

CDP Data Visualization Top Tasks

Date published: 2020-10-30

Date modified: 2024-04-29

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

Deploying Data Visualization in CDP.....	4
Deploying Data Visualization in CDW.....	4
Deploying a CDV application in CML.....	7
Creating a CML project with CDV Runtime.....	8
Adding CDV Runtime to a CML project.....	8
Creating CDV application in CML.....	9
Deploying a CDV application in CDSW.....	9
Creating a CDSW project with CDV Runtime.....	10
Creating Data Visualization application in CDSW.....	10
Starting Data Visualization in CDSW.....	11
 Connecting to a data source in Cloudera Data Visualization.....	 12
Creating a CML data connection to Impala.....	12
Creating a CML data connection to a Hive data warehouse.....	20
Creating a CDW data connection in Data Visualization.....	26
Creating a CDSW data connection to a data warehouse.....	29
 Creating a visual.....	 31

Deploying Data Visualization in CDP

The first step to start using Cloudera Data Visualization (CDV) in Cloudera Data Platform (CDP) is to deploy it on your chosen platform. CDV provides consistent visualization capabilities across various CDP deployments and services. Once your CDV instance is up and running, you can start creating visuals, build customized dashboards, and develop analytical applications to enable users to explore data across the whole CDP data lifecycle.

Data Visualization is integrated with CDP, allowing you to log in using your CDP credentials and start analyzing data that you have access to. In the Public Cloud, Data Visualization is included as a core capability in Cloudera Data Warehouse (CDW) and Cloudera Machine Learning (CML) both for Azure and AWS. Data Visualization is also available in CDP Private Cloud for Data Warehouse and Machine Learning, and it can be used in Cloudera Data Science Workbench (CDSW), certified with CDH, HDP and CDP Private Cloud Base.

Deploying Data Visualization in CDW

Learn how to create a Cloudera Data Visualization (CDV) instance in the Cloudera Data Warehouse (CDW) data service and how to start the CDV application. You can use CDV with CDW either in CDP Public Cloud or with the CDP Private Cloud Data Service to explore the data stored in the database catalogs and use visualization to communicate insights across the whole data lifecycle.

About this task

If you want to create visuals based data stored in CDW, you have to create a CDV instance and connect it to Hive or Impala Virtual Warehouse(s).

Data Visualization is not tied to a particular Virtual Warehouse (VW). You can access data for your visualizations from multiple Data Catalogs using multiple Hive or Impala Virtual Warehouses in various environments. With multiple Data Visualization instances attached to an environment, you can connect to different data sets, create dashboards for different groups and share your visualizations with different users.



Note: When you delete a Virtual Warehouse, your visual artifacts remain intact as long as the CDV instance is not deleted.

Before you begin

- You are logged into the CDP web interface and you have opened the Data Warehouse service
- You have DWAdmin role in CDW
- You have activated your environment
- You have a Hive/Impala warehouse in running state
- If you are using CDV in CDW Private Cloud: You have an admin group created in Management Console User Management . To log in using LDAP, select the Sync Groups on Login option from Management Console Administration Authentication , so that the associated groups can be imported when you log in to CDV.

Procedure

1. In Cloudera Data Warehouse, click Data Visualization in the left navigation panel.
A list of existing Data Visualization instances appears, if there are any.

- If you are creating the first Data Visualization instance in CDW, click **CREATE**.

Overview

Database Catalogs

Virtual Warehouses

Data Visualization

Real Time Event Store Analy...

Data Visualization

Search

ADD NEW

NAME	DATA VISUALIZATION ID	Environment ID	VERSION	CPU	MEMORY	UPTIME	CREATED BY
pyrwviz1	viz-1664890117-qpc5c	env-r5dcq9	7.0.2-35	2	8 GB	an hour	pyrwgucs Data Viz
se-aws-visualization	viz-1664883821-ju98	env-t8hlf	7.0.2-35	2	8 GB	5 hours	cristina.sanchez Data Viz
telco-demo-streaming-dataviz	viz-1664877374-t9h6	env-t8hlf	7.0.2-35	2	8 GB	5 hours	frothkoeffer Data Viz
sf-dataviz1	viz-1664804575-9unj	env-t8hlf	7.0.2-35	2	8 GB	a day	sfransen Data Viz

1

3. Provide the following information in the **New Data Visualization** modal window:

Mandatory fields are marked with an asterisk.

- a) Name* – Specify a name for the instance.
- b) Environments* – Select the environment that you want CDV to be connected to.
- c) User Groups – Add user groups to allow user access to Data Visualization for selected groups. If no group is added, all CDP users will have non-admin access.
- d) Admin Groups* – Add admin groups to allow configuration access to Data Visualization for selected groups.

For more information on Data Visualization permission granularity, see the *Security model*.

- e) Tagging – Enter keys and values to apply tags to your resources for organizing them into a taxonomy.



