

Connecting to Data

Date published: 2020-10-30

Date modified: 2024-10-30

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

Connecting to a data source in Cloudera Data Visualization.....	4
Creating a Cloudera Machine Learning data connection to Impala.....	4
Creating a Cloudera Machine Learning data connection to a Hive data warehouse.....	12
Creating a Cloudera Data Warehouse data connection in Cloudera Data Visualization.....	18
Creating a Cloudera Data Science Workbench data connection to a data warehouse.....	21
Managing data connections.....	23
Editing a data connection.....	23
Setting concurrency for a data connection.....	26
Setting a warning if the query loads for too long.....	27
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
Adding data.....	36
Importing data in CSV format.....	37
Adding data through URL.....	40
Changing the locale setting of an imported file.....	43
Preparing Excel files for data import.....	44

Connecting to a data source in Cloudera Data Visualization

Cloudera Data Visualization allows you to create connections to many types of external data sources, enhancing your data analysis and visualization capabilities.

Cloudera Data Visualization currently supports the following connection types:

- Hive
- Impala
- MariaDB
- MySQL
- PostgreSQL
- Solr [Technical Preview]
- Spark SQL
- SQLite (not supported in CDW)
- Snowflake [Technical Preview]

This range of supported connection types provides flexibility and versatility in connecting to various data sources for comprehensive data analysis.

When using Cloudera Data Visualization with Cloudera Data Warehouse, the data connection is automatically set up, but you can connect to various other data sources as well to suit your specific requirements.

In Cloudera Machine Learning, you can connect to an Impala or a Hive data warehouse, or tie in data from predictive Cloudera Machine Learning models.

Related Information

[Connections](#)

[Datasets](#)

[Data modeling](#)

Creating a Cloudera Machine Learning data connection to Impala

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

About this task

Before you start using data modeling and visualization functions, you must connect to your data. The following steps show you how to create a new Cloudera Machine Learning data connection in Cloudera Data Visualization to an Impala data warehouse.



Note:

You must have the Manage data connections privilege or be an admin to be able to create and manage connections in Cloudera Data Visualization.

Setting user privileges requires administrator-level access. You can log in as an administrator using the default admin account with the following credentials:

- Username: vizapps_admin
- Password: vizapps_admin

When you create a connection, you automatically get privileges to create and manage the associated datasets. You can also build dashboards and visuals within these datasets.

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

Before you begin

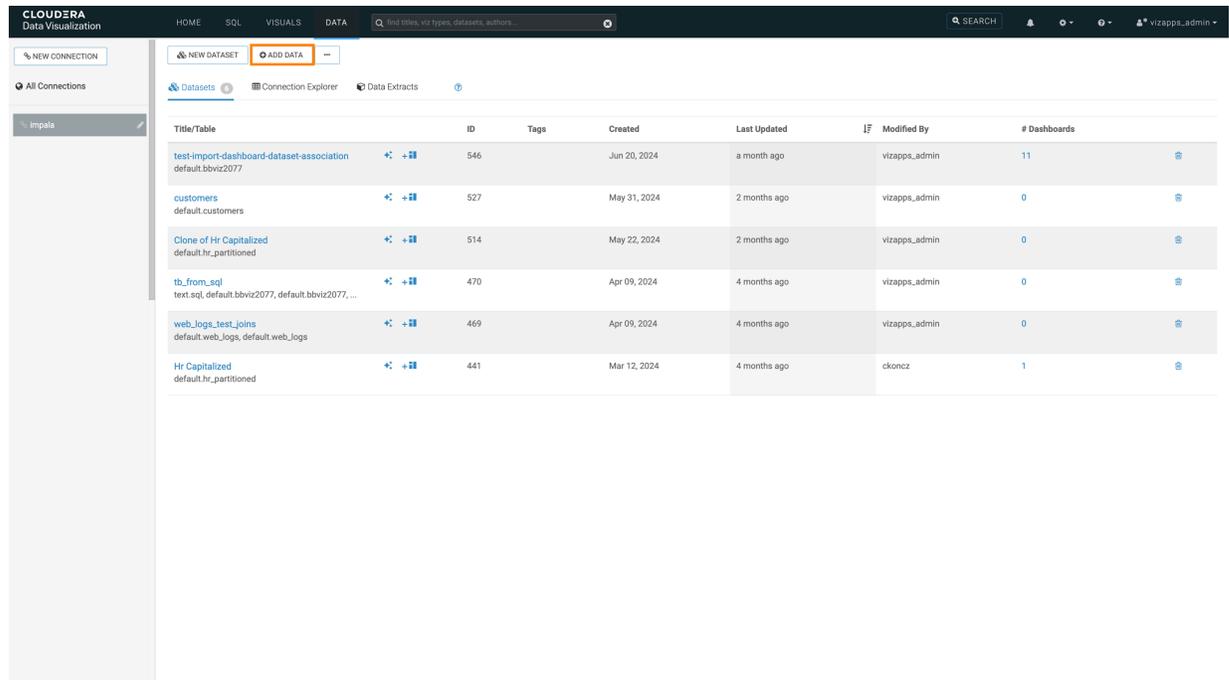
If you are using a Cloudera Private Cloud Base cluster running Impala with Kerberos for authentication, make sure that Kerberos credentials are configured in Cloudera Machine Learning before creating a Cloudera Machine Learning data connection to the Impala data warehouse. This ensures seamless integration and authentication between Cloudera Data Visualization and the Impala cluster. If you add Kerberos credentials after launching the Cloudera Data Visualization app, you need to restart the app for the changes to take effect.

For more information on using Kerberos for authentication in Cloudera Machine Learning, see *Hadoop Authentication for AI Workspaces*.

Procedure

1. On the main navigation bar, click DATA.

The DATA interface opens, displaying the Datasets tab.



Title/Table	ID	Tags	Created	Last Updated	JF	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago		vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago		vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago		vizapps_admin	0
tbl_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago		vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago		vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago		ckoncz	1

2. On the side menu bar, click NEW CONNECTION.



Note: Only users with Manage data connections privilege or administrators can access the NEW CONNECTION button.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. A search bar is present on the right. The left sidebar contains a 'NEW CONNECTION' button (highlighted with an orange box) and a list of connections, including '2050_CloneSamples'. The main content area displays a table of datasets with columns for 'Title/Table', 'ID', and 'Tags'. The first row is highlighted and includes an 'Extract Source' button.

Title/Table	ID	Tags
Cereals main.cereals Data Connection: samples	11	Extract Source
cereal main.cereals, main.chicago_govt_pay Data Connection: samples	191	
cereal main.cereals, main.chicago_govt_pay Data Connection: 2050_CloneSamples	341	
csimport Solr.csaba_import_20240119 Data Connection: SolrTestConnection	417	

The Create New Data Connection modal window appears.

3. Choose Impala from the Connection type drop-down list and assign a name to your connection.

Create New Data Connection

Connection type: Impala

Connection name: doc-test

Basic | Advanced | Parameters | Data

Hostname or IP address: Enter IP address of the server where your data resides (example: prod_db.yourcompany.com or 10.0.1.20)

Port #: 443

Credentials

Username: [Empty]

Password: [Masked]

TEST | CONNECT

In this example, the Impala connection is made through Knox. Knox always uses TLS encryption and port 443 is the default HTTPS port.

