

..

Connecting to Data

Date published: 2020-10-30

Date modified: 2022-09-21

CLOUDBERA

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

Connecting to a data source in CDP Data Visualization.....	4
Creating a CML data connection to an Impala data warehouse.....	4
Creating a CML data connection to a Hive data warehouse.....	9
Creating a CDW data connection in Data Visualization.....	14
Creating a CDSW data connection to a data warehouse.....	20
Editing a data connection.....	26
Deleting a data connection.....	29
Using the Connection Explorer.....	30
Discovering the Connection Explorer interface.....	31
Previewing data table details.....	33
Sample data.....	33
Datasets.....	33
Using the Direct Access interface.....	34
Running a SQL query in Direct Access.....	34
Downloading the results of a Direct Access query.....	36

Connecting to a data source in CDP Data Visualization

Data Visualization allows you to create connections to many types of external data sources.

When using Data Visualization in CDP Public Cloud with Cloudera Data Warehouse (CDW), the data connection is automatically set up, but you can connect to other data sources as well. In Cloudera Machine Learning (CML), you can connect to an Impala or a Hive data warehouse, or tie in data from predictive CML models.

The supported connection types are:

- Impala
- Hive
- Druid
- SQLite
- MySQL
- MariaDB
- PostgreSQL

Related Information

[Connections](#)

[Datasets](#)

[Data modeling](#)

Creating a CML data connection to an Impala data warehouse

Learn how to connect natively to data stored in Impala when using CDP Data Visualization in Cloudera Machine Learning (CML).

About this task

You must connect to your data prior to using the data modeling and visualization functions. The following steps show you how to create a new CML data connection to an Impala data warehouse.



Note:

Only users who have the Manage data connections privilege or administrators can create and manage connections in CDP Data Visualization.

You must be an administrator to be able to set privileges for a user. If you want to log in as an administrator, you can use the default admin account:

- username: vizapps_admin
- password: vizapps_admin

When you create a connection, you automatically have privileges to create and manage datasets on this connection, and to build dashboards and visuals in these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a role, see *Setting role privileges*.
- For instructions on how to assign the administrator role to a user, see *Promoting a user to administrator*.

Procedure

1. On the main navigation bar, click DATA.

The DATA interface appears, open on the Datasets tab.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA' (which is highlighted). Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'Datasets' tab is selected and highlighted with a red box. The main content area displays a table of datasets with the following columns: Title/Table, ID, Created, Last Updated, Modified By, and # Visuals. The table lists various datasets such as 'Food Stores Inspection in NYC', 'Cereals', 'Earthquake Data January 2019', 'World Life Expectancy', 'US County Population', 'US State Populations Over Time', 'Global Information Security Threats', 'Restaurant Inspection SF', 'Iris', and 'NYC Taxicab Rides Detail'.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infoseq_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

2. In the Data side menu bar, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with Manage data connections privilege and to administrators.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar has a 'NEW CONNECTION' button highlighted with an orange box, along with 'All Connections' and 'samples'. The main content area shows a 'NEW DATASET' and 'ADD DATA' button, a 'Datasets' list with 13 items, and a 'Connection Explorer' button. Below this is a table of datasets.

Title/Table	ID	Created
Test Dataset main.census_pop	13	Dec 08, 2021
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021
Cereals main.cereals	11	Nov 22, 2021
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021

The Create New Data Connection modal window appears.

Create New Data Connection

Connection type

Connection name

Basic **Cache**

Hostname or IP address

Port #

Database

Credentials

Username

Password

3. Select the Impala Connection type from the drop-down list and enter the hostname or IP address of the running coordinator.
You can get the coordinator hostname from the JDBC URL of the Impala DW.
4. Enter 443 in the Port # field.
5. Enter your workload username and password as credentials.

6. Click the Advanced tab and make the selections below:

Create New Data Connection

Connection type

Connection name

Basic **Advanced** Parameters Cache

Connection mode Binary HTTP

HTTP Path

Socket type Normal SSL SSL with certificate

Authentication mode NoSasl Plain LDAP Kerberos

Query Timeout

Session Timeout

Socket Timeout

Queue Depth

Impersonation Enabled

Trusted Impersonation Enabled

7. Locate the Impala Endpoint for the data hub.

Name	URL
CM-API	https://jingalls-test-dm-gateway.euph-aw.a465-9q4k.cloudera.site/jingalls-test-dm/cdp-proxy-api/cm-api/
Impala	https://jingalls-test-dm-gateway.euph-aw.a465-9q4k.cloudera.site/jingalls-test-dm/cdp-proxy-api/impala/
Impala	jdbc:impala://jingalls-test-dm-gateway.euph-aw.a465-9q4k.cloudera.site:443/;ssl=1;transportMode=http;httpPath=jingalls-test-dm/cdp-proxy-api/impala;AuthMech=3;

8. Copy it and paste it into the HTTP Path field.

9. Click TEST to test the connection.

10. Click CONNECT to create the connection.

Results

You have set up a connection to a running Impala DW.