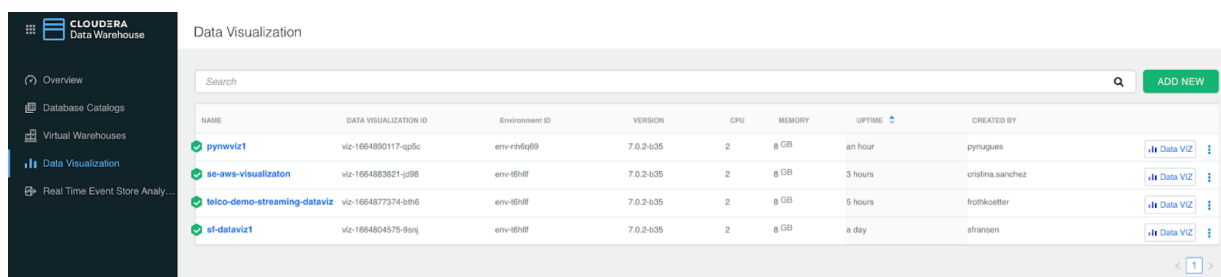
Note: This field is not available in CDW Private Cloud.

- f) Size* – Select the size of the compute instance from the drop-down menu.

4. Click CREATE.

Instance creation starts. Wait until the CDV instance is in running state.


5. You can find the list of Data Visualization instances and environments appears under the Data Visualization menu that you can open from the left navigation panel.



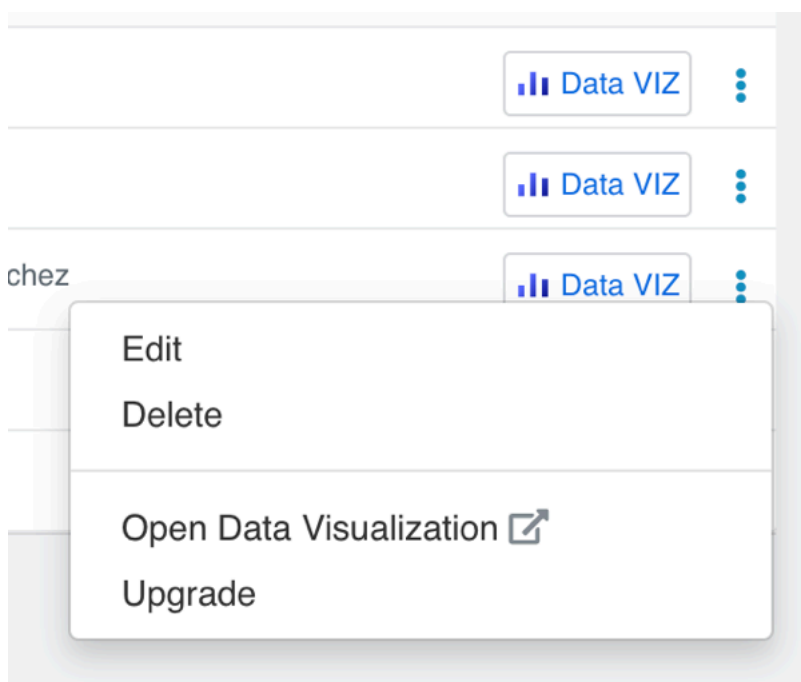
The screenshot shows the Cloudera Data Warehouse interface. On the left is a dark navigation panel with the 'Data Visualization' menu item highlighted. The main area is titled 'Data Visualization' and contains a table of instances. At the top right of the table is a search bar and an 'ADD NEW' button. The table has columns for NAME, DATA VISUALIZATION ID, Environment ID, VERSION, CPU, MEMORY, UPTIME, and CREATED BY. There are four rows of data, each with a 'Data VIZ' button and a three-dot menu icon at the end of the row.

NAME	DATA VISUALIZATION ID	Environment ID	VERSION	CPU	MEMORY	UPTIME	CREATED BY
pymwiz1	viz-1664880117-qp5c	env-rh6q89	7.0.2-b35	2	8 GB	an hour	pymugues
se-aws-visualization	viz-1664883821-jd98	env-16hllf	7.0.2-b35	2	8 GB	3 hours	crisina.sanchez
telco-demo-streaming-dataviz	viz-1664877374-b8f6	env-16hllf	7.0.2-b35	2	8 GB	5 hours	frithkoetter
sf-dataviz1	viz-1664804575-9apj	env-16hllf	7.0.2-b35	2	8 GB	a day	sfransen

6.

Select one from the list of running CDV instances and click  to start Data Visualization.

Alternatively, you can click the launch/options menu (three dots) at the end of the row and click Open Data Visualization in the pop-up menu.



SSO authentication is enabled by default both in Public Cloud and Private Cloud. If you are logged in to the Control Plane, you will be logged into Cloudera Data Visualization automatically. If not, log into the Control Plane using your LDAP credentials. Data Visualization opens in a new browser tab and you land on the Data Visualization homepage, where you can explore sample dashboards and access the in-tool *Get Started guide* for help.

