Cloudera Streaming Analytics 1.8.0

Data Providers

Date published: 2019-12-17
Date modified: 2022-09-28

https://docs.cloudera.com/

Legal Notice

Contents

Registering Data Providers in SSB

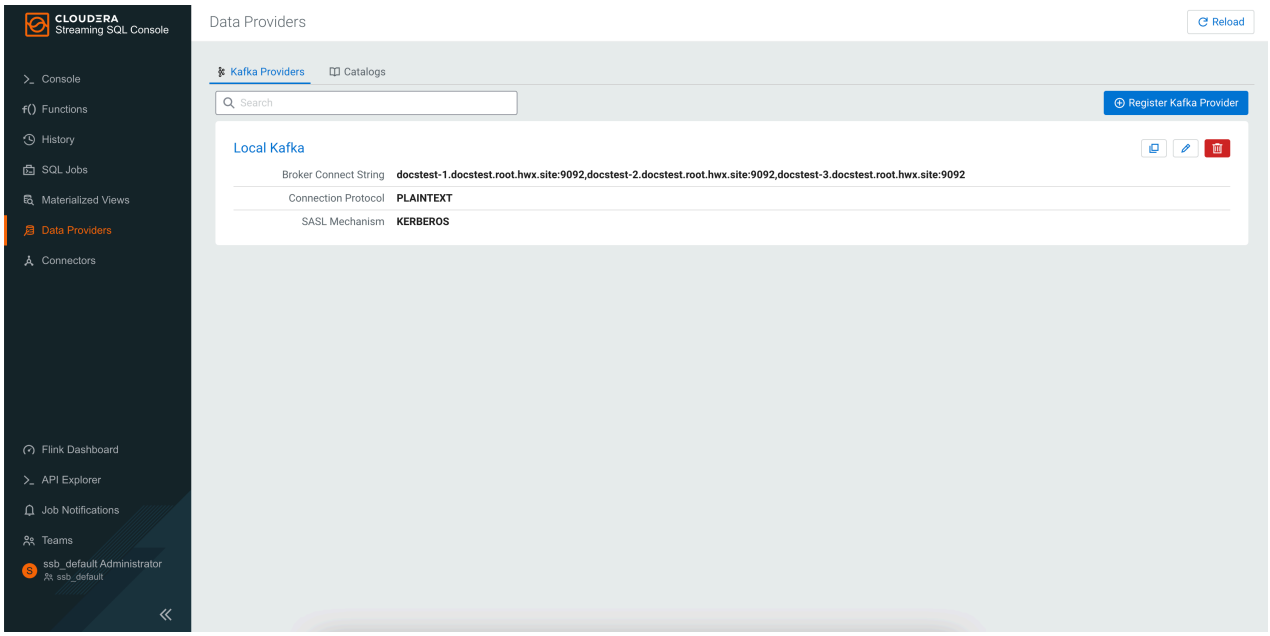
Data Providers are a set of data endpoints to be used as sources, sinks and catalogs. Data Providers allow you to connect to an already installed component on your cluster, then use that provider for adding tables in SQL Stream Builder (SSB).

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.

