

1.19.

## Configuring Data Connectors

Date published: 2023-04-05

Date modified: 2023-04-13

# CLOUDERA

# 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

<b>Using Ozone storage with Cloudera Data Engineering Private Cloud.....</b>	<b>4</b>
Adding Ozone data connector for Cloudera Data Engineering service.....	4
Deleting Ozone data connector for Cloudera Data Engineering service.....	5
Backing up and Restoring Data Connectors.....	6
Backing up the data connector.....	6
Restoring the data connector.....	7

# Using Ozone storage with Cloudera Data Engineering Private Cloud

Apache Ozone is an object store available on the CDP Private Cloud Base cluster which enables you to optimize storage for big data workloads. You can configure Ozone as the backend storage for workloads of CDE clusters.

Data connectors enable you to access different storage using only a few configurations specific to storage. Data Connectors are bound to a CDE service. Hence, you must first create a CDE service before configuring Ozone storage through Data Connectors. All the virtual clusters managed by that CDE service can use the same data connectors.

## Related Information

[Adding a Cloudera Data Engineering service](#)


## Adding Ozone data connector for Cloudera Data Engineering service

You can configure Ozone as the backend storage for Cloudera Data Engineering (CDE) workloads.

### Before you begin

Data connectors are bound to a CDE service and you must first create a CDE service before configuring Ozone storage.

### Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Click Administration in the left navigation menu. The Administration page displays.
3. In the Services column, click  for the CDE service you want to configure for accessing the Ozone file system.
4. Click the Data Connectors tab. On the Data Connectors tab, you can view details about the existing data connectors, if any, including the service name and CDP environment.
5. Click Create. The Add a Connector pane is displayed.
6. Specify the connector's name.

7. Select Ozone from the Type of Connector drop-down list.

The screenshot shows the 'Add a Data Connector' interface. At the top, there are four tabs: 'Configuration', 'Charts', 'Logs', and 'Data Connectors'. The 'Data Connectors' tab is selected. Below the tabs, the title 'Add a Data Connector' is displayed. The form contains two main fields: 'Name' with a red asterisk, which has a text input field containing 'ozone-2'; and 'Type of connector' with a red asterisk, which is a dropdown menu currently showing 'OZONE'. At the bottom of the form, there are two buttons: a light blue 'Cancel' button and a blue 'Create' button. A mouse cursor is pointing at the 'Create' button.

8. Review the summary and click Create.

### Results

After configuring the CDE service to access the Ozone filesystem, use this data connector during job creation to read and write from Apache Ozone Object Store through your workload Spark job.

### Related Information

[Adding a Cloudera Data Engineering service](#)



[Creating jobs in Cloudera Data Engineering](#)

## Deleting Ozone data connector for Cloudera Data Engineering service

You can delete an existing data connector.

### Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.

2. Click Administration in the left navigation menu. The Administration page displays.
3. In the Services column, click  icon for the CDE service you want to configure for accessing the Ozone file system.
4. Click the Data Connectors tab. On the Data Connectors tab, you can view details about the existing data connectors, if any, including the service name and CDP environment.
5. Click  in the Actions column next to the data connector, and then click Delete.

## Backing up and Restoring Data Connectors

Before you upgrade your clusters or delete an existing cluster to create a new cluster, you can back up the data connectors and restore those data connectors post-upgrade on the new cluster. The data connector pre-upgrade jobs work seamlessly on the upgraded cluster after you restore those jobs.


### Backing up the data connector

You must back up your data connectors before you delete or upgrade your cluster.

#### Before you begin

Install `jq` on the host machine where you want to run the backup and restore commands.


#### Procedure

1. Get the CDE Service URL.
  - a. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
  - b. In the CDE Services column, click  for the CDE service.
  - c. Click GRAFANA CHARTS and get the domain name.
  - d. Note the URL. For example, if the URL copied is `https://service.cde-n82xsqpr.apps.apps.shared-rke-dev-01.kcloud.example.com/grafana/d/sK1XDusZz/kubernetes`, use only the `https://service.cde-n82xsqpr.apps.apps.shared-rke-dev-01.kcloud.example.com` as the CDE Service URL.
2. Get the CDE token for your CDE service URL.

```
export CDE_TOKEN=$(curl -k -L -u <user-name>:<password> https://<cde service url>/gateway/authnkn/knoxtoken/api/v1/token | jq -r .access_token)
```

