

Monitoring Kafka Connect using Streams Messaging Manager

Date published: 2020-05-06

Date modified: 2020-08-10



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

Kafka Connect Overview.....	4
Default view of Kafka Connect in the SMM UI.....	4
Creating a Connector.....	5
Modifying a Connector.....	8
Deleting a Connector.....	9
Monitoring Connectors.....	9
Monitoring Connector Profile.....	10
Monitoring Connector Settings.....	10
Monitoring Cluster Profile.....	10

Kafka Connect Overview

Kafka Connect is a framework for connecting Kafka with external systems. You can use Kafka Connect to deploy connector implementations for common data sources and sinks to move data to and from Kafka.

You can create the following types of connectors through Kafka Connect in SMM:

- **Source Connector.** A source connector delivers data from external systems like databases to Kafka topics.
- **Sink Connector.** A sink connector delivers data from Kafka topics to external systems, for example, S3 and Hadoop.

You can use the Kafka Connect option in SMM to create connectors in your cluster, and read data from or write data to Kafka topics.

Default view of Kafka Connect in the SMM UI

Learn where the Kafka Connect option lies in the SMM UI and what are the different options that you can use while connecting Kafka with external systems.

After you configure Kafka Connect in SMM, you can see the Connect option in the SMM UI.

The screenshot shows the 'Overview' page of the SMM UI for 'Cluster: Cluster 1'. At the top, there are four summary cards: Producers (1), Brokers (3), Topics (18), and Consumer Groups (1). Below these, there are tabs for 'TOPICS (18)' and 'BROKERS (3)'. The 'TOPICS (18)' tab is active, displaying a table of topics. The table has columns: NAME, DATA IN, DATA OUT, MESSAGES IN, and CONSUMER GROUPS. The topics listed are: connect-status, connect-offsets, connect-configs, avro_topic, __smm-app-smm-producer-ta..., and __smm-app-smm-producer-ta... (repeated). Each topic has a green checkmark icon in the NAME column. To the right of the table, there is a 'Consumer Groups (1)' section with tabs for ACTIVE (0), PASSIVE (1), and ALL. The 'Connect' option is visible in the left sidebar.

When you click the Connect option, the Connect Cluster page appears.

The screenshot shows the 'Connect Cluster' page for 'Cluster: Cluster 1'. The page title is 'connect-default-cluster'. There is a 'New Connector' button in the top right. Below the title, there is a 'Connectors' section with a summary of connector status: TOTAL CONNECTORS (0), RUNNING CONNECTORS (0), FAILED CONNECTORS (0), DEGRADED CONNECTORS (0), and PAUSED CONNECTORS (0). Below this, there are three main sections: 'Source Connectors (0)', 'Topics (0)', and 'Sink Connectors (0)'. Each section has a search bar and a filter dropdown (All, Running, Paused, Failed). The 'Source Connectors' and 'Sink Connectors' sections also have a 'Tasks' column. The 'Topics' section has a 'Name' column.

The Connect Cluster page shows the connector and cluster details.

At the top-right corner of the Connect Cluster page, you can see the name of the cluster.

The New Connector option below the cluster name enables you to create connectors in your cluster.

In the Connectors section, you can view the number of total connectors, running connectors, failed connectors, degraded connectors, and paused connectors.

In the Connectors tab, you can view details of the source connectors, topics, and sink connectors in the cluster. The Source Connectors and Sink Connectors sections show all, running, paused, and failed connectors with connector name and associated task details. Both Source Connectors and Sink Connectors sections contain a Search option that enables you to search for particular connector details. The Topics section shows the Kafka topic names where data is read from or written to. For more details on connectors, see the *Monitoring Connectors* section.

In the Cluster Profile tab, you can view details of the cluster and workers. For more details on cluster profile, see the *Monitoring Cluster Profile* section.

Related Information

[Monitoring Connectors](#)

[Monitoring Cluster Profile](#)

Creating a Connector

This section describes how to create a connector.