4. Enter the hostname or IP address of the running coordinator.
You can retrieve this information from the JDBC URL of the Impala DW.
5. Add 443 in the Port # field.
6. Enter your workload username and password as credentials.

- Click the Advanced tab to configure additional details.

Create New Data Connection
✕

Connection type

Connection name

Impala

doc-test

Basic
Advanced
Parameters
Data

Connection mode Binary HTTP

HTTP Path

Socket type Normal SSL SSL with certificate

Authentication mode NoSasl Plain LDAP Kerberos

Socket Timeout

Impersonation Enabled

Trusted Impersonation Enabled

Application Name

TEST

CONNECT

- For HTTP connection mode, 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;

- Copy and paste it into the HTTP Path field.

- c) Set any additional details as required.

- 8. Check the Parameters and Data tabs for more configuration options.

Create New Data Connection ✕

Connection type

Connection name

Basic Advanced **Parameters** Data

Parameter Name	Parameter Value	
Add new row		

Create New Data Connection ✕

Connection type

Connection name

Basic Advanced Parameters **Data**

Concurrency

Concurrency Per User

Query Timeout (Minutes)

Query Still Loading Warning (Seconds)

Row upload limit

Result Cache Enabled

Cache Retention Time (seconds)

9. Once you finish configuring the settings, click TEST to check the connection.

10. Click CONNECT to establish the connection.

Results

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

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

[Hadoop Authentication for ML Workspaces](#)

Creating a Cloudera Machine Learning data connection to a Hive data warehouse

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

About this task

Before you start using data modeling and visualization functions, you must connect to your data. The following steps show you how to create a new Cloudera Machine Learning data connection in Cloudera Data Visualization to a Hive data warehouse.



Note:

You must have the Manage data connections privilege or be an admin to be able to create and manage connections in Cloudera Data Visualization.

Setting user privileges requires administrator-level access. You can log in as an administrator, using the default admin account with the following credentials:

- Username: vizapps_admin
- Password: vizapps_admin

When you create a connection, you automatically get privileges to create and manage the associated datasets. You can also build dashboards and visuals within these datasets.

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

Before you begin

If you are using a Cloudera Private Cloud Base cluster running Hive with Kerberos for authentication, make sure that Kerberos credentials are configured in Cloudera Machine Learning before creating a Cloudera Machine Learning data connection to the Hive data warehouse. This ensures seamless integration and authentication between Cloudera Data Visualization and the Hive cluster. If you add Kerberos credentials after launching the Cloudera Data Visualization app, you need to restart the app for the changes to take effect.

Procedure

1. On the main navigation bar, click DATA.

The DATA opens, displaying the Datasets tab.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes HOME, SQL, VISUALS, and DATA. The 'DATA' tab is active, displaying a list of datasets. The left sidebar contains a 'NEW CONNECTION' button, which is highlighted with an orange box. Below the navigation bar, there are tabs for 'NEW DATASET', 'ADD DATA', and a minus sign. The main content area shows a table of datasets with columns: Title/Table, ID, Tags, Created, Last Updated, Modified By, and # Dashboards.

Title/Table	ID	Tags	Created	Last Updated	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago	vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago	vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago	vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago	vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago	vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago	ckoncz	1

2. On the side menu bar, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with the 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 'DATA' tab is active, displaying a list of datasets. The left sidebar contains a 'NEW CONNECTION' button, which is highlighted with an orange box. Below the navigation bar, there are tabs for 'NEW DATASET', 'ADD DATA', and a minus sign. The main content area shows a table of datasets with columns: 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.

3. Choose Hive from the Connection type drop-down list and assign a name to your connection.

Create New Data Connection ✕

Connection type

Connection name

Basic **Advanced** **Parameters** **Data**

Hostname or IP address

Port #

Credentials

Username

Password

4. Enter the hostname or IP address of the running coordinator.
You can get the coordinator hostname from the JDBC URL of the Hive DW.
5. Use port 443.
6. Enter your workload username and password as credentials.

7. Click the Advanced tab to configure the additional details.

Create New Data Connection ✕

Connection type

Connection name

Basic **Advanced** Parameters Data

Connection mode Binary HTTP

HTTP Path

Access Token

Socket type Normal SSL SSL with certificate

Authentication mode NoSasl Plain LDAP Kerberos

Socket Timeout

Impersonation Enabled

Trusted Impersonation Enabled

Application Name

- 8. 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** Data

Parameter Name	Parameter Value	
hive.server2.async.exec.async.compile	false	
Add new row		

9. Check the Data tab for more configuration options.

Create New Data Connection ✕

Connection type

Connection name

Basic Advanced Parameters **Data**

Concurrency

Concurrency Per User

Query Timeout (Minutes)

Query Still Loading Warning (Seconds)

Row upload limit

Result Cache Enabled

Cache Retention Time (seconds)

10. Once you finish configuring the settings, click TEST to test the connection.

11. Click CONNECT to establish the connection.

Results

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

Related Information

[RBAC permissions](#)

[Setting role privileges](#)

[Promoting a user to administrator](#)

Creating a Cloudera Data Warehouse data connection in Cloudera Data Visualization

Learn how to connect to data when using Cloudera Data Visualization in Cloudera Data Warehouse data service. You can connect Cloudera 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 Cloudera Data Warehouse URL has the same compute instance ID as your Virtual Warehouse.



Note:

To create and manage connections in Cloudera Data Visualization, you must have the Manage data connections privilege or hold administrative privileges. In Cloudera Data Warehouse, these are the members of the Admin Groups associated with the Cloudera Data Visualization instance.

When you create a connection, you automatically gain privileges to create and manage datasets associated with this connection, and to build dashboards and visuals within these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a specific 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 Cloudera Data Visualization from the left navigation panel in Cloudera Data Warehouse.
2. On the main navigation bar, click DATA.

The DATA interface appears, open on the Datasets tab.

Title/Table	ID	Tags	Created	Last Updated	IF	Modified By	# Dashboards
test-import-dash-board-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago		vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago		vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago		vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago		vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago		vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago		ckoncz	1

3. In the side menu bar of DATA, click NEW CONNECTION.



Note: The NEW CONNECTION button is only accessible to users assigned to roles with the 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 'DATA' tab is active, and a search bar is present. 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 displays a table of datasets with columns for 'Title/Table', 'ID', and 'Created'. The table lists several datasets, including 'Test Dataset', 'Food Stores Inspection in NYC', 'Cereals', 'Earthquake Data January 2019', and 'World Life Expectancy'.

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.

4. In Connection type, select CDW Hive or CDW Impala, and provide a name for your connection.



Note: SQLite connection is not supported in Cloudera Data Warehouse, and the option is disabled in the connection list.

Create New Data Connection ✕

Connection type

Connection name

Basic **Advanced** **Parameters** **Data**

Hostname or IP address

Port #

Credentials

Username

Password

5. Select a Cloudera Data Warehouse 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

6. Click the Advanced tab and configure the additional details.



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.

7. Click TEST.

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

8. 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 Cloudera Data Science Workbench data connection to a data warehouse

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

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 Cloudera Data Science Workbench data connection to a running Impala system.