Adding Kafka Data Provider

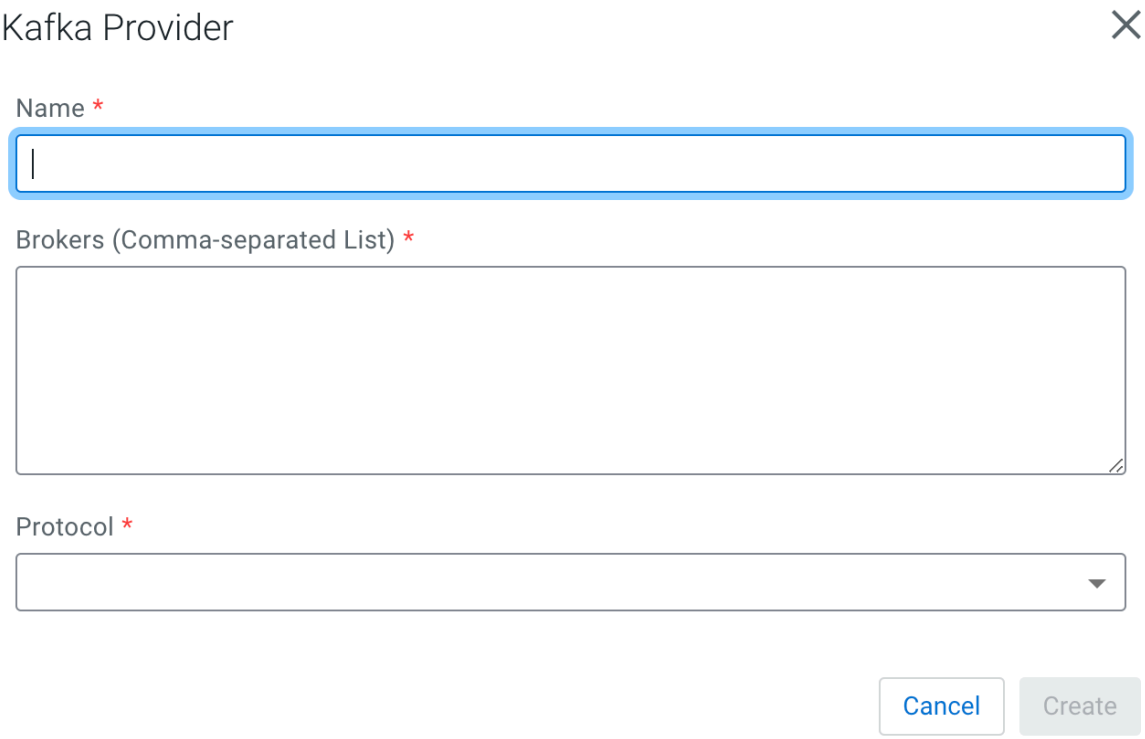
You need to register Kafka as a Data Provider using the Streaming SQL Console to create Kafka tables in SQL Stream Builder (SSB).

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:

For TLS disabled

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

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

For TLS enabled

```
docs-test-1.vpc.cloudera.com:9093,
docs-test-2.vpc.cloudera.com:9093
```

6. Select the security 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:

   For PLAINTEXT

   a. Click Create.

   For SSL

   a. Provide the path to the Kafka TrustStore.
   b. Click Create.

   For SASL SSL

   a. Provide the path to the Kafka TrustStore.
   b. Choose an SASL Mechanism.
   c. Provide the Username for SASL.
   d. Provide the Password for SASL.
   e. Click Create.

   For SASL PLAINTEXT

   a. Choose an SASL Mechanism.
   b. Provide the Username for SASL.
   c. Provide the Password for SASL.
   d. Click Create.

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 Catalogs

You need to add Schema Registry, Kudu, Hive or other services as a Catalog using the Streaming SQL Console in SQL Stream Builder (SSB) to use them with Flink DDL.

Before you begin

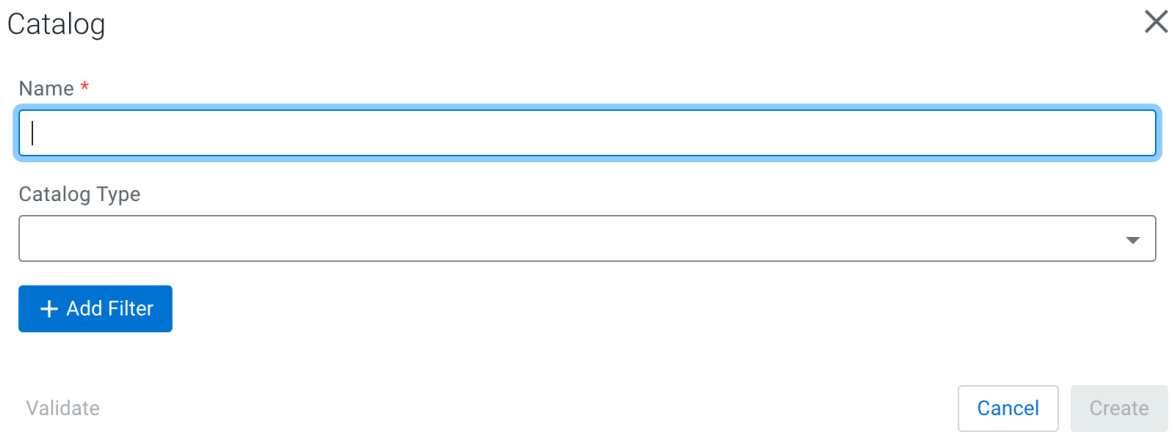
- Make sure that you have required service on your cluster.
- Make sure that you have the right permissions set in Ranger for SSB and the services.

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 Catalogs tab.
4. Click Register Catalog.

   The Add Catalog window appears.



5. Add a Name to your catalog.

6. Select the Catalog Type from the following options:

For Schema Registry

a. Select Schema Registry from the Catalog Type drop-down.

b. Select the Kafka cluster you registered as Data Provider.

c. Enable TLS, if needed for the communication.