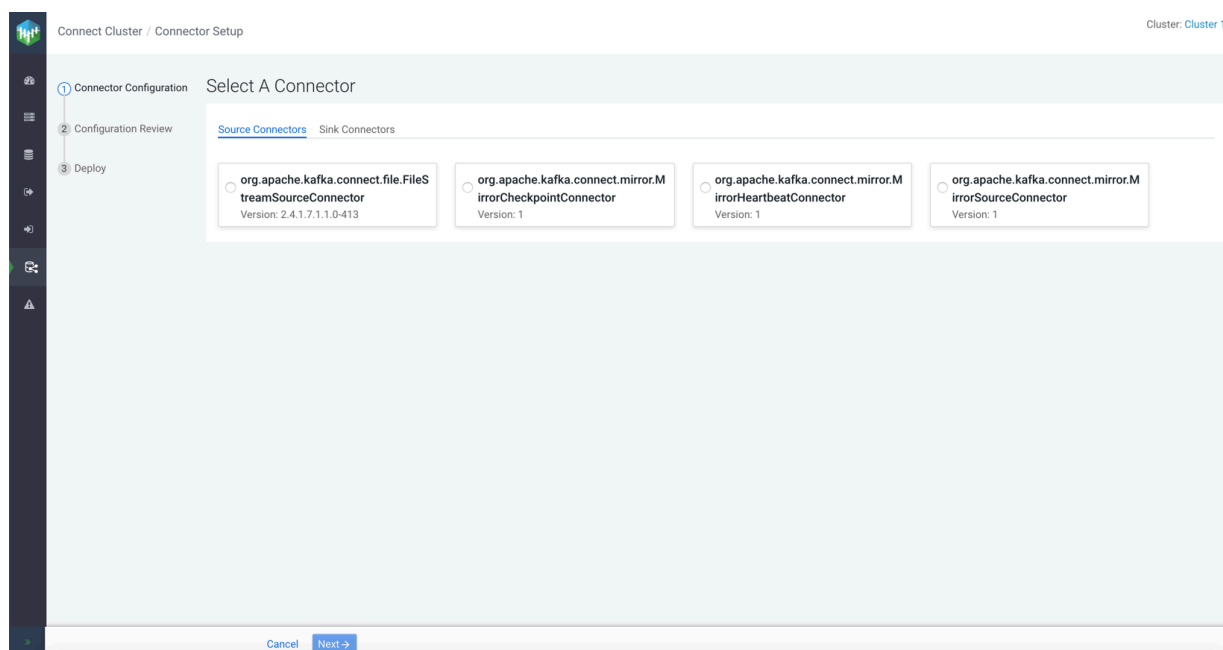
About this task

Perform the following steps to create a connector:

Procedure

1. Click the New Connector option in the upper right corner of the Connect Cluster page.

The Select A Connector screen appears.



- Go to Source Connectors or Sink Connectors tab based on your requirement.

Both the tabs show the list of available connectors.

- Select a connector.

The configuration options appear.

Connect Cluster / Connector Setup Cluster: Cluster 1

1 Connector Configuration 2 Configuration Review 3 Deploy

Select A Connector

Source Connectors Sink Connectors

☒ **org.apache.kafka.connect.file.FileStreamSourceConnector**
Version: 2.4.1.7.1.1.0-413

☐ org.apache.kafka.connect.mirror.MirrorCheckpointConnector
Version: 1

☐ org.apache.kafka.connect.mirror.MirrorHeartbeatConnector
Version: 1

☐ org.apache.kafka.connect.mirror.MirrorSourceConnector
Version: 1

Connector Configuration

Enter Name

1

JSON Validation ✓ Validate

Cancel Next →

- Enter a name for the connector in the Connector Configuration section.

A sample configuration appears for some connectors.

The following image shows the sample configuration for HDFS sink connector:

Connect Cluster / Connector Setup Cluster: Cluster 1

1 Connector Configuration 2 Configuration Review 3 Deploy

Select A Connector

Source Connectors Sink Connectors

☒ **com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector**
Version: 0.0.1-SNAPSHOT

☐ com.cloudera.dim.kafka.connect.s3.S3SinkConnector
Version: 0.0.1-SNAPSHOT

☐ org.apache.kafka.connect.file.FileStreamSinkConnector
Version: 2.4.1.7.1.1.0-413