New users logging in to Cloudera Data Visualization are automatically added to the viz_guest_group group. You can assign the System Admin role to this group to grant new users the permission to create new connections. For more information, see *Assigning roles to users*.

Deploying a CDV application in CML

Learn how to deploy Cloudera Data Visualization (CDV) in Cloudera Machine Learning (CML).

Creating a CML project with CDV Runtime

Learn how to create a Cloudera Machine Learning (CML) project with Cloudera Data Visualization (CDV) Runtime as the default runtime.

About this task

If you know that your project is going to be running CDV, you can add the Cloudera Data Visualization Runtime when setting up the project.

Procedure

1. Click Projects on the left sidebar of Cloudera Machine Learning data service.
2. Click New Project.
3. Enter a project name.
4. You can add a description for your project.
5. You can also select the visibility of the project.
6. Under Initial Setup, you can either create a blank project, or select a source for your project files.
7. Set Cloudera Data Visualization Runtime for your project in the Runtime setup section.

You can use the Basic view to select the Cloudera Data Visualization Runtime and add all of its Standard Edition ML Runtimes for that kernel. In the Advanced view, you can add ML Runtimes based on more detailed Editor, Kernel, Edition, and Version criteria.

8. Click Create Project.

What to do next

After the project is created, you can start creating your application. If you added Cloudera Data Visualization Runtime as the only runtime during project set-up, it will be the only available runtime when creating your application. You can select which edition you want to use for the application.

Adding CDV Runtime to a CML project

Learn how to add Cloudera Data Visualization (CDV) Runtime to an existing Cloudera Machine Learning (CML) project.

About this task

You need to add Cloudera Data Visualization Runtime to your project if the workspace of your project is not set to use this CDV Runtime by default.

Procedure

1. Click Projects on the left sidebar of Cloudera Machine Learning data service.
2. Click the project where you want to add Cloudera Data Visualization Runtime.
3. Open Project Settings from the left navigation bar and click the Runtime/Engine tab.
4. Click Add Runtime.

The Add new runtime to project modal window opens.

5. Select Cloudera Data Visualization from the Kernel drop-down list.
6. Select the edition of the CDV you want to add to your project.

The selected edition automatically defines the version of the runtime variant.



Note: New Cloudera Data Visualization Runtime versions are automatically added to the list of available editions, if internet connection is available.

7. Click Submit.

Results

You should now be able to select the Cloudera Data Visualization Runtime when creating a CDV application.

What to do next

Create a CDV application.

Creating CDV application in CML

Learn how to create a Cloudera Data Visualization (CDV) application in Cloudera Machine Learning (CML).

Before you begin

Ensure that CDV Runtime is available the CML project where you want to create the CDV application.

For more information about ML Runtimes, see [Managing ML Runtimes](#) and [Using Runtime Catalog](#).

Procedure

1. Go to your CML project's Overview page.
2. On the left sidebar, click Applications.
3. Click New Application.
4. Provide the following details for your new application:
 - Name – Enter a name for the application.
 - Subdomain – Enter a subdomain that will be used to construct the URL for the web application. Make sure it contains URL friendly characters.
 - Description – Enter a description for the application.
 - Script – Select the path to the startup script.



Note: Use the script in the location: /opt/vizapps/tools/arcviz/startup_app.py

- Runtime
 - Editor – Workbench
 - Kernel – Select Cloudera Data Visualization for the kernel supported by the Runtime variant of the CML project.
 - Edition – Select the edition of the Runtime variant you want to use for your application.



Note: For Data Visualization, the selected edition automatically defines the version of the Runtime variant.

5. Click Create Application.

Results

In a few minutes, you should see the application status change from Starting to Running on the Applications page. Your application has a hosted, fully-functional Data Visualization platform.

You can restart, stop, or delete a CDSW application from the supplemental menu of the application. If you want to make changes to the application, click Application Details and go to the Settings tab.

What to do next

Start Data Visualization.

Deploying a CDV application in CDSW

Learn how to deploy Cloudera Data Visualization (CDV) in Cloudera Data Science Warehouse (CDSW).

Creating a CDSW project with CDV Runtime

Learn how to create a Cloudera Data Science Workbench (CDSW) project with Cloudera Data Visualization (CDV) Runtime as the default runtime.

Procedure

1. Click Projects on the left sidebar of CDSW.
2. Click New Project on the Projects page.
3. Enter a project name.
4. Select the visibility of the project.
5. Under Initial Setup, you can either create a blank project, or select a source for your project files.
6. Click Create Project.

The created project uses Cloudera ML Runtimes that are managed in the Runtime Catalog.

What to do next

After the project is created, you can start creating an application. You can select which runtime edition you want to use for the application.