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

Creating a CML data connection to a Hive data warehouse

Learn how to connect natively to data stored in Hive when using CDP Data Visualization in Cloudera Machine Learning (CML).

About this task

You must connect to your data prior to using the data modeling and visualization functionalities. The following steps show you how to create a new CML data connection to a Hive data warehouse.



Note:

Only users who have the Manage data connections privilege or administrators can create and manage connections in CDP Data Visualization.

You must be an administrator to be able to set privileges for a user. If you want to log in as an administrator, you can use the default admin account:

- username: vizapps_admin
- password: vizapps_admin

When you create a connection, you automatically have privileges to create and manage datasets on this connection, and to build dashboards and visuals in these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a role, see *Setting role privileges*.
- For instructions on how to assign the administrator role to a user, see *Promoting a user to administrator*.

Procedure

1. On the main navigation bar, click DATA.

The DATA interface appears, open on the Datasets tab.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA' (which is highlighted). Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'Datasets' tab is selected and highlighted with a red box. The main content area displays a table of datasets with the following columns: Title/Table, ID, Created, Last Updated, Modified By, and # Visuals. The table lists various datasets such as 'Food Stores Inspection in NYC', 'Cereals', 'Earthquake Data January 2019', 'World Life Expectancy', 'US County Population', 'US State Populations Over Time', 'Global Information Security Threats', 'Restaurant Inspection SF', 'Iris', and 'NYC Taxicab Rides Detail'.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infoseq_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

2. In the Data side menu bar, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with Manage data connections privilege and to administrators.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar has a 'NEW CONNECTION' button highlighted with an orange box. Below it are 'All Connections' and a search bar with 'samples' entered. The main content area shows a 'NEW DATASET' and 'ADD DATA' button, followed by a 'Datasets' section with a 'Connection Explorer' button. A table lists datasets with columns for Title/Table, ID, and Created.

Title/Table	ID	Created
Test Dataset main.census_pop	13	Dec 08, 2021
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021
Cereals main.cereals	11	Nov 22, 2021
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021

The Create New Data Connection modal window appears.

Create New Data Connection

Connection type	<input type="text" value="Select an Option"/>
Connection name	<input type="text"/>

Basic	Cache
-------	-------

Hostname or IP address	<input type="text" value="Enter IP address of the server where your data resides (example: prod_db.yourcompany.com or 10.0.1.20)"/>
Port #	<input type="text" value="0"/>
Database	<input type="text"/>

Credentials

Username	<input type="text"/>
Password	<input type="text"/>

<input type="button" value="TEST"/>	<input type="button" value="CANCEL"/>	<input type="button" value="CONNECT"/>
-------------------------------------	---------------------------------------	--

3. Select the Hive Connection type from the drop-down list and enter the hostname or IP address of the running coordinator.

You can get the coordinator hostname from the JDBC URL of the Hive DW.

4. Use port 443.

5. Click the Advanced tab and make the selections below:

Create New Data Connection

Connection type

Connection name

Basic **Advanced** Parameters Cache

Connection mode Binary HTTP

HTTP Path

Access Token

Socket type Normal SSL SSL with certificate

Authentication mode NoSasl Plain LDAP Kerberos

Query Timeout

Session Timeout

Socket Timeout

Queue Depth

Impersonation Enabled

Trusted Impersonation Enabled


- Click the Parameters tab and set the `hive.server2.async.exec.async.compile` parameter to false.

Create New Data Connection

Connection type:

Connection name:

Basic Advanced **Parameters** Cache

Parameter Name	Parameter Value	
<code>hive.server2.async.exec.async.compile</code>	false	
Add new row		

TEST CANCEL **CONNECT**

- Use your workload username and password as credentials.
- Click TEST and then CONNECT to create the connection.

Results

You have set up a connection to a running Hive DW.