Connector Configuration

Enter Name

```

1 {
2   "connector.class": "com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector",
3   "tasks.max": "1",
4   "key.converter": "org.apache.kafka.connect.storage.StringConverter",
5   "value.converter": "com.cloudera.dim.kafka.connect.converters.AvroConverter",
6   "value.converter.passthrough.enabled": "true",
7   "value.converter.schema.registry.url": "http://schema-registry:9090",
8   "topics": "avro_topic",
9   "hdfs.uri": "hdfs://namenode:9000",
10  "hdfs.output": "/tmp/topics_output/",
11  "output.storage": "com.cloudera.dim.kafka.connect.hdfs.HdfsPartitionsStorage",
12  "output.writer": "com.cloudera.dim.kafka.connect.partition.writers.avro.AvroPartitionWriter",
13  "output.avro.passthrough.enabled": "true",
14  "hdfs.kerberos.authentication": "true",
15  "hdfs.kerberos.user.principal": "${cm-agent:ENV:kafka_connect_service}",
16  "hdfs.kerberos.keytab.path": "${cm-agent:keytab}",
17  "hdfs.kerberos.namenode.principal": "hdfs/HOST@REALM",

```

JSON Validation ✓ Validate

Cancel Next →

5. Optional. Modify the sample configuration as per your requirement.

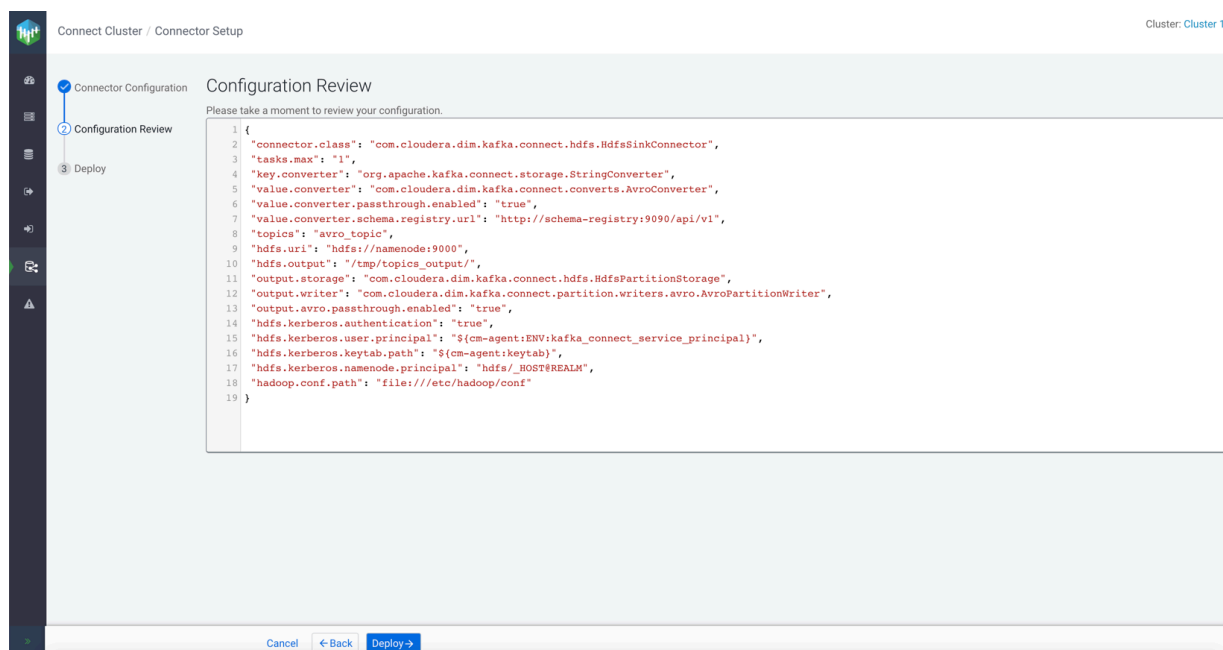
You can refer to environment variables by using a syntax similar to `${cm-agent:ENV:ENVIRONMENT_VARIABLE_NAME}`. You should replace the `ENVIRONMENT_VARIABLE_NAME` with the exact environment variable name.

Also, as shown in the sample configuration, you can refer to the keytab path by specifying `${cm-agent:keytab}`.