   1. If you enabled TLS, provide the Schema Registry Truststore location and password to the SR TrustStore and SR TrustStore Password field.

d. Add the Schema Registry URL.

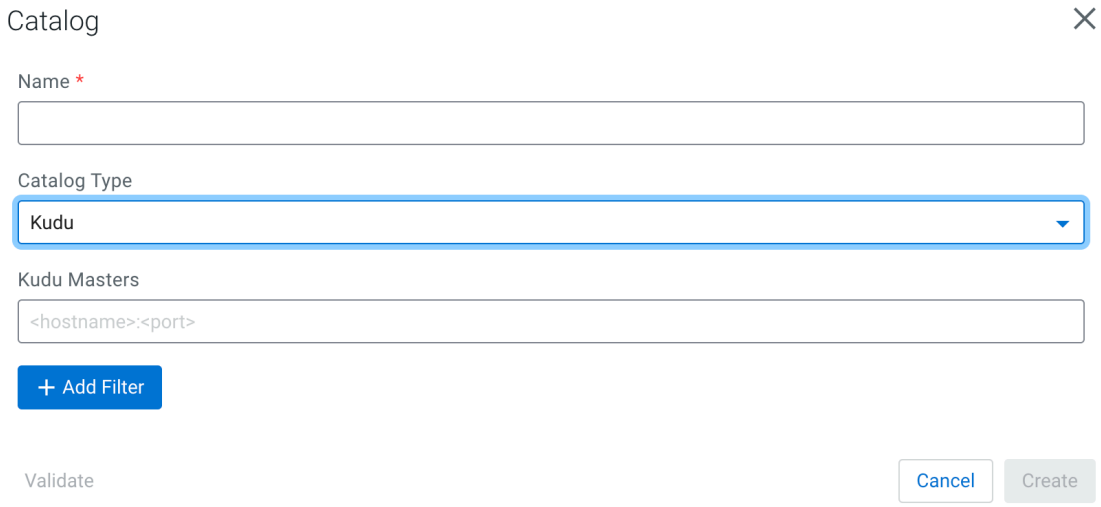
   1. Go to your cluster in Cloudera Manager.
   2. Select Schema Registry from the list of services.
   3. Click on Instances.
   4. Copy the Hostname of Schema Registry.
   5. Add the default port of Schema Registry after the hostname.

      Example:

      ```
      http://docs-test-1.vpc.cloudera.com:7788/api/v1
      ```

For Kudu

a. Select Kudu from the Catalog Type drop-down.



b. Add the host URL of Kudu Masters.

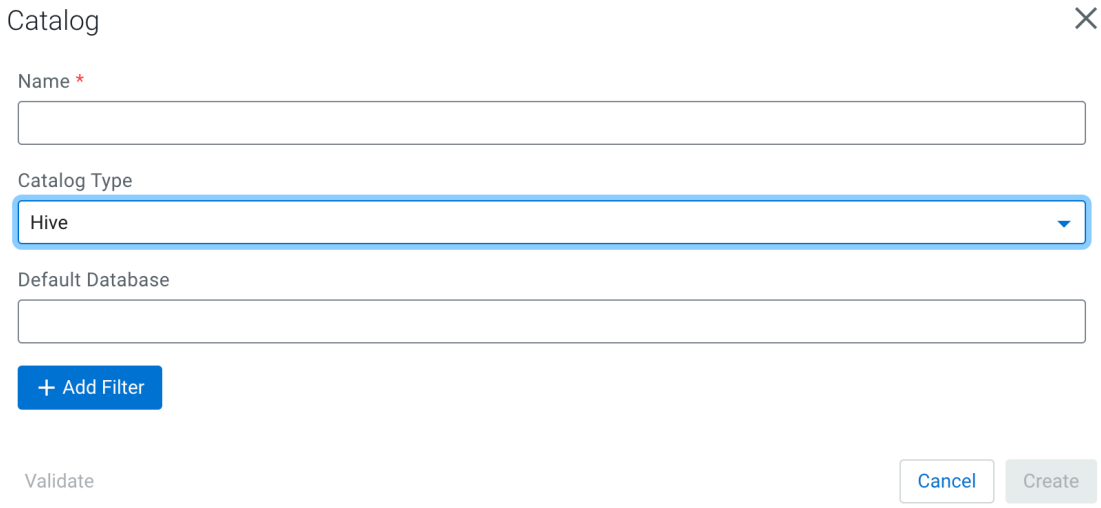
1. Go to your cluster in Cloudera Manager.
2. Select Kudu from the list of services.
3. Click on Instances.
4. Copy the Hostname of the Master Default Group.
5. Add the default port of Kudu after the hostname.

