQLStreamBuilder Console.

   The Streaming SQL Console opens in a new window.
4. Click Data Providers on the main menu.

   You are redirected to the Data Providers page.

You can register Kafka as a data provider, or Kudu, Hive and Schema Registry as a catalog. When registering the components, SSB can access the already existing topics from Kafka, tables from Kudu and Hive, and the schema in Schema Registry. This also means that when you update a data provider, for example add new topics, tables and schemas, SSB automatically detects the changes.

You can also manage your data providers after registering them. You can Edit the providers if there is any change in the connection. You can also Delete them when you no longer need the specific provider.



Adding Kafka as Data Provider

You need to register Kafka as a Data Provider using the Streaming SQL Console to create Kafka tables in SQL Stream Builder (SSB).

Before you begin

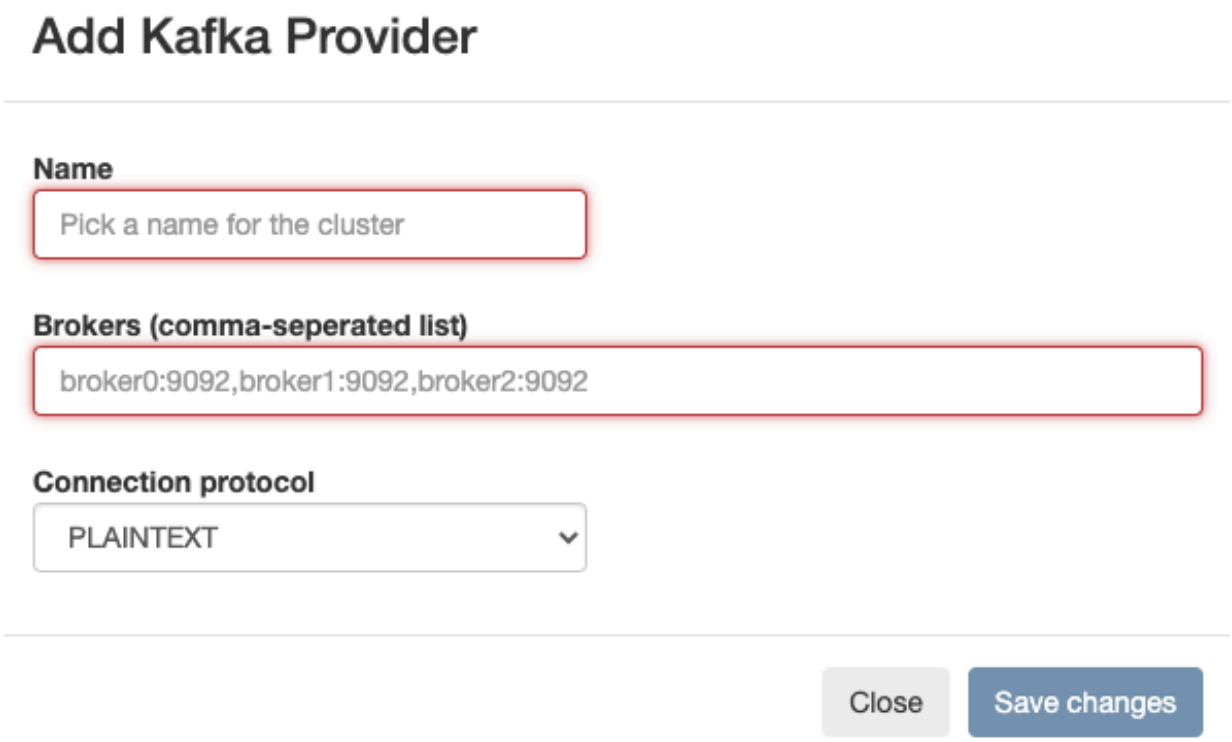
- Make sure that you have Kafka service on your cluster.
- Make sure that you have the right permissions set in Ranger.

Procedure

1. Navigate to the Streaming SQL Console.
   a) Go to your cluster in Cloudera Manager.
   b) Select SQL Stream Builder from the list of services.
   c) Click SQLStreamBuilder Console.

   The **Streaming SQL Console** opens in a new window.
2. Click Data Providers from the main menu.
3. Click Register Kafka Provider.

   The Add Kafka Provider window appears.



4. Add a Name to your Kafka provider.
5. Add the broker host name(s) to Brokers.

   You need to copy the Kafka broker name(s) from Cloudera Manager.
   a) Go to your cluster in Cloudera Manager.
   b) Click Kafka from the list of services.
   c) Click Instances.
   d) Copy the hostname of the Kafka broker(s) you want to use.
   e) Go back to the Add Kafka Provider page.
   f) Paste the broker hostname to the Brokers field.

   Note: You can add more than one broker hostname by separating them by commas.

   g) Add the default Kafka port after the hostname(s).

   Example:

```
docs-test-1.vpc.cloudera.com:9092,
docs-test-2.vpc.cloudera.com:9092
```

6. Select the Connection Protocol.

   The connection protocol must be the same as it is configured for the Kafka cluster in Cloudera Manager.

   You can choose from the following protocols:

   a) Select Plaintext, and click Save Changes.
   b) Select SSL, and click Save Changes.
   c) Select SASL/SSL, and choose an SASL Mechanism.

      1. Select Kerberos, and provide the Kafka Truststore location. Click Save Changes.
      2. Select Plain, and provide the SASL username and password. Click Save Changes.

Results

You have registered Kafka as a data provider to be able to add Kafka as a table in your SQL query. The already existing Kafka topics can be selected when adding Kafka as a table.