Note:

To create and manage connections in Cloudera Data Visualization, you must have the Manage data connections privilege or hold administrative privileges.

Setting user privileges requires administrator-level access. To log in as an administrator, you can use the default admin account with the following credentials:

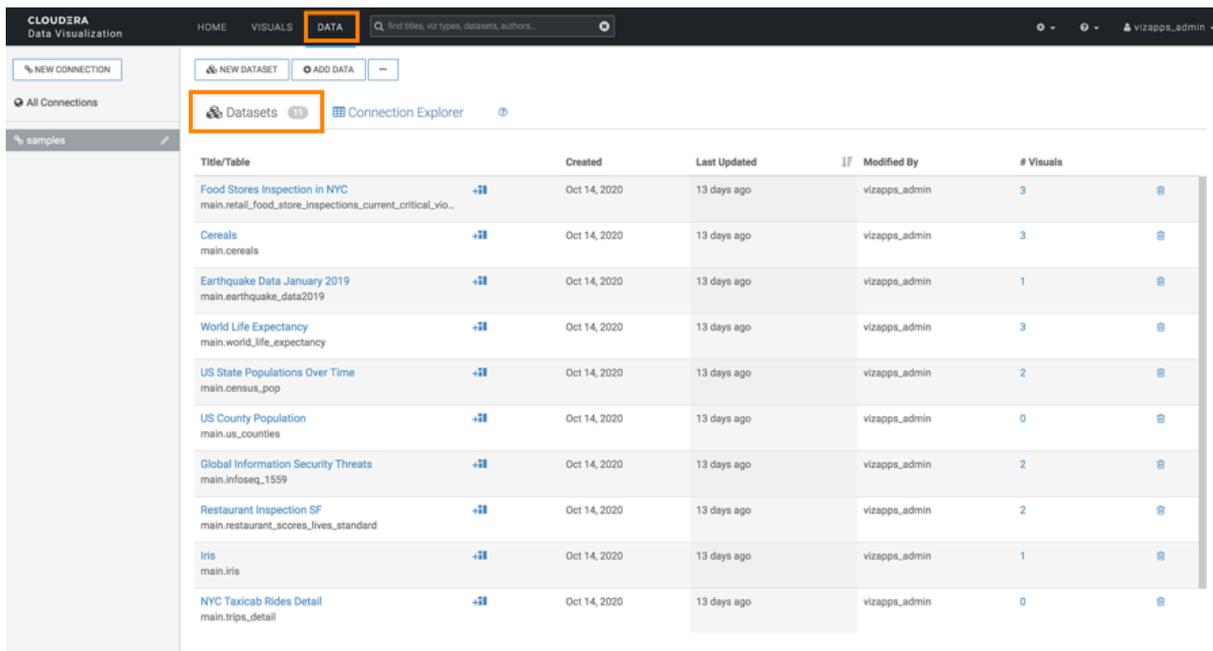
- Username: vizapps_admin
- Password: vizapps_admin

When you create a connection, you automatically gain privileges to create and manage datasets associated with this connection, and to build dashboards and visuals within these datasets.

- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a specific 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.



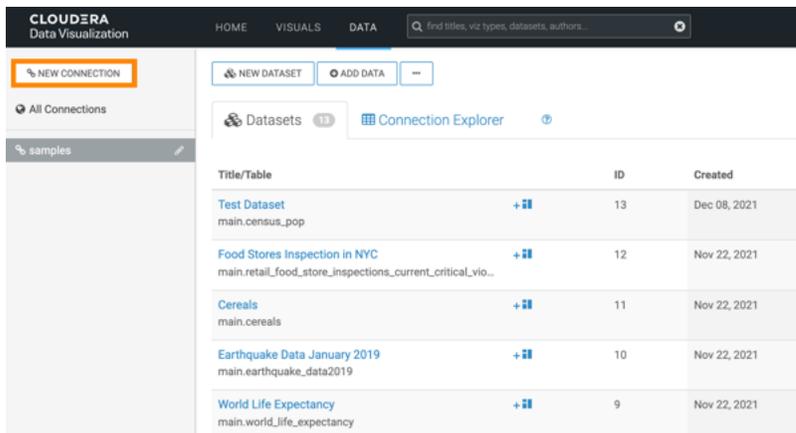
The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA' (highlighted with an orange box). Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'Datasets' tab is selected and highlighted with an orange box. The main content area displays a table of datasets with the following columns: 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 the Manage data connections privilege and to administrators.



The screenshot shows the Cloudera Data Visualization interface with the 'NEW CONNECTION' button highlighted in the side menu. The main content area displays a table of datasets with the following columns: 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.

3. Select a Connection type from the drop-down list, and provide a name for the connection.
4. Enter the hostname or IP address of the running coordinator.
5. Under Port #, enter the port number.
6. Use your workload username and password as credentials.
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.

8. Click TEST.

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

- 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](#)

Managing data connections

You can change the properties of existing data connections and you can also delete connections that are not associated with any dataset.

Editing a data connection

Learn how to edit a data connection in Cloudera 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

- On the main navigation bar, click DATA.

The Data view appears, open on the Datasets tab.

Title/Table	ID	Tags	Created	Last Updated	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago	vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago	vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago	vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago	vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago	vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago	ckoncz	1

- 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

Connection Verified! ×

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

Results

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

Setting concurrency for a data connection

You can improve resource management in Cloudera Data Visualization by restricting the number of simultaneous connections on specific datasets.

About this task



Note: This setting is only applicable to Impala, Hive, and SQLite connections.

Procedure

1. Open the Edit Data Connection modal.
For more information, see *Editing a data connection*.
2. Click the Data tab.
3. In the Concurrency field, provide a number to set the maximum limit for simultaneous requests.



Note: The default maximum concurrency setting per connection is 100. You can change it to a different value in Site Settings Advanced Settings . If you change the default value, ensure that the new maximum concurrency aligns with the requirements of your data connection.

```
DATACONNECTION_CONCURRENCY_DEFAULT = 100  
DATACONNECTION_CONCURRENCY_MAX = 100
```

4. In the Concurrency Per User field, provide a number to limit the maximum number of simultaneous requests a user can send.



Note: You can use this setting to prevent a single user from blocking access for a connection. The default concurrency setting per user is 5, with a default maximum concurrency setting per user at 100. You can adjust these values in Site Settings Advanced Settings .

```
DATACONNECTION_CONCURRENCY_USER_DEFAULT = 5  
DATACONNECTION_CONCURRENCY_USER_MAX = 100
```

5. Click SAVE to apply the changes.

Edit Data Connection ✕

Connection type SQLite ▾

Connection name samples

Basic **Data**

Concurrency ⓘ 30

Concurrency Per User ⓘ 5

Query Timeout (Minutes) ⓘ 50

Query Still Loading Warning (Seconds) 1

Row upload limit ⓘ 10000

Result Cache Enabled

Cache Retention Time (seconds) 8000000

TEST **CLONE** **DELETE CONNECTION** **SAVE**

Related Information

[Editing a data connection](#)

Setting a warning if the query loads for too long

A data connection admin user can set the time limit for every connection before a warning is shown when data is loaded through a query.