Example:

```
export CDE_TOKEN=$(curl -k -L -u dexssoadmin:Password@123 https://service.cde-4x6s29mx.apps.apps.shared-os-qe-01.kcloud.example.com/gateway/authnkn/knoxtoken/api/v1/token | jq -r .access_token)
```

3. Note the cluster-ID.
  - a. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
  - b. In the CDE Services column, click  for the CDE service.
  - c. Click GRAFANA CHARTS and get the domain name.
  - d. Note the Cluster ID shown on the page.
4. Provide the CDE token and Cluster ID and run the data connector backup command.

```
curl --location --request POST 'https://<cde service url>/data-connectors/v1/backupDataConnections' \
--header "Authorization: Bearer ${CDE_TOKEN}" \
--data-raw '{
  "clusterId": "<cluster id of cluster that needs backing up>"
}
```

```
}' --insecure | jq .zipFile | tr -d '"' | base64 -d > backup-dataconnections.zip
```

Example:

```
curl --location --request POST 'https://service.cde-4x6s29mx.apps.apps.shared-os-qe-01.kcloud.cloudera.com/data-connectors/v1/backupDataConnections' \
--header "Authorization: Bearer ${CDE_TOKEN}" \
--data-raw '{
  "clusterId": "cluster-4x6s29mx"
}' --insecure | jq .zipFile | tr -d '"' | base64 -d > backup-dataconnections.zip
```


## Related Information

[Installing jq](#)

## Restoring the data connector

You can restore the backed-up data connectors post-upgrade or on the new CDE Service in the same cluster or across clusters.

### Procedure

1. Get the CDE Service URL.
  - a. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
  - b. In the CDE Services column, click  for the CDE service.
  - c. Click GRAFANA CHARTS and get the domain name.
  - d. Note the URL. For example, if the URL copied is `https://service.cde-n82xsqpr.apps.apps.shared-rke-dev-01.kcloud.example.com/grafana/d/sK1XDusZz/kubernetes`, use only the `https://service.cde-n82xsqpr.apps.apps.shared-rke-dev-01.kcloud.example.com` as the CDE Service URL.
2. Get the CDE token for the CDE Service that you want to restore.

```
export CDE_TOKEN=$(curl -k -L -u <user-name>:<password> https://<cde service url>/gateway/authkn/knoxtoken/api/v1/token | jq -r .access_token)
```

Example:

```
export CDE_TOKEN=$(curl -k -L -u dexssoadmin:Password@123 https://service.cde-4x6s29mx.apps.apps.shared-os-qe-01.kcloud.example.com/gateway/authkn/knoxtoken/api/v1/token | jq -r .access_token)
```


3. Encode the backed-up zip file to base64 to use it in the restore process.

```
base64 -i <backup-dataconnections.zip>
```

Example:

```
base64 -i backup-dataconnections.zip
dGhpcyBpcyBhbiBleGFtcGxlIGVuY29kZWQgc3RyaW5n
```

dGhpcyBpcyBhbiBleGFtcGxlIGVuY29kZWQgc3RyaW5n is the output.

4. Note the Cluster ID.
  - a. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
  - b. In the CDE Services column, click  for the CDE service.
  - c. Click GRAFANA CHARTS and get the domain name.
  - d. Note the Cluster ID shown on the page.
5. Run the restore API to restore the data connector.

```
curl --location --request POST '<CDE Service URL>/data-connectors/v1/restoreDataConnections' \
--header "Authorization: Bearer ${CDE_TOKEN}" \
--data-raw '{
  "clusterId": "<cluster id to restore into>",
  "handleDuplicates":false,
  "backupZipFile": "<base64 string from last step>"
}'--insecure
```

Example:

```
curl --location --request POST 'https://service.cde-wwkd776f.apps.apps.s
hared-os-ge-01.kcloud.cloudera.com/data-connectors/v1/restoreDataConnect
ions' \
--header "Authorization: Bearer ${CDE_TOKEN}" \
--data-raw '{
  "clusterId": "cluster-wwkd776f",
  "handleDuplicates": false,
  "backupZipFile": "dGhpcyBpcyBhbiBleGFtcGxlIGVuY29kZWQgc3RyaW5n"
}' --insecure
```

The `handleDuplicates` parameter controls the restore behavior if data connectors with the same name are found on the cluster during the restore process. If there are data connectors with the same name and `handleDuplicates` is set to `false`, the restoration fails with an error. If `handleDuplicates` is set to `true`, data connectors are restored with a different name and the format of the new name is `<old-name>-restored-<random string of length 3>`.