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

Creating a CDW data connection in Data Visualization

Learn how to connect to data when using CDP Data Visualization in Cloudera Data Warehouse (CDW) data service. You can connect Data Visualization to a Virtual Warehouse to visualize your data. Similar to using a BI client, you can configure and connect to Virtual Warehouses from different clusters.

About this task

You must connect to your data prior to using the data modeling and visualization functions. You make the connection to the Virtual Warehouse when you select your warehouse in the steps below. The CDW Warehouse URL has the same compute instance ID as your Virtual Warehouse.

**Note:**

Only users who have the Manage data connections privilege or administrators can create and manage connections in CDP Data Visualization. In CDW, these are the members of the Admin Groups associated with the CDV instance.

When you create a connection, you automatically have privileges to create and manage datasets on this connection, and to build dashboards and visuals in these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a role, see *Setting role privileges*.
- For instructions on how to assign the administrator role to a user, see *Promoting a user to administrator*.

When you are creating a Hive or Impala data connection within the same cluster, the connection is considered secure and trusted, and the connection details can be auto populated with a default authentication user.

Procedure

1. Start Data Visualization from the left navigation panel.
2. On the main navigation bar, click DATA.

The DATA interface appears, open on the Datasets tab.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infosec_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

- In the Data side menu bar, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with Manage data connections privilege and to administrators.

The screenshot shows the Cloudera Data Visualization interface. In the left sidebar, the 'NEW CONNECTION' button is highlighted with an orange box. Below it, there are sections for 'All Connections' and 'samples'. The main content area shows a 'NEW DATASET' and 'ADD DATA' button, followed by a 'Datasets' section with a 'Connection Explorer' link. A table lists several datasets:

Title/Table	ID	Created
Test Dataset main.census_pop	13	Dec 08, 2021
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021
Cereals main.cereals	11	Nov 22, 2021
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021

The **Create New Data Connection** modal window appears.

The 'Create New Data Connection' modal window is shown with the following configuration:

- Connection type:** CDW Impala
- Connection name:** khahn-connection
- CDW Warehouse:** impala-1629029940-xx6t
- Hostname or IP address:** coordinator.impala-1629029940-xx6t.svc.cluster.local
(example: prod_db.yourcompany.com or 10.0.1.20)
- Port #:** 28000

The modal has tabs for 'Basic', 'Advanced', 'Parameters', and 'Data', with 'Basic' currently selected.

- In Connection type, select CDW Hive or CDW Impala.
- Provide a name for the connection.

6. Select a CDW Warehouse to connect to.

For Data connection within the same cluster

The following fields are auto populated:

- Hostname or IP address
- Port #
- Username

For Data connection outside the cluster

Enter the following information:

- Hostname or IP address
- Port #
- Username
- Password

7. Click the Advanced tab and make the appropriate selections.



Important: Depending on the type of connection you are creating, there can be additional tabs in the Create New Data Connection modal window where you have to adjust further settings.

The selections below are correct for the Cloudera iedh system.

Create New Data Connection

Connection type	Impala
Connection name	ImpalaConnection

Basic Advanced Parameters Cache

Connection mode	<input type="radio"/> Binary <input checked="" type="radio"/> HTTP
HTTP Path	SQL path (default cliservice)
Socket type	<input type="radio"/> Normal <input checked="" type="radio"/> SSL <input type="radio"/> SSL with certificate
Authentication mode	<input type="radio"/> NoSasl <input type="radio"/> Plain <input type="radio"/> LDAP <input checked="" type="radio"/> Kerberos
Kerberos service name	Kerberos service name
Query Timeout	60
Session Timeout	60
Socket Timeout	60
Queue Depth	
Impersonation	<input type="checkbox"/> Enabled
Trusted Impersonation	<input type="checkbox"/> Enabled

TEST	CANCEL	CONNECT
------	--------	---------

8. Click TEST.

If the connection is valid, the system returns a Connection Verified message.

9. Click CONNECT.**What to do next**

You can create a data set, and then start creating visuals, dashboards, and applications. For more information, see *Creating datasets* and *Creating a visual*.

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

[Creating datasets](#)

[Creating a visual](#)

Creating a CDSW data connection to a data warehouse

Learn how to connect natively to data stored in a data warehouse when using CDP Data Visualization in Cloudera Data Science Workbench (CDSW).

About this task

You must connect to your data prior to using the data modeling and visualization functionalities. The following steps show you how to create a new CDSW data connection to a running Impala system.