About this task

For more information on creating or editing data connections, see *Connection to a data source* and *Editing a data connection*.

Procedure

1. Select the Data tab in the Data Connection modal.
2. Enter the number of seconds you want to pass before the warning appears in Query Still Loading Warning (Seconds).

The screenshot shows the 'Data' tab of a configuration modal. It contains several settings, each with a label, an information icon, and a value field:

Basic	Advanced	Parameters	Data
			Concurrency ⓘ
			10
			Concurrency Per User ⓘ
			5
			Query Timeout (Minutes) ⓘ
			10
			Query Still Loading Warning (Seconds) ⓘ
			20
			Row upload limit ⓘ
			1000000
			Result Cache <input type="checkbox"/> Enabled
			Cache Retention Time (seconds)
			86400

It is also possible to have this message appear when viewing a dashboard, see *Dashboards*.

Related Information

[Dashboards](#)

[Connecting to a data source in Cloudera Data Visualization](#)

[Editing a data connection](#)

Deleting a data connection

Learn how you can remove an existing data connection in Cloudera Data Visualization.

About this task



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	Tags	Created	Last Updated	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago	vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago	vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago	vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago	vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago	vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago	ckoncz	1

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 the 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 the 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

Cloudera 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 Cloudera 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.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The 'DATA' tab is active. Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. The 'ADD DATA' button is highlighted with an orange box. The main area displays a table of datasets for the 'impala' connection.

Title/Table	ID	Tags	Created	Last Updated	IF	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago		vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago		vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago		vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago		vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago		vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago		ckoncz	1

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

The screenshot shows the Cloudera Data Visualization interface with the 'Connection Explorer' tab selected. The 'ADD DATA' button is still highlighted with an orange box. The main area displays a table of datasets for the 'samples' connection.

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, where you can explore the databases and tables available on the connection and you can also manage all functions related to the data on the connection. 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, there's a sidebar with 'samples' selected. The main area shows 'main' selected, displaying a table of datasets. The table has columns for 'Table Name' and '# Datasets'. The 'generalTestForExploreCall' table is highlighted. Below the table, there's a 'Sample Data' section with a table showing columns: id, name, age, cost, debt, altname, favoriteanimal, and heartrate. The data rows are as follows:

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 with the supplemental menu open. The menu options are: Create Analytical View, Clear result cache, Import Visual Artifacts, and Direct Access. The main area shows the same dataset table as in the previous screenshot.

- 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 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.
 - a. For Impala connections, clicking the Clear result cache option under the supplemental menu reloads the full table definition.
 - b. Import Visual Artifacts option under the supplemental menu enables you to restore or import visual artifacts from a backup JSON file.
 - c. 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.

- 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.
- In the list of tables, the # Datasets column lists the number of datasets that use the particular table as their initial definition.
- 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 shows a tree view of databases, with 'main' selected. Below the tree, a table lists tables and their corresponding number of datasets. The 'iris' table is highlighted, and a 'Sample Data' view is open below it, showing a table with columns: id, name, age, cost, debt, altname, favoriteanimal, and heartrate. The data rows are:

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

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.

Table Name	# Datasets	
census_pop	1	New dataset
cereals	1	New dataset
chicago_govt_pay	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

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

Using the Direct Access interface

The Direct Access interface of Cloudera 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 Cloudera Data Visualization.

Procedure

1. On the main navigation bar, click SQL.

Alternatively, you can click DATA **☰** > Direct Access .

2. Select the database in the left-side panel.

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.

4. Click RUN to initiate the SQL query.

After the query runs, the results tab shows the query results.

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

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

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.

For more information about the content available on the other tabs, see *Cloudera Data Visualization, SQL interface*.

Related Information

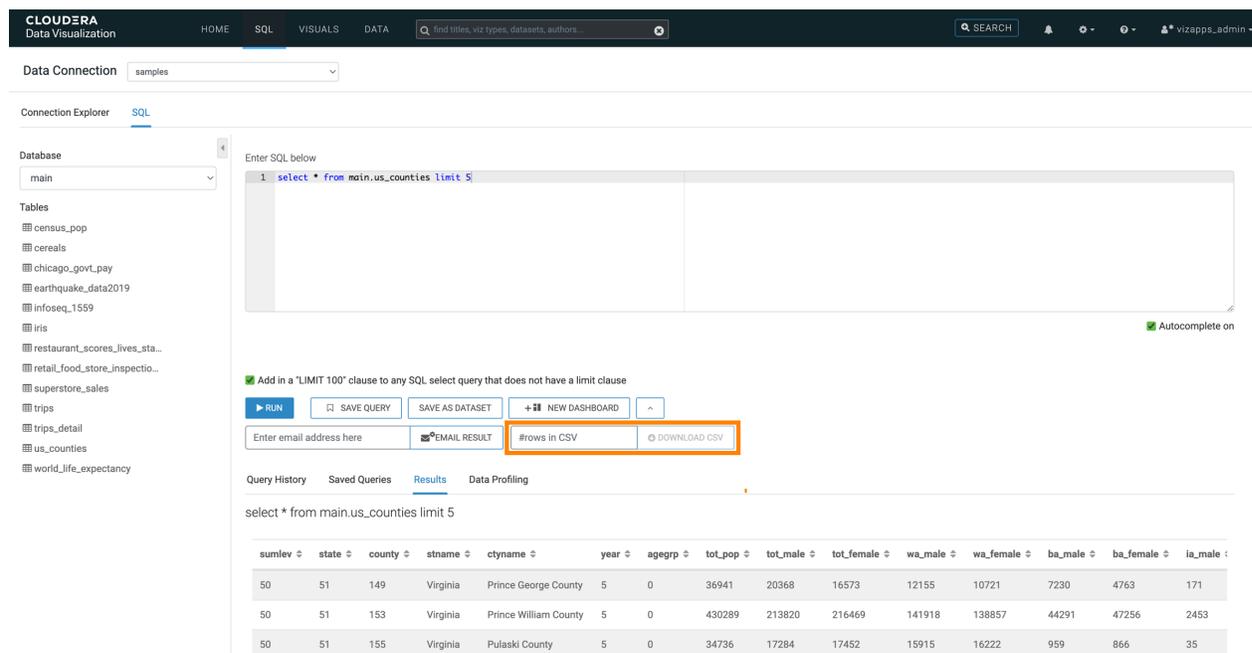
[Cloudera Data Visualization, SQL interface](#)

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.