Creating Data Visualization application in CDSW

Learn how to create a Data Visualization application in Cloudera Data Science Workbench (CDSW).

Procedure

1. Go to your CDSW project's Overview page.
2. On the left sidebar, click Applications.
3. Click New Application.

4. Provide the following details for your new application:

- Name – Enter a name for the application.
- Subdomain – Enter a subdomain that will be used to construct the URL for the web application. Make sure it contains URL friendly characters.
- Description – Enter a description for the application.
- Enable Unauthenticated Access – Mark the checkbox if you want to allow unauthenticated access to your application. You can also update this setting later on the Settings page of the application.



Note: To create public applications on an ML workspace, an Admin user needs to turn on the feature flag in Admin Security by selecting Allow applications to be configured with unauthenticated access. For more information on working with public applications, see *Application Limitations* in the Cloudera Data Science Workbench documentation.

- Script – Select the path to the startup script.



Note: Use the script in the location: `/opt/vizapps/tools/arcviz/startup_app.py`

- Runtime
 - Editor – Workbench
 - Kernel – Select Cloudera Data Visualization for the kernel supported by the Runtime variant of the CDSW project.
 - Edition – Select the edition of the Runtime variant you want to use for your application.



Note: For Data Visualization, the selected edition automatically defines the version of the Runtime variant.

All available runtimes are listed in the CDSW Runtime Catalog. For more information about ML Runtimes, see [Managing ML Runtimes](#), and [Using Runtime Catalog](#).



Note: New ML Runtime releases are automatically added to the deployment, if internet connection is available.

5. Click Create Application.

Results

In a few minutes, you should see the application status change from Starting to Running on the Applications page. Your application has a hosted, fully-functional Data Visualization platform.

You can restart, stop, or delete a CDSW application from the supplemental menu of the application. If you want to make changes to the application, click Application Details and go to the Settings tab.

What to do next

Start Data Visualization in CDSW.

Starting Data Visualization in CDSW

Learn how to start the Cloudera Data Visualization (CDV) application you have created in Cloudera Data Science Workbench (CDSW).

Procedure

1. On the Applications page, click the name of your Data Visualization application to access the login interface.

2. Log in to Data Visualization by entering your username and password. Use your workload credentials.

If you want to log in as an administrator, you can use the following default credentials:

- username: vizapps_admin
- password: vizapps_admin

When using the default credentials to log in to Data Visualization, you are prompted to change your password at the first login.



Important: If you use the default credentials, security issues may arise. Cloudera recommends that you change the default username and password.

SSO authentication is disabled by default. See [Authentication](#) for information on how to permit user authentication with the CDSW login credentials and log users in automatically.

After logging in, you land on the homepage view of Cloudera Data Visualization. Here you can explore some sample dashboards or access the in-tool Get Started guide for help.

Connecting to a data source in Cloudera Data Visualization

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

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 Data Visualization with Cloudera Data Warehouse (CDW), 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 (CML), you can connect to an Impala or a Hive data warehouse, or tie in data from predictive CML models.

Creating a CML data connection to Impala

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

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

For more information on using Kerberos for authentication in CML, see *Hadoop Authentication for ML 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
cereals main.cereals Data Connection: samples	11	Extract Source	May 03, 2021	19 hours ago		juno	69
cereal main.cereals, main.chicago.govt_pay Data Connection: samples	191		Sep 28, 2022	19 hours ago		juno	24
cereal main.cereals, main.chicago.govt_pay Data Connection: 2050_CloneSamples	341		May 23, 2023	19 hours ago		vizapps_admin	19
csimport Solr.csaba_import_20240119 Data Connection: SolrTestConnection	417		Jan 23, 2024	2 days ago		juno	1
Dataset-2-test main.census_pop Data Connection: samples	416		Jan 23, 2024	2 days ago		vizapps_admin	1
Dataset-test-1 main.census_pop Data Connection: samples	415		Jan 23, 2024	2 days ago		vizapps_admin	1
vector Solr.cdv_embeddings_out_1536_100_rows_1703174544 Data Connection: SolrTestConnection	414		Jan 23, 2024	2 days ago		vizapps_admin	1
asd main.cdv_1937 Data Connection: SQLite	413	Extract Target	Jan 22, 2024	2 days ago		juno	0
report main.reports_report, main.reports_appgroup, main.rep... Data Connection: meta	412		Jan 22, 2024	2 days ago		vizapps_admin	0

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.

CLUSTERA
Data Visualization

HOME SQL VISUALS DATA

find titles, viz types, datasets, authors...

NEW CONNECTION

All Connections

2050_CloneSamples

Datasets 356 Data Extracts

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

Password

.....

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.