**Note:**

Only users who have the Manage data connections privilege or administrators can create and manage connections in CDP Data Visualization. You must be an administrator to be able to set privileges for a user. If you want to log in as an administrator, you can use the default admin account:

- username: vizapps_admin
- password: vizapps_admin

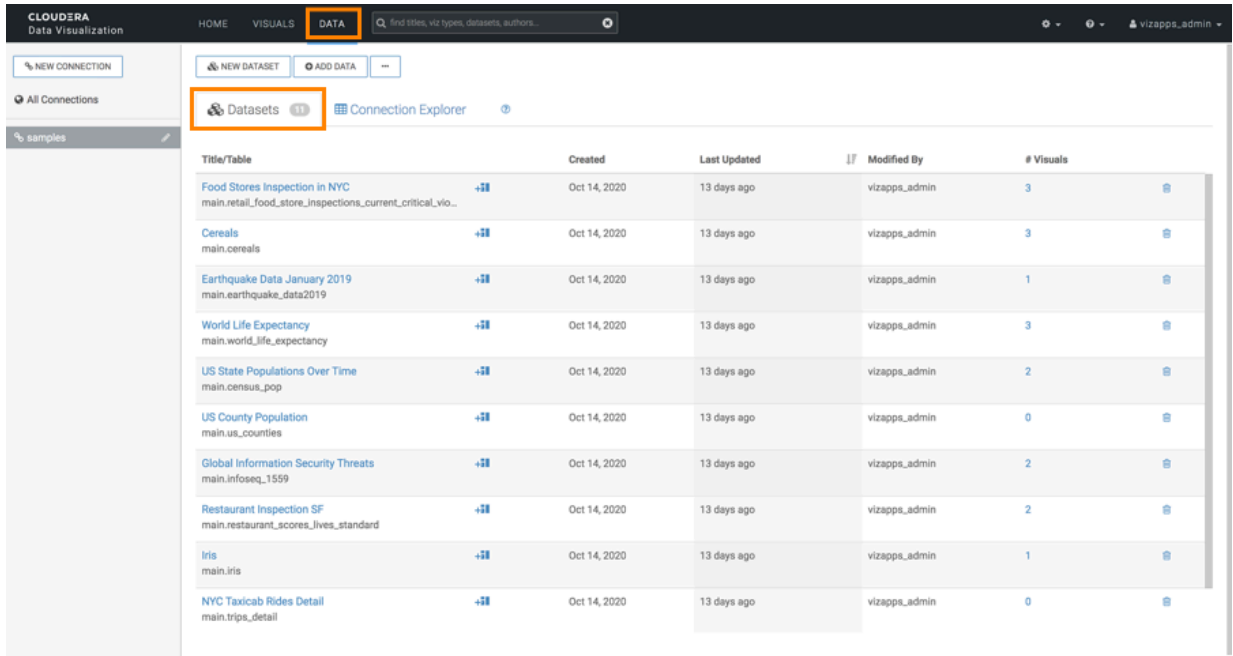
When you create a connection, you automatically have privileges to create and manage datasets on this connection, and to build dashboards and visuals in these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a role, see *Setting role privileges*.
- For instructions on how to assign the administrator role to a user, see *Promoting a user to administrator*.

Procedure

1. On the main navigation bar, click DATA.

The DATA interface appears, open on the Datasets tab.



The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA' (highlighted with an orange box). Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'Datasets' tab is selected and highlighted with an orange box. The main content area displays a table of datasets with columns for Title/Table, Created, Last Updated, Modified By, and # Visuals.

Title/Table	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	Oct 14, 2020	13 days ago	vizapps_admin	3
Cereals main.cereals	Oct 14, 2020	13 days ago	vizapps_admin	3
Earthquake Data January 2019 main.earthquake_data2019	Oct 14, 2020	13 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	Oct 14, 2020	13 days ago	vizapps_admin	3
US State Populations Over Time main.census_pop	Oct 14, 2020	13 days ago	vizapps_admin	2
US County Population main.us_counties	Oct 14, 2020	13 days ago	vizapps_admin	0
Global Information Security Threats main.infoseq_1559	Oct 14, 2020	13 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	Oct 14, 2020	13 days ago	vizapps_admin	2
Iris main.iris	Oct 14, 2020	13 days ago	vizapps_admin	1
NYC Taxicab Rides Detail main.trips_detail	Oct 14, 2020	13 days ago	vizapps_admin	0

2. In the Data side menu bar, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with Manage data connections privilege and to administrators.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar has a 'NEW CONNECTION' button highlighted with an orange box. Below it are 'All Connections' and 'samples'. The main content area shows a 'NEW DATASET' and 'ADD DATA' button, a 'Datasets' section with a count of 13, and a 'Connection Explorer' button. A table lists datasets with columns for Title/Table, ID, and Created.