6. Click Validate.

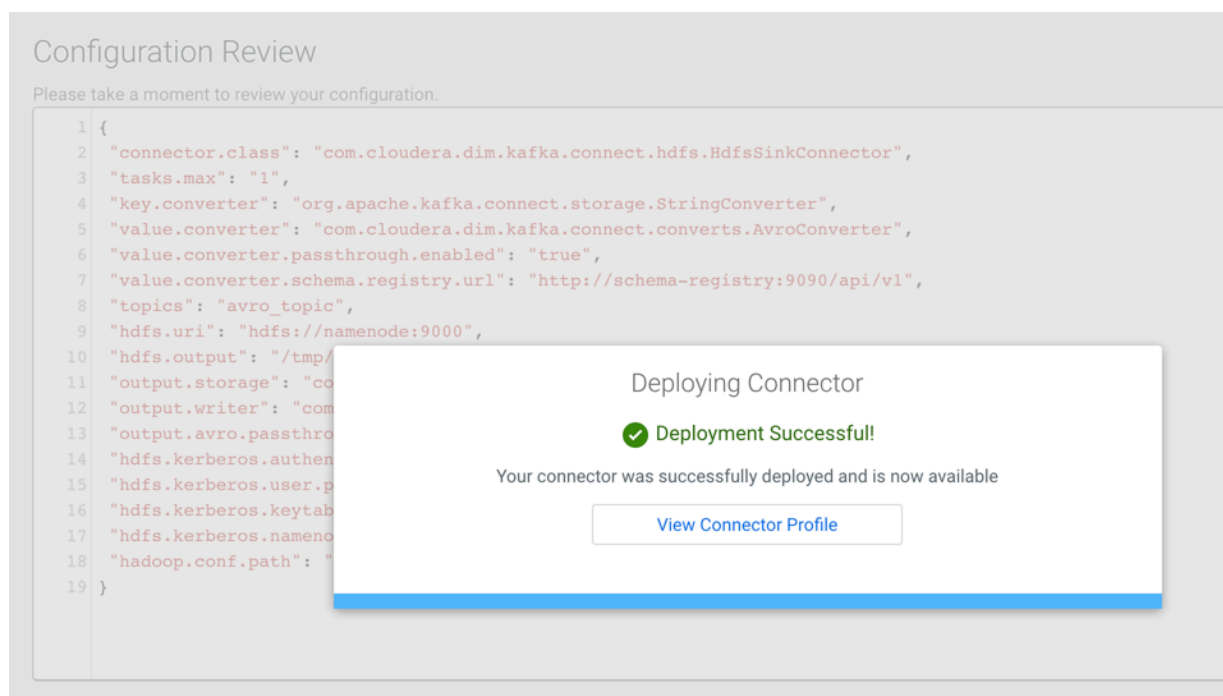
7. Click Next.

The Configuration Review page appears.



8. Review your connector configuration and click Deploy.

The Deploying Connector dialog appears.



9. Click View Connector Profile.

The Connector Profile page appears where you can view your connector details.

Modifying a Connector

This section describes how to modify a connector.

About this task

Perform the following steps to modify a connector settings:

Procedure

1. Go to the Connect Cluster page.
2. Identify the connector you want to modify.
3. Click Profile beside the connector.

The Connector Profile page appears.

The screenshot shows the 'Connector Profile' page for a connector named 'testsinkconnector'. The page header includes 'Connect Cluster / Connector Profile' and 'Cluster: Cluster 1'. Below the header, there are buttons for 'Pause', 'Resume', 'Restart', 'Delete', and 'New Connector'. The main content area is titled 'Connector Profile' and displays the following information:

- CLASSNAME:** com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector
- ASSIGNED WORKER:** 172.27.125.67:38083
- STATUS:** RUNNING (indicated by a green checkmark)
- TOTAL TASKS:** 1
- RUNNING TASKS:** 0
- FAILED TASKS:** 1
- PAUSED TASKS:** 0

Below this information is a 'Tasks' section with a search bar and a 'Restart' button. A table lists the tasks:

Status	Worker ID	Task ID	Put Batch Avg Time	Sink Record Send Rate	Partition Count
●	172.27.125.67	0	NA	NA	NA