7. Click the Advanced tab to configure additional details.

Create New Data Connection

Connection type

Impala

Connection name

doc-test

Basic

Advanced

Parameters

Data

Connection mode

☐ Binary
 ☒ HTTP

HTTP Path

SQL path (default cliservice)

Socket type

☐ Normal
 ☒ SSL
 ☐ SSL with certificate

Authentication mode

☐ NoSasl
 ☒ Plain
 ☐ LDAP
 ☐ Kerberos

Socket Timeout

60

Impersonation ⓘ

☐ Enabled

Trusted ⓘ
Impersonation

☐ Enabled

Application Name

viz

TEST

CONNECT

- a) For HTTP connection mode, locate the Impala Endpoint for the Data Hub.

Event History Autoscale Endpoints (3) Tags (8) Hardware Network Telemetry Repository Details Image Details Recipes (0)	
Database Upgrade	
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;

- b) 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

Impala

Connection name

doc-test

Basic

Advanced

Parameters

Data

Parameter Name	Parameter Value	
Add new row		

TEST

CONNECT

Create New Data Connection

Connection type

Impala

Connection name

doc-test

Basic

Advanced

Parameters

Data

Concurrency ⓘ

55

Concurrency Per ⓘ
User

5

Query Timeout ⓘ
(Minutes)

22

Query Still Loading
Warning (Seconds)

20

Row upload limit ⓘ

10000

Result Cache

☐ Enabled

Cache Retention
Time (seconds)

86400

TEST

CONNECT

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](#)

Creating a CML data connection to a Hive data warehouse

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

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 CML data connection in Cloudera Data Visualization (CDV) 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.

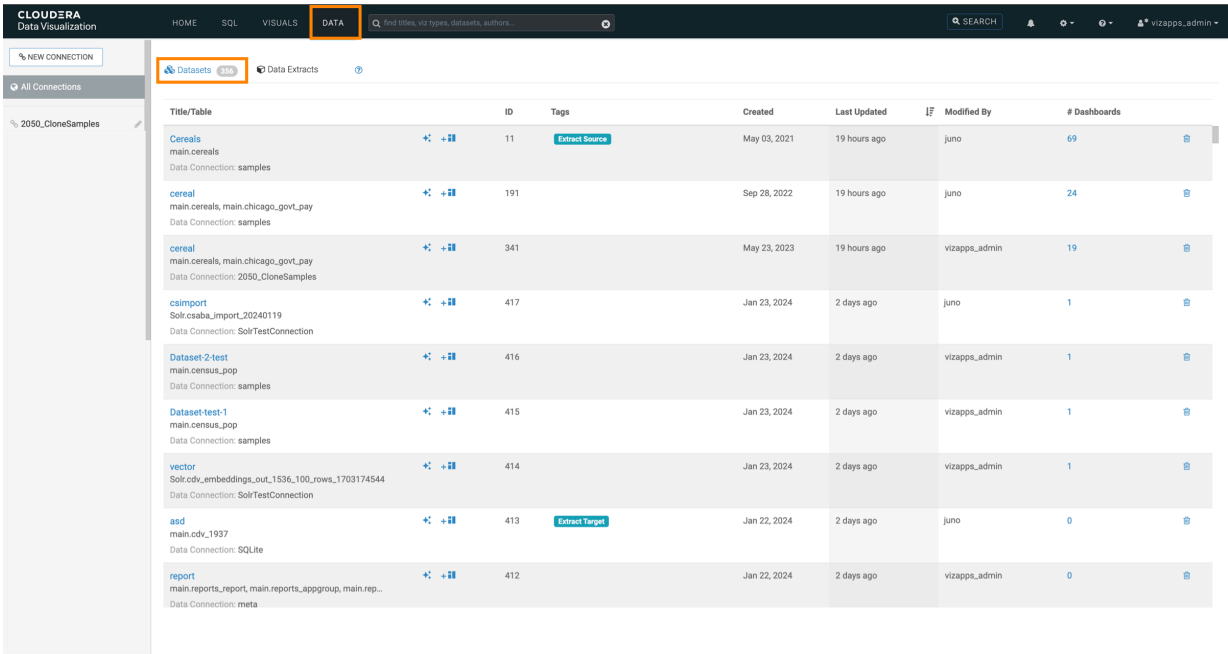
- For more information on the Manage data connections privilege, see *RBAC permissions*.
- For instructions on how to define privileges for a specific role, see *Setting role privileges*.
- For instructions on how to assign the administrator role to a user, see *Promoting a user to administrator*.