The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The 'Data Connection' is set to 'samples'. The 'Connection Explorer' shows the 'SQL' tab. The 'Database' is set to 'main'. The 'Tables' list includes 'census_pop', 'cereals', 'chicago_govt_pay', 'earthquake_data2019', 'infoseq_1559', 'iris', 'restaurant_scores_lives_sta...', 'retail_food_store_inspectio...', 'superstore_sales', 'trips', 'trips_detail', 'us_counties', and 'world_life_expectancy'. The 'Enter SQL below' field contains the query: `1 select * from main.us_counties limit 5`. Below the query field, there are buttons for 'RUN', 'SAVE QUERY', 'SAVE AS DATASET', and 'NEW DASHBOARD'. There is also an input field for 'Enter email address here' and a button for 'EMAIL RESULT'. The 'DOWNLOAD CSV' button is highlighted with an orange box. Below the query execution area, there is a 'Query History' section with tabs for 'Saved Queries', 'Results', and 'Data Profiling'. The 'Results' tab is active, showing the query: `select * from main.us_counties limit 5`. The results are displayed in a table with the following columns: sumlev, state, county, stname, ctyname, year, agegrp, tot_pop, tot_male, tot_female, wa_male, wa_female, ba_male, ba_female, and ia_male. The table contains three rows of data.

sumlev	state	county	stname	ctyname	year	agegrp	tot_pop	tot_male	tot_female	wa_male	wa_female	ba_male	ba_female	ia_male
50	51	149	Virginia	Prince George County	5	0	36941	20368	16573	12155	10721	7230	4763	171
50	51	153	Virginia	Prince William County	5	0	430289	213820	216469	141918	138857	44291	47256	2453
50	51	155	Virginia	Pulaski County	5	0	34736	17284	17452	15915	16222	959	866	35

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

Adding data

Experience the flexibility of importing and integrating data to enhance your data visualization projects. In Cloudera Data Visualization, you have the capability to add data from flat files. The imported data is processed as a new table and it is integrated into an existing data repository. Subsequently, you can use this data as a supplementary information source for your existing datasets by establishing new table joins. You can also create new datasets that reference the imported data as their primary tables.

Data import is available for the following connections:

- Hive
- Impala
- MariaDB
- MySQL
- PostgreSQL
- Spark SQL



Note: Data import is not supported in Cloudera Data Warehouse as the connection is read-only.

- SQLite



Note: This connection is not supported in Cloudera Data Warehouse.

- Snowflake [Technical Preview]

Cloudera Data Visualization supports two primary forms of data sources: CSV and URL.

You can use the Import Data functionality to enhance your existing datasets, incorporating this data by creating joins within the data model. The choice of creating a new dataset based on the imported data depends on your specific business requirements.

For importing data stored in Excel format, see *Preparing Excel files for data import*.

Importing data in CSV format

Learn how to add data to Cloudera Data Visualization from flat CSV files.

About this task

You can use comma-delimited files or files with other delimiting characters. If you want to import data stored in Excel format, see *Preparing Excel files for data import*.

This feature is available on the following connections:

Enabled by default:	Disabled by default:
<ul style="list-style-type: none"> • SQLite • Hive • Impala • Snowflake [Technical Preview] • Solr [Technical Preview] • Spark SQL 	<ul style="list-style-type: none"> • DuckDB • MS SQL • Teradata

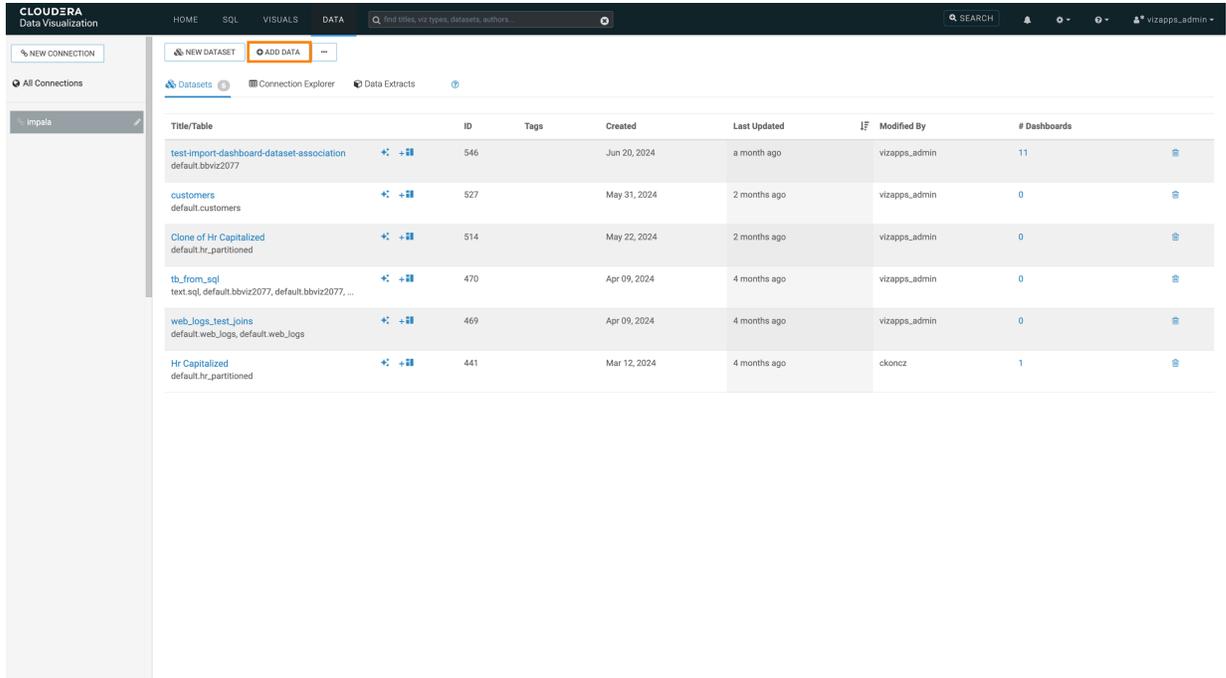


Note: To import data into Cloudera Data Visualization, ensure that you use a data connection with write permissions.

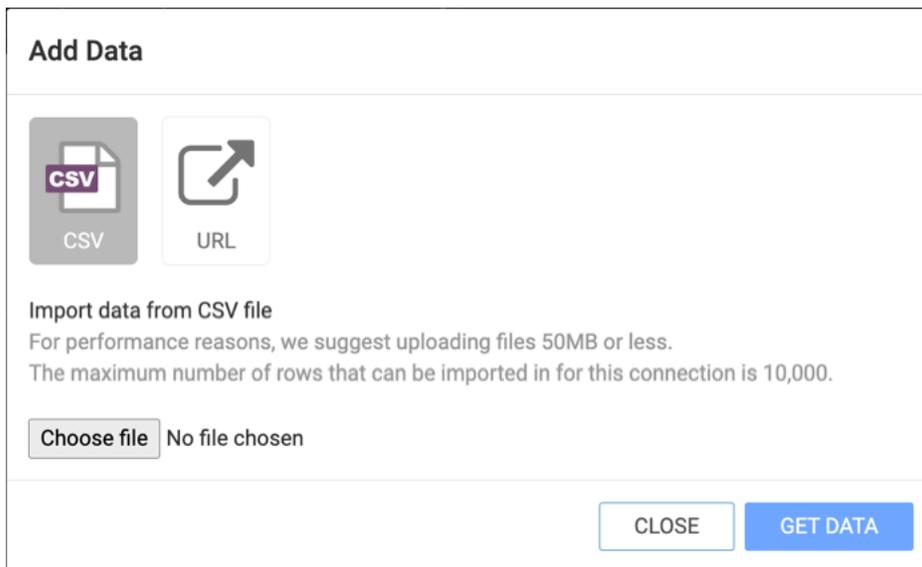
Procedure

1. On the main navigation bar, click DATA.
The Data view appears, showing the Datasets tab.

2. Click ADD DATA.



The Add Data modal window appears.