4. Go to the Connector Settings tab.

The screenshot shows the 'Connector Settings' page for the same connector 'testsinkconnector'. The page header is the same as the previous screenshot. Below the header, there are buttons for 'Pause', 'Resume', 'Restart', 'Delete', and 'New Connector'. The main content area is titled 'Connector Configuration' and displays a JSON configuration for the connector. To the right of the configuration is a 'JSON Validation' section with a 'Validate' button and a message 'JSON is valid'.

```

1 {
2   "connector.class": "com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector",
3   "hdfs.uri": "hdfs://namenode:9000",
4   "tasks.max": "1",
5   "topics": "avro_topic",
6   "hdfs.kerberos.authentication": "true",
7   "hdfs.kerberos.user.principal": "${cm-agent:ENV:kafka_connect_service_principal}",
8   "hdfs.kerberos.keytab.path": "${cm-agent:keytab}",
9   "value.converter.schema.registry.url": "http://schema-registry:9090/api/v1",
10  "hdfs.kerberos.namenode.principal": "hdfs/_HOST@REALM",
11  "value.converter.passthrough.enabled": "true",
12  "hdfs.output": "/tmp/topics_output/",
13  "output.avro.passthrough.enabled": "true",
14  "hadoop.conf.path": "file:///etc/hadoop/conf",
15  "name": "testsinkconnector",
16  "output.writer": "com.cloudera.dim.kafka.connect.partition.writers.avro.AvroPartitionWriter",
17  "value.converter": "com.cloudera.dim.kafka.connect.converts.AvroConverter",

```


5. Click Edit at the bottom-left corner of the page.
6. Modify your connector configuration, and click Validate.
7. Click Update.

The Deploying Connector dialog appears.

8. Click View Connector Profile.

The Connector Profile page appears where you can view your connector details.

Deleting a Connector

This section describes how to delete a connector.

About this task

Perform the following steps to delete a connector:

Procedure

1. Go to the Connect Cluster page.
2. Identify the connector you want to delete.
3. Click the Profile icon beside the connector.

The Connector Profile page appears.

The screenshot shows the 'Connector Profile' page for a connector named 'testsinkconnector' in 'Cluster: Cluster 1'. The page has a sidebar on the left with navigation icons. The main content area shows the connector's details:

- Connector Name:** testsinkconnector
- Buttons:** Pause, Resume, Restart, Delete, New Connector
- Connector Profile Section:**
 - CLASSNAME:** com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector
 - ASSIGNED WORKER:** 172.27.125.67:38083
 - STATUS:** RUNNING (indicated by a green checkmark)
 - TOTAL TASKS:** 1
 - RUNNING TASKS:** 0
 - FAILED TASKS:** 1
 - PAUSED TASKS:** 0
- Tasks Section:**
 - Search by host:** (input field with a search icon and a Restart button)
 - Table:**

Status	Worker ID ↓	Task ID	Put Batch Avg Time	Sink Record Send Rate	Partition Count
❌	172.27.125.67	0	NA	NA	NA

4. Click Delete at the top-right of the page.
5. Click Yes.

Monitoring Connectors

After you create a connector, you can monitor the details of the connector in the Connector Profile and Connector Settings pages.

If you hover the mouse over a connector in the Connect Cluster page, you can see the class name and status of the connector. You can also see the options to pause, resume, and restart the connector.

If you click the connector name in the Connect Cluster page, you can see the complete workflow for that connector.

To monitor a connector, click the Profile option. The Connector Profile page appears.

Monitoring Connector Profile

The Connector Profile tab enables you to monitor details of the connector and task.

In the Connector Profile section, you can view and monitor Classname, Assigned Worker, Status, Total Tasks, Running Tasks, Failed Tasks, and Paused Tasks.

In the Tasks section, you can view and monitor Status, Worker ID, Task ID, Put Batch Avg Time, Sink Record Send Date, and Partition Count.

Click the metrics arrow to monitor Running Ratio, Offset Commits, Additional Sink Record Metrics, and Additional Sink Start Metrics for the selected task.

The Restart option enables you to restart the task. To restart a particular task, select the task and click Restart.

The Tasks section contains a Search option that enables you to search for particular task details.

Monitoring Connector Settings

The Connector Settings tab enables you to monitor connector settings.

In the Connector Settings section, you can view and monitor the Connector Configuration and JSON Validation. You can edit the connector configuration details here and click Validate.

Use the Pause, Resume, and Restart options to pause, resume, and restart a connector respectively.

Monitoring Cluster Profile

The Cluster Profile tab enables you to monitor details of the cluster and workers.

In the Cluster Details section, you can monitor the Cloudera Manager URL, number of workers, and uptime details.

In the Workers section, you can monitor details of worker host, connector count, connector startup attempts, connector startup failures, task count, task startup attempts, and task startup failures.

If you click the metrics arrow, you can monitor the details and metrics of workers. The worker metrics showcases the hostname, details of the connector assigned to the worker, connector metrics, task metrics, and worker rebalance metrics.

The Workers section contains a Search option that enables you to search for particular worker details.