Before you begin

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

Procedure

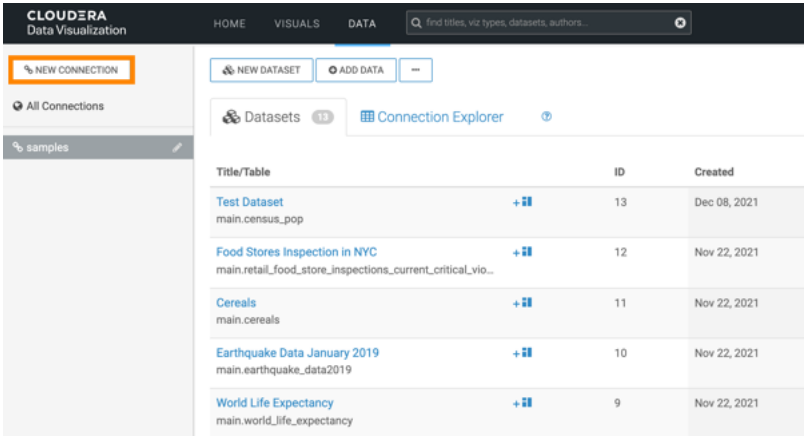
- 1. On the main navigation bar, click DATA.
The DATA opens, displaying the Datasets tab.



- 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 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

Hive

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

Password

.....

TEST

CONNECT

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.

- Click the Advanced tab to configure the additional details.

Create New Data Connection

Connection type

Hive

Connection name

doc-test

Basic

Advanced

Parameters

Data

Connection mode

☐ Binary

☒ HTTP

HTTP Path

SQL path (default cliservice)

Access Token

Access token (optional)

Socket type

☐ Normal

☒ SSL

☐ SSL with certificate

Authentication mode

☐ NoSasl

☒ Plain

☐ LDAP

☐ Kerberos

Socket Timeout

60

Impersonation ⓘ

☐ Enabled

Trusted ⓘ
Impersonation

☐ Enabled

Application Name

viz

TEST

CONNECT

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

Create New Data Connection

Connection type

Hive

Connection name


doc-test

Basic

Advanced

Parameters

Data

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

TEST

CONNECT

9. Check the Data tab for more configuration options.

Create New Data Connection

Connection type

Hive

Connection name

doc-test

Basic

Advanced

Parameters

Data

Concurrency ⓘ

50

Concurrency Per ⓘ
User

5

Query Timeout ⓘ
(Minutes)

22

Query Still Loading
Warning (Seconds)

20

Row upload limit ⓘ

10000

Result Cache

☐ Enabled

Cache Retention
Time (seconds)

86400

TEST

CONNECT

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 CDW data connection in Data Visualization

Learn how to connect to data when using Cloudera 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:

To create and manage connections in Cloudera Data Visualization, you must have the Manage data connections privilege or hold administrative privileges. In CDW, these are the members of the Admin Groups associated with the CDV 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 Data Visualization from the left navigation panel in CDW.
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
Cereals main.cereals Data Connection: samples	11	Extract Source	May 03, 2021	19 hours ago		juno	69
cereal main.cereals, main.chicago_govt_pay Data Connection: samples	191		Sep 28, 2022	19 hours ago		juno	24
cereal main.cereals, main.chicago_govt_pay Data Connection: 2050_CloneSamples	341		May 23, 2023	19 hours ago		vizapps_admin	19
csimport Solr.csaba_import_20240119 Data Connection: SolrTestConnection	417		Jan 23, 2024	2 days ago		juno	1
Dataset-2-test main.census_pop Data Connection: samples	416		Jan 23, 2024	2 days ago		vizapps_admin	1
Dataset-test-1 main.census_pop Data Connection: samples	415		Jan 23, 2024	2 days ago		vizapps_admin	1
vector Solr.cdv_embeddings_out_1536_100_rows_1703174544 Data Connection: SolrTestConnection	414		Jan 23, 2024	2 days ago		vizapps_admin	1
asd main.cdv_1937 Data Connection: SQLite	413	Extract Target	Jan 22, 2024	2 days ago		juno	0
report main.reports_report, main.reports_appgroup, main.rep... Data Connection: meta	412		Jan 22, 2024	2 days ago		vizapps_admin	0

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 links for 'All Connections' and 'samples'. The main content area shows a 'Datasets' section with a 'Connection Explorer' button. A table lists several datasets with their titles, IDs, and creation dates.

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 CDW, and the option is disabled in the connection list.

Create New Data Connection

Connection type

CDW Hive

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

Password

.....

TEST

CONNECT

5. 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

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 CDSW data connection to a data warehouse

Learn how to connect natively to data stored in a data warehouse when using 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:

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 has tabs for HOME, VISUALS, and DATA. The DATA tab is selected. Below the navigation bar, there are buttons for NEW CONNECTION, NEW DATASET, and ADD DATA. The left sidebar shows a list of connections, with 'samples' selected. The main area displays a table of datasets.

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. The top navigation bar has tabs for HOME, VISUALS, and DATA. The DATA tab is selected. Below the navigation bar, there are buttons for NEW DATASET, ADD DATA, and NEW CONNECTION. The left sidebar shows a list of connections, with 'samples' selected. The main area displays 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.

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.