3. Add the CSV file that contains the data you want to import.
 - a) Select the CSV option and click Choose file.
 - b) Using your computer's file browser, locate and select the data file.
 - c) Click GET DATA.

4. In the Imported File modal window, ensure that the data you are adding is configured correctly.
 - a) Verify the name of the file.
 - b) Under Database, select which database you want to add data to.
 - c) Under Table Name, specify a table. The system usually assigns a numerical suffix, which you can remove.
 - d) Under Upon Import, select whether you only want to import data or create a dataset, or create a dataset and a dashboard based on the data.
 - If you select Create Dataset, you are directed to the Detail page of the dataset you have created after the import is completed.
 - If you select Create Dataset and Dashboard, you are directed to the newly created dashboard when the import is completed.
 - e) For Column Delimiter, define the delimiter used in the source file. It can be Comma, Tab, Space, Semicolon, Colon, Pipe, Control A, or Other.
 - f) Under Locale Setting, you can select one of the common locales: United States, United Kingdom, Sweden, Norway. Or you can select Other to add a separate field next to the locale setting, where you can specify the supported locale. For more information, see *Changing the locale setting of an imported file*.
 - g) Under Options, select the data clean-up tasks that Cloudera Data Visualization provides prior to data import. These include Create "EXTERNAL" table, File contains headers, Fill missing columns, Skip malformed rows, and Use '\ ' as escape character.



Note: Cloudera Data Visualization can detect many items, for example the presence of header rows, automatically.

- h) In the Data Table preview, each column of the table appears with its data type and name. The data types available are Boolean, Integer, Real, String, and Timestamp.

Cloudera Data Visualization detects the data type automatically, but you can modify them as needed (for example store numerical categories as strings).

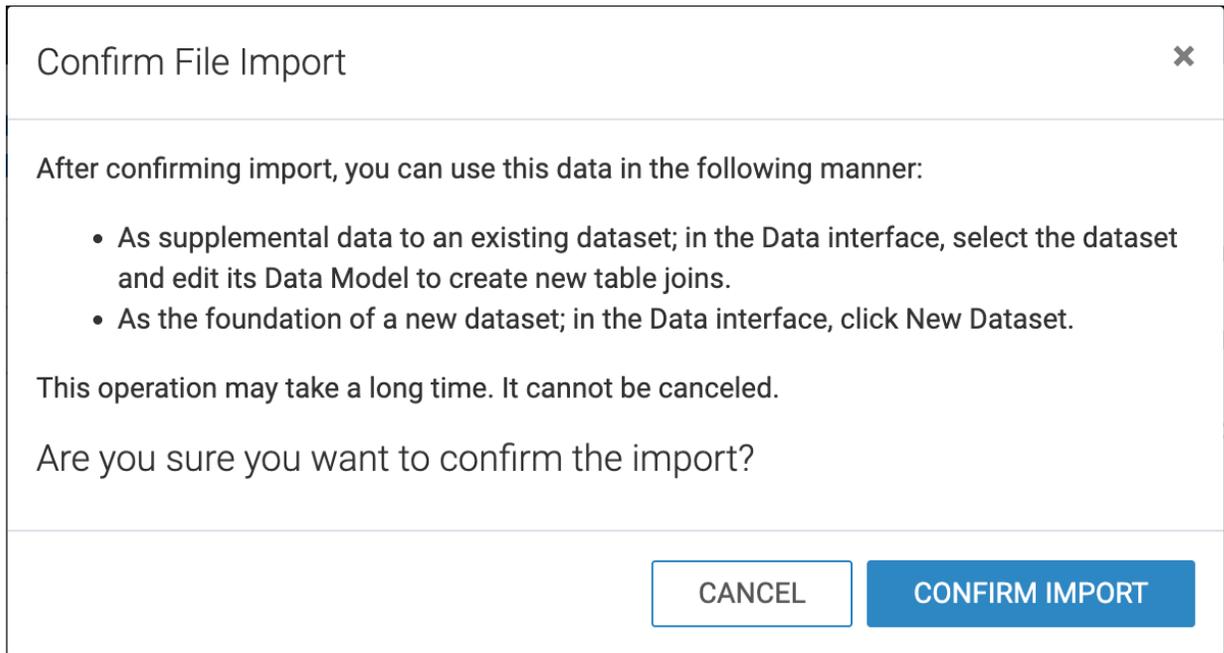
You can change the name of the table column before importing.

- i) Click APPLY CHANGES to save the adjustments you have made.

#	year	industry_aggr	industry_code	industry_name	units	variable_code	variable_name	variable_category	value	industry_code
2021	Level 1	99999	All industries	Dollars (millions)	H01	Total income	Financial performance	757,504	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H04	Sales, government funding, grants and subsidies	Financial performance	674,890	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	49,593	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H07	Non-operating income	Financial performance	33,020	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H08	Total expenditure	Financial performance	654,404	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H09	Interest and donations	Financial performance	26,138	ANZSIC06 divisions A-S (excluding classes K6330, L6711, 07552, 0760, 0771, 0772, S9540, S9601, S9602, and S9603)	
2021	Level 1	99999	All industries	Dollars (millions)	H10	Indirect taxes	Financial performance	6,991	ANZSIC06 divisions A-S (excluding	

5. Click CONFIRM IMPORT.

The Confirm File Import modal window appears, where you have to click CONFIRM IMPORT again to finalize the import.



6. To verify that the data has been imported correctly into your system, check the Connection Explorer interface.

Related Information

[Preparing Excel files for data import](#)

[Changing the locale setting of an imported file](#)

Adding data through URL

Learn how to how to add a data file using a URL in Cloudera Data Visualization.

About this task

Data files are typically in JSON format. The files may be comma-delimited or they use other delimiting characters.



Note: This feature is available on the following connections:

Enabled by default:

- SQLite
- Hive
- Impala
- Snowflake
- Spark SQL

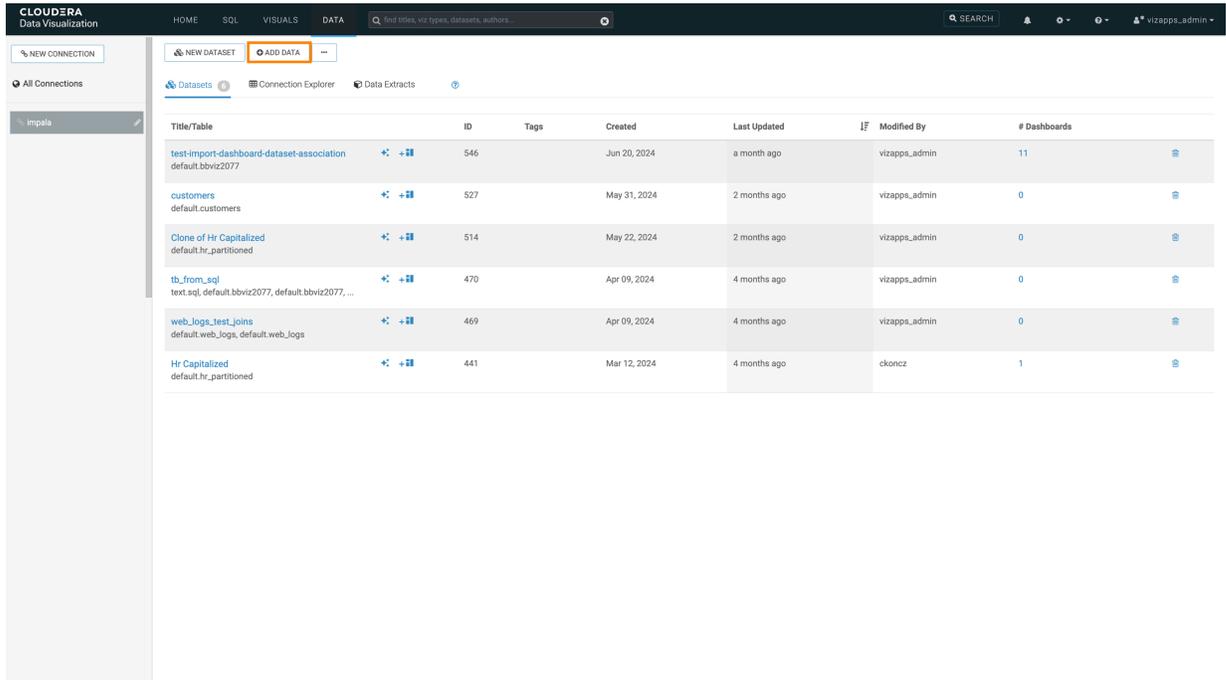
Disabled by default:

- DuckDB
- MS SQL
- Teradata

Procedure

1. On the main navigation bar, click DATA.

The Data view appears, open on the Datasets tab.

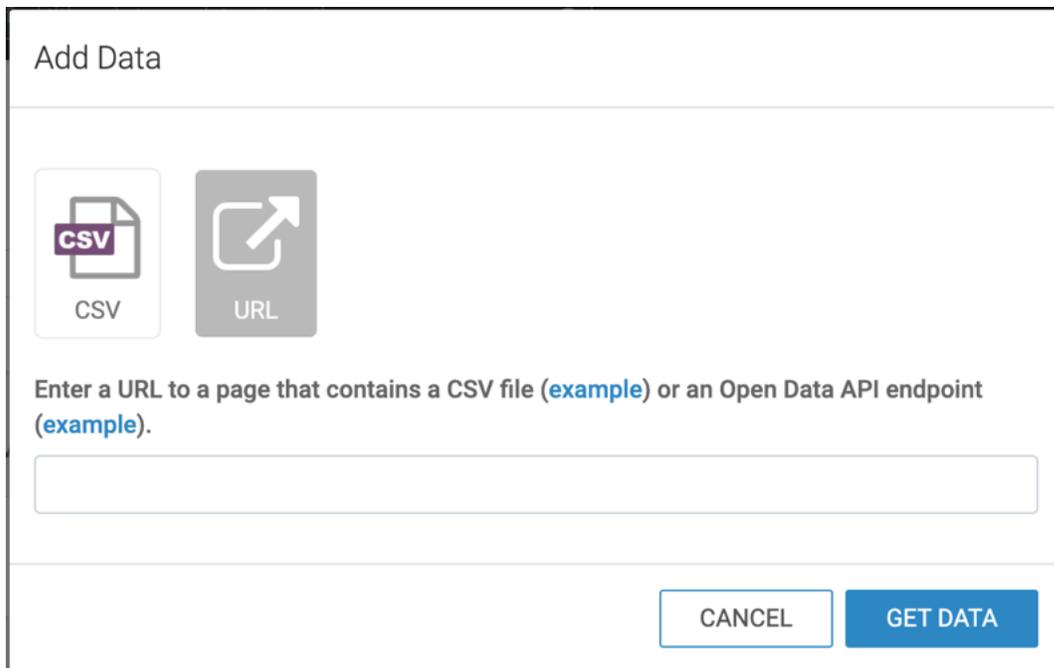


The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes HOME, SQL, VISUALS, and DATA. The 'DATA' tab is active, and the 'ADD DATA' button is highlighted with a red box. Below the navigation bar, there are tabs for 'Datasets', 'Connection Explorer', and 'Data Extracts'. The 'Datasets' tab is selected, displaying a table of datasets. The table has columns for Title/Table, ID, Tags, Created, Last Updated, JF, Modified By, and # Dashboards. The datasets listed are:

Title/Table	ID	Tags	Created	Last Updated	JF	Modified By	# Dashboards
test-import-dashboard-dataset-association default.bbviz2077	546		Jun 20, 2024	a month ago		vizapps_admin	11
customers default.customers	527		May 31, 2024	2 months ago		vizapps_admin	0
Clone of Hr Capitalized default.hr_partitioned	514		May 22, 2024	2 months ago		vizapps_admin	0
tb_from_sql text.sql, default.bbviz2077, default.bbviz2077, ...	470		Apr 09, 2024	4 months ago		vizapps_admin	0
web_logs_test_joins default.web_logs, default.web_logs	469		Apr 09, 2024	4 months ago		vizapps_admin	0
Hr Capitalized default.hr_partitioned	441		Mar 12, 2024	4 months ago		ckoncz	1

2. Click ADD DATA.

The Add Data modal window appears.



The 'Add Data' modal window is displayed. It features two main options: 'CSV' and 'URL'. Below these options, there is a text input field for entering a URL. At the bottom of the modal, there are two buttons: 'CANCEL' and 'GET DATA'.

3. Select the URL option, and add the address of the file to the text window, and click GET DATA.

4. Ensure that your data is configured correctly before confirming the import:
 - a. Under Database, specify documentation. Alternatively, select another database.
 - b. Under Table Name, specify city_chicago_staff. The system typically assigns a numerical suffix that you can remove.
 - c. Under Upon Import, select whether you only want to import the data or create a dataset and a dashboard based on the data.
 - If you select Create Dataset, you are taken to the Detail page of the dataset you have created when the import is completed.
 - If you select Create Dataset and Dashboard, you are taken to the newly created dashboard when the import is completed.
 - d. For Column Delimiter, define the delimiter used in the source file. In this case, the column delimiter is Comma, but it can also be Tab, Space, Semicolon, Colon, Pipe, Control A, or Other. Make the necessary adjustments to this field.
 - e. Under Locale Setting, you can select one of the common options: United States, United Kingdom, Sweden, Norway, or Other. Selecting Other adds a separate field next to the locale setting where you can specify the supported locale. For more information, see *Changing the locale setting of an imported file*.
 - f. Under Options, select the data clean-up tasks that Cloudera Data Visualization provides prior to data import. These include File contains headers, Fill missing columns, Skip malformed rows, and Use '\ ' as escape character.



Note: Cloudera Data Visualization can detect many items automatically, such as the presence of header rows.

- g. In the Data Table preview, each column of the table appears with its data type and name. Cloudera Data Visualization detects the data type automatically, but you can change it in this interface prior to import. For example, you may wish to store numerical categories as strings. The data types are Boolean, Integer, Real, String, and Timestamp.
- h. You can change the name of the table column before importing.