Adding Schema Registry as Catalog

You need to add Schema Registry as a Catalog using the Streaming SQL Console in SQL Stream Builder (SSB) to use Schema Registry with Flink DDL.

Before you begin

- Make sure that you have Schema Registry service on your cluster.
- Make sure that you have the right permissions set in Ranger.

Procedure

1. Navigate to the Streaming SQL Console.
   a) Go to your cluster in Cloudera Manager.
   b) Select SQL Stream Builder from the list of services.
   c) Click SQLStreamBuilder Console.

   The **Streaming SQL Console** opens in a new window.
2. Click Data Providers from the main menu.
3. Click Register Catalog.

   The Add Catalog window appears.
4. Add a Name to your catalog.
5. Select Schema Registry from the Catalog Type drop-down.
6. Select the Kafka cluster you registered as Data Provider.
7. Add the Schema Registry URL.
   a) Go to your cluster in Cloudera Manager.
   b) Select Schema Registry from the list of services.
   c) Click on Instances.
   d) Copy the Hostname of Schema Registry.
   e) Add the default port of Schema Registry after the hostname.

      Example:

      ```
      http://docs-test-1.vpc.cloudera.com:7788/api/v1
      ```
8. Select if you wan to enable TLS for Schema Registry.
   a) If yes, provide the Schema Registry Truststore location and password.

9. Click on Add Filter.
   a) Provide a Database and Table filter if you want to select specific tables to use from the catalog.
10. Click on Validate.
11. If the validation is successful, click Add Tables.

Results

Schema Registry is added as a Catalog and ready to be used in Flink DDL. The already existing schemas in Schema Registry are automatically imported to SSB.

Adding Kudu as Catalog

You need to add Kudu as a Catalog using the Streaming SQL Console in SQL Stream Builder (SSB) to create Kudu tables with Flink DDL.

Before you begin

- Make sure that you have Kudu service on your cluster.
- Make sure that you have the right permissions set in Ranger.

Procedure

1. Navigate to the Streaming SQL Console.
   a) Go to your cluster in Cloudera Manager.
   b) Select SQL Stream Builder from the list of services.
   c) Click SQLStreamBuilder Console.

   The **Streaming SQL Console** opens in a new window.
2. Click Data Providers from the main menu.
3. Click Register Catalog.

   The Add Catalog window appears.
4. Add a Name to your catalog.
5. Select Kudu from the Catalog Type drop-down.
6. Add the Host URL of Kudu Masters.
   a) Go to your cluster in Cloudera Manager.
   b) Select Kudu from the list of services.
   c) Click on Instances.
   d) Copy the Hostname of the Master Default Group.
   e) Add the default port of Kudu after the hostname.

   Example:

   ```
   http://docs-test-1.vpc.cloudera.com:7051
   ```
7. Click on Add Filter.
   a) Provide a Database and Table filter if you want to select specific tables to use from the catalog.
8. Click on Validate.
9. If the validation is successful, click Add Tables.

Results

Kudu is added as a Catalog and ready to be used in Flink DDL. The already existing tables in Kudu are automatically imported to SSB.

Adding Hive as Catalog

You need to add Hive as a Catalog using the Streaming SQL Console in SQL Stream Builder (SSB) to create Hive tables with Flink DDL.

Before you begin

- Make sure that you have Hive service on your cluster.
- Make sure that you have the right permissions set in Ranger.

Procedure

1. Navigate to the Streaming SQL Console.
   a) Go to your cluster in Cloudera Manager.
   b) Select SQL Stream Builder from the list of services.
   c) Click SQLStreamBuilder Console.

   The **Streaming SQL Console** opens in a new window.
2. Click Data Providers from the main menu.
3. Click Register Catalog.

   The Add Catalog window appears.
4. Add a Name to your catalog.
5. Select Hive from the Catalog Type drop-down.
6. Click on Add Filter.
   a) Provide a Database and Table filter if you want to select specific tables to use from the catalog.
7. Click on Validate.
8. If the validation is successful, click Add Tables.

Results

Hive is added as a Catalog and ready to be used in Flink DDL. The already existing tables in Hive are automatically imported to SSB.

Adding Custom Catalogs

You can use custom catalogs in cases when you need to add a catalog that is not predefined in the Streaming SQL Console, and when more advanced settings need to be specified for the catalogs. This case you need to provide the configuration directly to Flink with the custom catalog option.

Procedure

1. Navigate to the Streaming SQL Console.
   a) Go to your cluster in Cloudera Manager.
   b) Select SQL Stream Builder from the list of services.
   c) Click SQLStreamBuilder Console.

   The **Streaming SQL Console** opens in a new window.
2. Click Data Providers from the main menu.
3. Click Register Catalog.

   The Add Catalog window appears.
4. Add a Name to your catalog.
5. Select Custom from the Catalog Type drop-down.
6. Provide a Property Name and Value for the Catalog.

7. Click on Add Filter.
   a) Provide a Database and Table filter if you want to select specific tables to use from the catalog.
8. Click on Validate.
9. If the validation is successful, click Add Tables.

Results
The custom catalog is added and ready to be used in Flink DDL.

Managing registered Data Providers

You can edit or delete the registered Data Providers if you need to change their configurations or if you no longer need them.

Editing registered Data Providers

1. Click Data Providers from the main menu.
2. Search for the Kafka provider or catalog you want to modify.
3. Click Edit.

   The Edit Provider or Catalog window appears.
4. Change the settings as required.

   Note: You must validate the modified catalog before saving the changes.
5. Click Save Changes.

Deleting registered Data Providers

1. Click Data Providers from the main menu.
2. Search for the Kafka provider or catalog you want to modify.
3. Click Delete.
4. Click Confirm to delete the provider or catalog.