9. 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](#)

Creating a visual

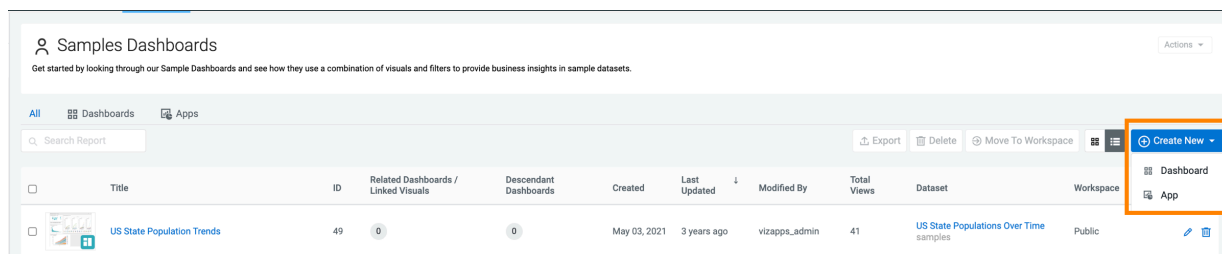
You can easily create new visual representations for your data in Cloudera Data Visualization.

About this task

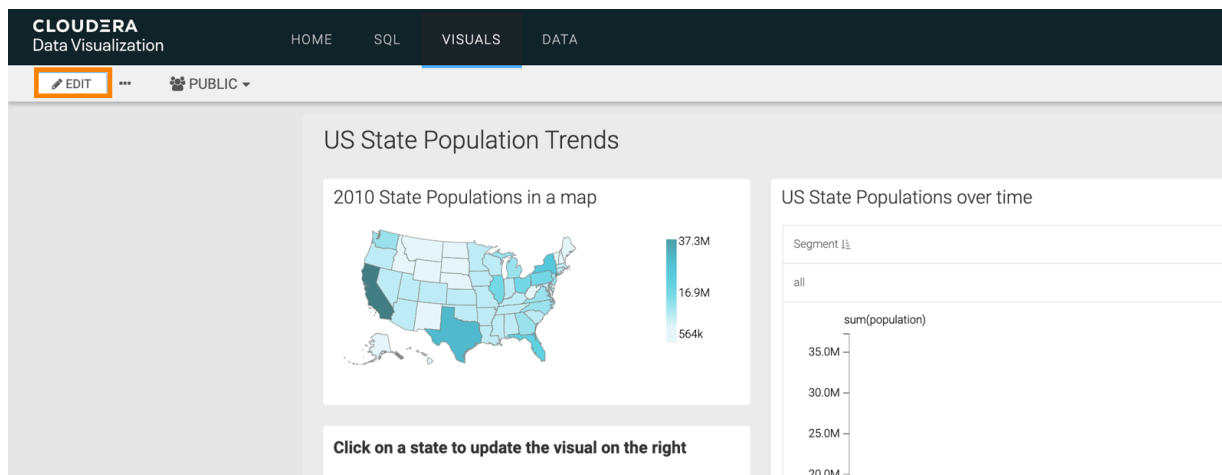
The following steps demonstrate how to create a new visual.

Procedure

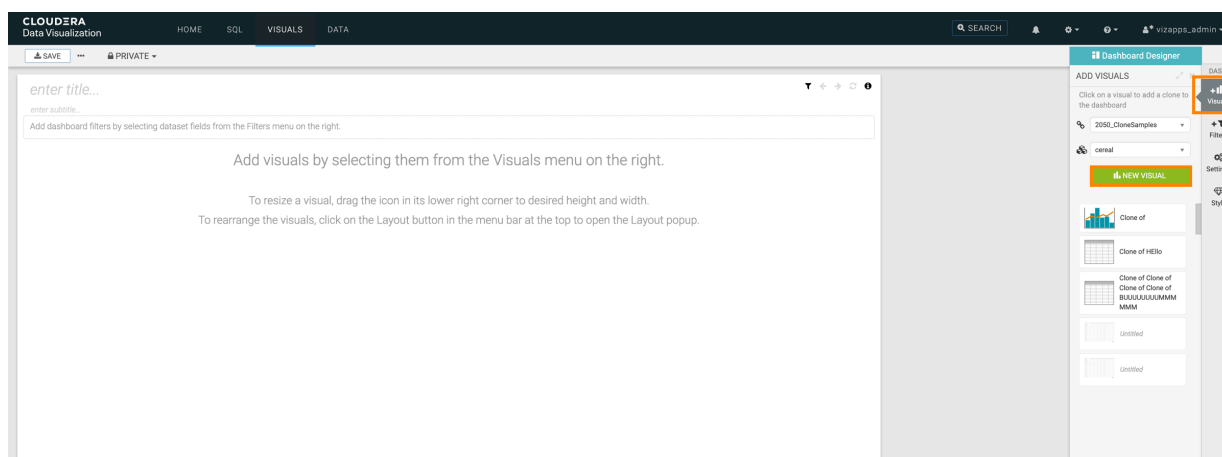
1. On the main navigation bar, click VISUALS.
2. Click **Create New Dashboard** to open a new, blank dashboard.