   Example:

```
docs-test-1.vpc.cloudera.com:7051
```

For Hive

a. Select Hive from the Catalog Type drop-down.

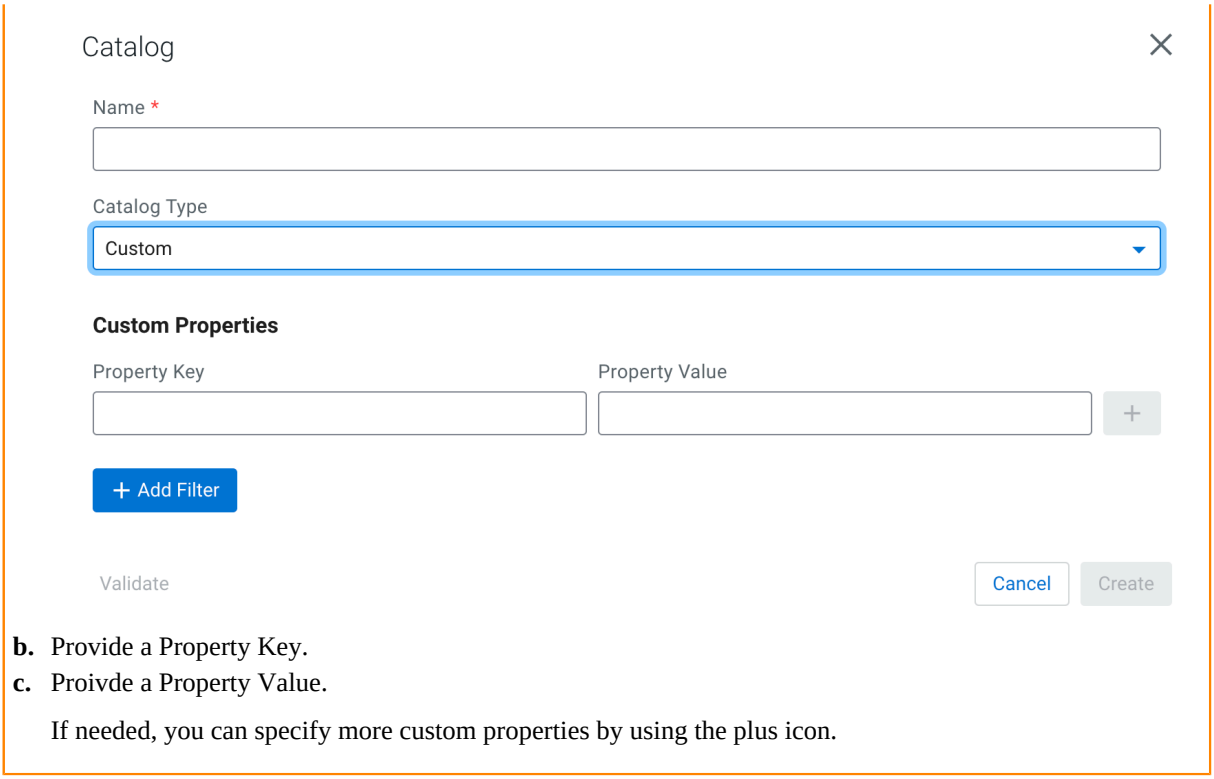


b. Provide a Name to the Hive catalog created in SSB.

c. Provide the name of the Hive Default Database.

For Custom

a. Select Custom from the Catalog Type drop-down.



b. Provide a Property Key.
c. Proivde a Property Value.

If needed, you can specify more custom properties by using the plus icon.

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 Create.

Results

You are ready to use the added catalog in SSB with Flink DDL. The already existing schemas in Schema Registry, tables in Kudu and Hive are automatically imported to SSB.

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. Select Kafka Provider or Catalogs tab.
3. Search for the Kafka provider or catalog you want to modify.
4. Hover over the data provider you need to edit.
5. Click Edit.

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

   Note: You must validate the modified catalog before saving the changes.
7. Click Update.

Deleting registered Data Providers

1. Click Data Providers from the main menu.
2. Select Kafka Provider or Catalogs tab.
3. Search for the Kafka provider or catalog you want to delete.
4. Hover over the data provider you need to edit.
5. Click Delete.
6. Click Delete to confirm the removal of the data provider or catalog.