Imported File: https://en.wikipedia.org/wiki/Food_safety_in_the_United_States APPLY CHANGES CONFIRM IMPORT

Database: Column Delimiter:

Table Name: Locale Setting:

Upon Import: Import only
 Create Dataset
 Create Dataset and Dashboard

Options: File contains headers
 Fill missing columns
 Skip malformed rows
 Use '\ ' as escape character

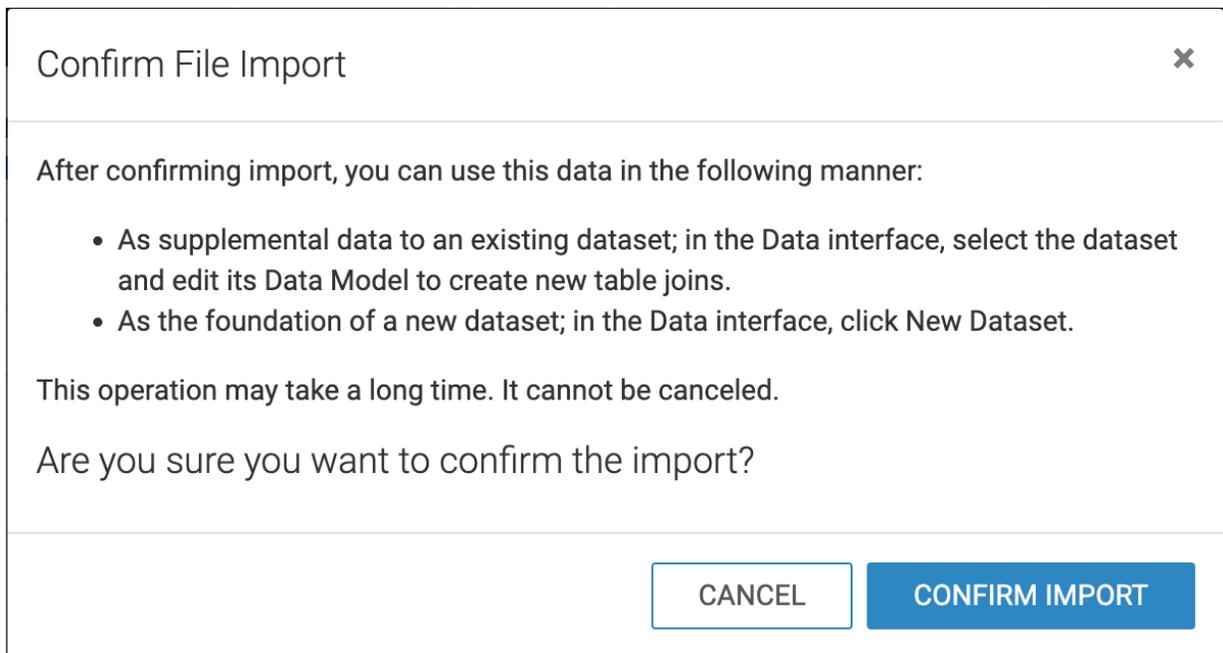
Data Table Rows 1-100

column_0	
<!DOCTYPE html>	
<html class="client-nojs vector-feature-language-in-header-enabled vector-feature-language-in-main-page-header-disabled vector-feature-language-alert-in-sidebar-enabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-pinned-enabled vector-feature-main-menu-pinned-disabled vector-feature-limited-width-enabled vector-feature-limited-width-content-enabled vector-feature-zebra-design-disabled" lang="en" dir="ltr">	
<head>	
<meta charset="UTF-8"/>	
<title>Food safety in the United States - Wikipedia</title>	
<script>document.documentElement.className="client-js vector-feature-language-in-header-enabled vector-feature-language-in-main-page-header-disabled vector-feature-language-alert-in-sidebar-enabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-pinned-enabled vector-feature-main-menu-pinned-disabled vector-feature-limited-width-enabled vector-feature-limited-width-content-enabled vector-feature-zebra-design-disabled";(function(){var cookie=document.cookie.match(/(?:\s ,)=wikimwclientprefs=[^"]+\$/);if(cookie){var featureName="cookie[1]";document.documentElement.className=document.documentElement.className.replace(featureName+"-enabled"	featureName+"-dir ();RLCONF= ("wgBreakFrames
1ab72189-b978-4f03-8ffb-facd76151589	wgCSPNonce: false
wgWMEschemaEditAttemptStepOversample: false	wgWMEPageLength
ext.visualEditor.desktopArticleTarget.noscript:"ready"	ext.wikimediaBad
<script>(RLQ=window.RLQ []).push(function(){mw.loader.implement("user.options@12a5f"	function(\$
<link rel="stylesheet" href="/w/load.php?lang=en&modules=ext.cite.styles%7Cext.visualEditor.desktopArticleTarget.noscript%7Cext.wikimediaBadges%7Cmediawiki.ui.button%2Cicon%7Cskins.vector.icons%2Cstyles%7Cwikibase.client.init&only=styles&skin=vector"	

5. Click APPLY CHANGES to ensure that all adjustments you made are saved.

6. Click CONFIRM IMPORT.

The Confirm File Import modal window appears, where you have to click CONFIRM IMPORT again.

**7. To verify that the data has been imported correctly into your system, check the Connection Explorer interface.****Related Information**

[Changing the locale setting of an imported file](#)

Changing the locale setting of an imported file

When importing a data file, you can specify the localization for the data. This will specifically handle differences in specifying floating-point number.

Under Locale Setting, which is the default for your system, select one of the common options: United States, United Kingdom, Sweden, Norway, or Other.

Selecting Other brings adds a new text entry box. Clicking on the information icon opens a separate menu, where you can select another supported locale. In our case, we selected Russian, ru_RU.

Column Delimiter

Locale Setting ?

Options

- File contains headers
- Fill missing columns
- Skip malformed rows
- Use '\' as escape character

Country	Code
Bulgarian	bg_BG
Chinese	zh_CN
Croatian	hr_HR
Czech	cs_CZ
Danish	da_DK
Dutch	nl_NL
English	en_US
Estonian	et_EE
Finnish	fi_FI
French	fr_FR
German	de_DE
Greek	el_GR
Hungarian	hu_HU
Italian	it_IT
Latvian	lv_LV
Lithuanian	lt_LT
Norwegian	no_NO
Polish	pl_PL
Portuguese	pt_PT
Romanian	ro_RO
Russian	ru_RU
Slovak	sk_SK
Slovenian	sl_SI
Spanish	es_ES
Swedish	sv_SE
Turkish	tr_TR

column_7	column_8	column_9	column_10
Issued	Submitted to tekom	Payment de	
Jun 30	Jul 12	N/A	
Sep 16	Sep 19	Sep 20	
Sep 29	Sep 29	Oct 7	
Oct 18	Oct 18	Oct 24	
Oct 19	Oct 20	Oct 27	
Oct 27	Oct 27	Oct 27	

Preparing Excel files for data import

Learn how you can import data from Microsoft Excel or other similar formats into Cloudera Data Visualization.

About this task

You must save the files in character-delimited format.

Procedure

1. In Microsoft Excel, open the file that contains the data you are importing.
2. Click Save as, then select Comma Separated Values (.csv).
3. Choose a file name, such as MyCSVdata, and click Save.