Title/Table	ID	Created
Test Dataset main.census_pop	13	Dec 08, 2021
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio_...	12	Nov 22, 2021
Cereals main.cereals	11	Nov 22, 2021
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021

The Create New Data Connection modal window appears.

Create New Data Connection

Connection type

Connection name

Basic Cache

Hostname or IP address

Port #

Database

Credentials

Username

Password

3. Select a Connection type from the drop-down list.
4. Provide a name for the connection.
5. Enter the hostname or IP address of the running coordinator.

In this example, you can see to create an Impala connection. You can get the coordinator hostname from the JDBC URL of the Impala DW.

6. Under Port #, enter the port number.
7. Use your workload username and password as credentials.

8. Click the Advanced tab and make the appropriate selections.



Important: Depending on the type of connection you are creating, there can be additional tabs in the Create New Data Connection modal window where you have to adjust further settings.

Create New Data Connection

Connection type	Impala
Connection name	ImpalaConnection

Basic Advanced Parameters Cache

Connection mode	<input type="radio"/> Binary <input checked="" type="radio"/> HTTP
HTTP Path	SQL path (default cliservice)
Socket type	<input type="radio"/> Normal <input checked="" type="radio"/> SSL <input type="radio"/> SSL with certificate
Authentication mode	<input type="radio"/> NoSasl <input type="radio"/> Plain <input type="radio"/> LDAP <input checked="" type="radio"/> Kerberos
Kerberos service name	Kerberos service name
Query Timeout	60
Session Timeout	60
Socket Timeout	60
Queue Depth	
Impersonation	<input type="checkbox"/> Enabled
Trusted Impersonation	<input type="checkbox"/> Enabled

TEST	CANCEL	CONNECT
------	--------	---------

9. Click TEST.

If the connection is valid, the system returns a Connection Verified message.

10. Click CONNECT.

Results

You have set up a connection to a running data warehouse.

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

Editing a data connection

Learn how to edit a data connection in CDP Data Visualization.

About this task

The following steps demonstrate how to edit existing data connections. The example shows changing an Impala connection to a Hive connection.

Procedure

1. On the main navigation bar, click DATA.

The Data view appears, open on the Datasets tab.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infoseq_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

2. In the side bar, click the Edit button to the right of the connection you want to change.

The Edit Data Connection modal window appears.

3. Edit the connection details according to the connection type change you want to implement.

In this example, an Impala connection is changed to a Hive connection.

Edit Data Connection

Connection type	Impala
Connection name	ImpalaConnection

Basic **Advanced** Parameters Cache

Hostname or IP address	10.02.40 <small>(example: prod_db.yourcompany.com or 10.0.1.20)</small>
Port #	21050

Credentials

Username	admin
Password

TEST	CANCEL	DELETE CONNECTION	SAVE
------	--------	-------------------	------

Edit Data Connection

Connection type

Connection name

Basic **Advanced** Parameters Cache

Hostname or IP address
(example: prod_db.yourcompany.com or 10.0.1.20)

Port #

Credentials

Username

Password

4. At the bottom of the modal, click TEST.
5. If the connection is verified, click SAVE.

Connection Verified!

Results

After this operation succeeds, the name of the new type of connection appears on the side navigation bar.

Deleting a data connection

Learn how you can remove a data connection in CDP Data Visualization.

About this task

The following steps show you how to delete an existing data connection.



Tip: You can only delete connections that are not associated with any datasets. To learn how to delete datasets, see *Deleting datasets*.

Procedure

1. On the main navigation bar, click DATA.

The Data view appears, open on the Datasets tab.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infoseq_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

2. In the side bar, click the Edit Connection (pencil) button to the right of the connection you want to delete.

The Edit Data Connection modal window appears.

3. At the bottom of Edit Data Connection modal window, click DELETE CONNECTION.

Edit Data Connection

Connection type

Connection name

Basic **Advanced** Parameters Cache

Hostname or IP address
(example: prod_db.yourcompany.com or 10.0.1.20)

Port #

Credentials

Username

Password

Results

After this operation succeeds, the connection is deleted and its name no longer appears on the side navigation bar.