If you already have an existing dashboard and you just want to add a new visual, open the dashboard and click **EDIT** to make changes to it.



3. In the Dashboard Designer interface, open the Visuals menu from the side menu bar and click NEW VISUAL.



4. The Visual Builder appears.
By default, a table visual is selected. For more information, see the *Visual Builder* overview.

- 5. Choose the visual type you want to build.
In this example, the Bars visual has been selected.

Dashboard Designer

VISUALS

Bars

1234 LABEL

WORD CLOUD

ACTION

SQL

X Axis

content

Y Axis

Record Count

Colors

drag fields to add here

Tooltips

drag fields to add here

Drill

drag fields to add here

Labels

drag fields to add here

Filters

drag fields to add here

REFRESH VISUAL

DATA

cm_logs_DL

Sample Mode: OFF

Search

Dimensions

cm_logs_goes_v2

content

timeoccurred

severity

Measures

cm_logs_goes_v2

Record Count

DASH.

Visuals

Filters

Settings

Style

VISUAL

Build

Settings

Style

6. You can switch on Sample Mode for the visual you are building.

With sample mode, you create the visual on a subset of your data to test it. You can define this subset as a percentage of the original data.



Note: This feature is only available on Impala, Hive, and Spark connections.

Run in Sample Mode

Running in sample mode will allow you to run on a subset of your data which will make the chart load faster.

Sample Mode: % of data

CANCEL

APPLY

7. Populate the shelves of the visual from the Dimensions and Measures fields:

In this example, the following fields have been added:

- state dimension on the X Axis shelf
- population measure on the Y Axis shelf



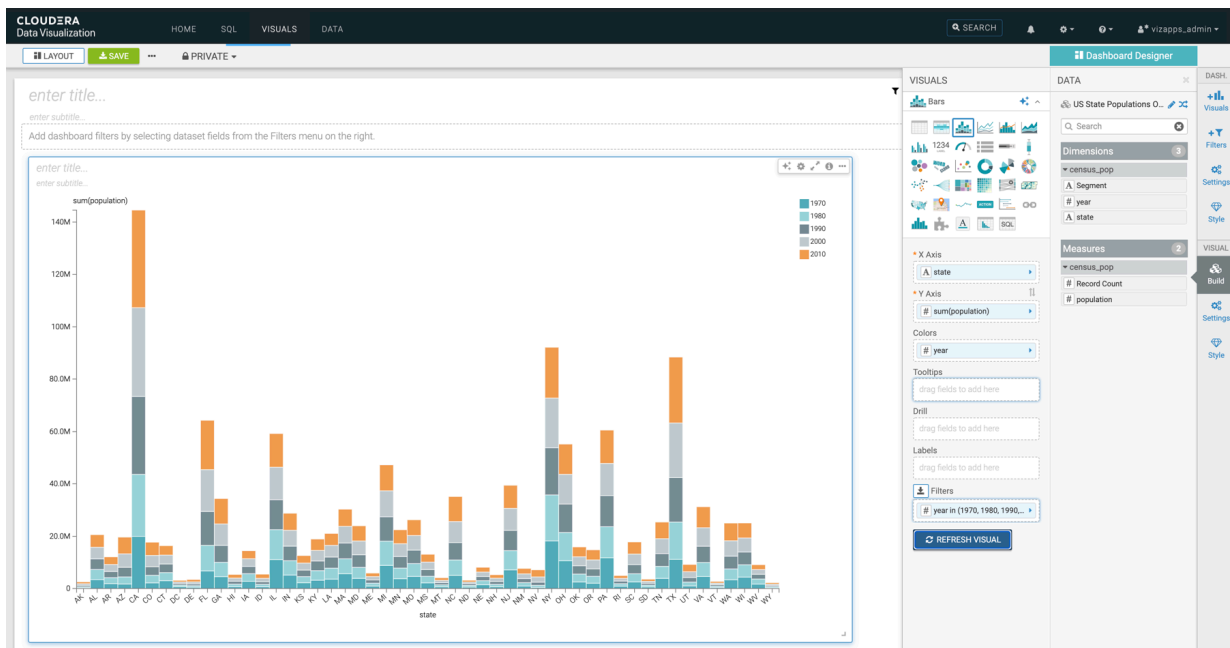
Note: The population measure appears with its default aggregation, as sum(population).

- year measure on the Colors shelf
- year measure on the Filters shelf



Note: The Filter for year module opens automatically. Years 1970, 1980, 1990, 2000, and 2010 have been selected.

8. Click REFRESH VISUAL.



9. Click the enter title... field and enter a name for the visual.

- In this example the title 'Recent US Census Results by State' has been added.
- You can also click the enter subtitle... field below the title of the visualization to add a brief description of the visual.

10. Click SAVE at the top left corner of the Dashboard Designer.