Related Information

[Deleting datasets](#)

Using the Connection Explorer

CDP Data Visualization enables you to view existing data connections and all data tables accessible through them. In the Connection Explorer interface, you can create new connections to data sources, preview that data, create new

datasets, navigate to these datasets, import supplemental data, and locate existing dashboards and visuals based on specific datasets.

Discovering the Connection Explorer interface

Learn how you can navigate to the Connection Explorer interface and use it to connect to data in CDP Data Visualization.

Procedure

1. On the main navigation bar, click DATA.

The Data view appears, open on the Datasets tab. The Datasets tab lists all existing datasets on the connection.

Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021	16 days ago	vizapps_admin	3
Cereals main.cereals	11	Nov 22, 2021	16 days ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021	16 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021	16 days ago	vizapps_admin	4
US County Population main.us_counties	8	Nov 22, 2021	16 days ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Nov 22, 2021	16 days ago	vizapps_admin	2
Global Information Security Threats main.infosec_1559	6	Nov 22, 2021	16 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	Nov 22, 2021	16 days ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	4	Nov 22, 2021	16 days ago	vizapps_admin	1
Iris main.iris	3	Nov 22, 2021	16 days ago	vizapps_admin	1
NYC Taxicab Rides Detail	2	Nov 22, 2021	16 days ago	vizapps_admin	0

2. In the main area, click the Connection Explorer tab.

Title/Table	ID	Created
Test Dataset main.census_pop	13	Dec 08, 2021
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	Nov 22, 2021
Cereals main.cereals	11	Nov 22, 2021
Earthquake Data January 2019 main.earthquake_data2019	10	Nov 22, 2021
World Life Expectancy main.world_life_expectancy	9	Nov 22, 2021

The Connection Explorer interface appears. The Connection Explorer tab (current) enables you to explore the databases and tables available on the connection, and to manage all functions related to the data on the connection.

In this view, you can click a connection in left navigation that you want to explore and select a database. You can also select a specific table from that database, and explore its details.

The screenshot shows the Cloudera Data Visualization interface. On the left, the 'samples' connection is selected. In the main area, the 'main' database is selected, and a list of tables is displayed. The table 'generalTestForExploreCall' is highlighted. Below the table list, a 'Sample Data' table is shown with the following data:

id	name	age	cost	debt	altname	favoriteanimal	heartrate
1	austen	21	10000	0	a-dog	horse	71
2	trevor	33	203	30	t-money	cat	82
3	nathaniel	22	393	845.84	nate	dog	56
4	james	34	456	944.23	j-dog	seagul	99
5	jake	31	543	23.45	nate	dog	87
6	john	22	34	34.4	nate	dog	66

Showing 1 - 6 of 6 rows

The Connection Explorer interface contains the following items:

The screenshot shows the Cloudera Data Visualization interface. On the left, the 'samples' connection is selected. In the main area, the 'main' database is selected, and a list of tables is displayed. A supplemental menu is open, showing the following options:

- Create Analytical View
- Clear result cache
- Import Visual Artifacts
- Direct Access

Click on a table row above to view sample data and table statistics

- New Connection is for connecting to any source of data.
- New Dataset is for creating datasets, which are necessary for developing dashboards and visuals. You can also use start a new dataset from a specified table.
- For SQLite connections, the Add Data option enables you to introduce data that enriches your datasets from outside sources.
- The Supplemental menu, under the (ellipsis) icon, opens new options.
 - For Impala connections, clicking the Clear result cache option under the supplemental menu reloads the full table definition.
 - Import Visual Artifacts option under the supplemental menu enables you to restore or import visual artifacts from a backup *.json file.
 - Direct Access enables you to access data directly by running SQL queries. You can build datasets from specific SQL queries, as opposed to starting with an existing table.

- e.
- f.
- g. The databases area of the screen shows all databases that you can access through the current connection. In our example, there is one called main (selected). Selecting a database shows its tables.
- h. In the list of tables, the # Datasets column lists the number of datasets that use the particular table as their initial definition.
- i. New Dataset is for creating a dataset on a specific table.

Previewing data table details

Learn how you can preview table details directly in the Connection Explorer interface.

To see more information about data tables in the Connection Explorer, click the row of a table. When you click a row, two tabs, Sample Data and Datasets appear below the list of tables.

Sample data

When you click a table, you can preview the table data in the Sample Data view.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The main area is divided into a left sidebar with 'NEW CONNECTION' and 'All Connections', and a main content area. The main content area has a 'NEW DATASET' and 'ADD DATA' button, and a 'Connection Explorer' tab. The 'Connection Explorer' shows a tree view with 'main' selected. Below this, a table lists tables and their corresponding number of datasets. The 'Iris' table is selected, and a 'Sample Data' view is open below it, showing a table with columns: id, name, age, cost, debt, alname, favoriteanimal, and heartrate. The data is as follows:

id	name	age	cost	debt	alname	favoriteanimal	heartrate
1	austen	21	10000	0	a-dog	horse	71
2	trevor	33	203	30	t-money	cat	82
3	nathaniel	22	393	845.84	nate	dog	56
4	james	34	456	944.23	j-dog	seagui	99
5	jake	31	543	23.45	nate	dog	87
6	john	22	34	34.4	nate	dog	66

Showing 1 - 6 of 6 rows

Datasets

When you click a table, you can check the following data in the Datasets view:

- Title/Table
- Created date
- Last Updated date
- Modified by username
- # Visuals for a link to the dashboards and visuals based on the dataset.

You can also perform the following actions:

- Navigate directly to the dataset interface, where you can rename the dataset, modify fields and other parameters, create joins, and so on.
- Start a new dashboard based on the dataset.

- Order datasets based on any of the table columns.
- Delete datasets.

The screenshot shows the Cloudira Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. A search bar is present with the text 'find sites, viz types, datasets, authors...'. The main content area is divided into several sections:

- NEW CONNECTION**: A button to create a new connection.
- All Connections**: A list of connections, currently showing 'samples'.
- NEW DATASET** and **ADD DATA**: Buttons to create or add data to a dataset.
- Datasets**: A tab showing a list of datasets. The list includes columns for 'Table Name' and '# Datasets'. The datasets listed are:

Table Name	# Datasets	Action
census_pop	1	New dataset
cereals	1	New dataset
chicago_govLpay	0	New dataset
earthquake_data2019	1	New dataset
generalTestForExploreCall	0	New dataset
generaltest_1516160078	0	New dataset
infoseq_1559	1	New dataset
iris	1	New dataset
- Connection Explorer**: A tree view showing the connection structure, currently showing 'main'.
- Sample Data** and **Datasets**: Tabs for viewing data. The 'Datasets' tab is active, showing a table of dataset details:

Title/Table	Created	Last Updated	Modified By	# Visuals
Earthquake Data January 2019 main.earthquake_data2019	Oct 14, 2020	13 days ago	vizappa_admin	1

Using the Direct Access interface

The Direct Access interface of CDP Data Visualization enables you to run SQL queries on data connections directly on the DATA page. You can quickly examine the structure of tables that are available on the connection, build a query using standard SQL syntax, preview its results, and then create a dataset on the query. You can also download the data, if needed.

Running a SQL query in Direct Access

Learn how you can run a SQL query in the Direct Access interface of CDP Data Visualization.

Procedure

1. On the main navigation bar, click DATA.

The DATA view appears, open on the Datasets tab.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA' (highlighted with an orange box). Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'Datasets' tab is selected and highlighted with an orange box. A table of datasets is displayed below, with columns for 'Title/Table', 'Created', 'Last Updated', 'Modified By', and '# Visuals'.

Title/Table	Created	Last Updated	Modified By	# Visuals
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	Oct 14, 2020	13 days ago	vizapps_admin	3
Cereals main.cereals	Oct 14, 2020	13 days ago	vizapps_admin	3
Earthquake Data January 2019 main.earthquake_data2019	Oct 14, 2020	13 days ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	Oct 14, 2020	13 days ago	vizapps_admin	3
US State Populations Over Time main.census_pop	Oct 14, 2020	13 days ago	vizapps_admin	2
US County Population main.us_counties	Oct 14, 2020	13 days ago	vizapps_admin	0
Global Information Security Threats main.infoseq_1559	Oct 14, 2020	13 days ago	vizapps_admin	2
Restaurant Inspection SF main.restaurant_scores_lives_standard	Oct 14, 2020	13 days ago	vizapps_admin	2
Iris main.iris	Oct 14, 2020	13 days ago	vizapps_admin	1
NYC Taxicab Rides Detail main.trips_detail	Oct 14, 2020	13 days ago	vizapps_admin	0

2. Open the Supplemental menu (ellipsis icon) and click >_ Direct Access.

The screenshot shows the Cloudera Data Visualization interface with the 'DATA' view selected. The 'Datasets' tab is active. A supplemental menu is open over the 'Datasets' tab, showing options: 'Create Analytical View', 'Clear result cache', 'Import Visual Artifacts', and '>_ Direct Access' (highlighted with an orange box). The table of datasets is partially visible below the menu.

Title/Table	Created	Last Updated
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	Oct 14, 2020	13 days ago
Cereals main.cereals	Oct 14, 2020	13 days ago
Earthquake Data January 2019 main.earthquake_data2019	Oct 14, 2020	13 days ago

The Direct Access interface appears, where you can select the database you want to access.

3. Add your SQL data query to the Enter SQL below code-enabled text box.

The Autocomplete option is on by default. It validates the syntax of the SQL query you enter.

If you want to limit the number of, you have two options:

- You can add a limit clause in the SQL query syntax.
- You can mark the Add in a "LIMIT 100" clause to any SQL select query that does not have a limit clause option. This limitation is on by default. If you set a record limit in the SQL query, it will override this default option, even if it is checked.

Data Connection: samples

Database:

main

Tables

- restaurant_scores_lives...
- retail_food_store_inspe...
- superstore_sales
- trips
- trips_detail
- us_counties
- world_life_expectancy

Autocomplete on

Add in a "LIMIT 100" clause to any SQL select query that does not have a limit clause

4. Click RUN to execute the SQL query.

After the query executes, the results area shows the query results.

In this example, the following query has been run: `select * from main.us_counties limit 5`

Data Connection: samples

Database:

main

Tables

- restaurant_scores_lives...
- retail_food_store_inspe...
- superstore_sales
- trips
- trips_detail
- us_counties
- world_life_expectancy

Autocomplete on

Add in a "LIMIT 100" clause to any SQL select query that does not have a limit clause

Number of rows in CSV:

select * from main.us_counties limit 5

sumlev	state	county	stname	ctyname	year	agegrp	tot_pop	tot_male	tot_female	wa_male	wa_female	ba_male	ba_female	ia_male	ia_female	aa_male
50	51	149	Virginia	Prince George County	5	0	36941	20368	16573	12155	10721	7230	4763	171	98	254
50	51	153	Virginia	Prince William County	5	0	430289	213820	216469	141918	138857	44291	47256	2453	2331	16249
50	51	155	Virginia	Pulaski County	5	0	34736	17284	17452	15915	16222	959	866	35	37	93
50	51	157	Virginia	Rappahannock County	5	0	7456	3694	3762	3420	3496	181	171	5	13	19
50	51	159	Virginia	Richmond County	5	0	9059	5066	3993	3138	2925	1799	961	24	15	31

Showing 1 - 5 of 5 rows



Note: If there is a error in the query the line number with the error will be displayed if the database returns it.

Downloading the results of a Direct Access query

After obtaining query results in the Direct Access interface, you can download the records in CSV format.

You can download the results in CSV format, by clicking **DOWNLOAD CSV**. The system saves the CSV file to your default download directory.

If you only want to download a subset of the query results, you can specify the Number of rows in CSV. This will limit the number of records in the CSV file.

Data Connection: samples

Database:

Enter SQL below: `select * from main.us_counties limit 5`

Tables: restaurant_scores_lives..., retail_food_store_inspe..., superstore_sales, trips, trips_detail, us_counties, world_life_expectancy

Add in a "LIMIT 100" clause to any SQL select query that does not have a limit clause

select * from main.us_counties limit 5

sumlev	state	county	stname	ctyname	year	agegrp	tot_pop	tot_male	tot_female	wa_male	wa_female	ba_male	ba_female	ia_male	ia_female	aa_male
50	51	149	Virginia	Prince George County	5	0	36941	20368	16573	12155	10721	7230	4763	171	98	254
50	51	153	Virginia	Prince William County	5	0	430289	213820	216469	141918	138857	44291	47256	2453	2331	16249
50	51	155	Virginia	Pulaski County	5	0	34736	17284	17452	15915	16222	959	866	35	37	93
50	51	157	Virginia	Rappahannock County	5	0	7456	3694	3762	3420	3496	181	171	5	13	19
50	51	159	Virginia	Richmond County	5	0	9059	5066	3993	3138	2925	1799	961	24	15	31

Showing 1 - 5 of 5 rows

The system saves the CSV file to your default download directory.