Cloudera Data Visualization 8.0.9

Cloudera Data Visualization Top Tasks

Date published: 2020-10-30
Date modified: 2025-12-16

https://docs.cloudera.com/

Legal Notice

Contents

Deploying Cloudera Data Visualization

To start using Cloudera Data Visualization, the first step is to deploy it on your preferred Cloudera platform. Cloudera Data Visualization provides a consistent set of visualization capabilities across various Cloudera services and deployment environments.

Once your Cloudera Data Visualization instance is up and running, you can log in using your Cloudera credentials and immediately start working with any datasets you have access to. Start creating visuals, build customized dashboards, and develop analytical applications to enable users to explore data across the entire data lifecycle.

Cloudera Data Visualization is available in both cloud and on-premises environments:

- Cloudera on cloud:

  Cloudera Data Visualization is included as a core capability in both Cloudera Data Warehouse and Cloudera AI on Azure and AWS.
- Cloudera on premises:

  Cloudera Data Visualization is available in Cloudera Base on premises, and in the Cloudera Data Warehouse and Cloudera AI data services.

Installing Cloudera Data Visualization in Cloudera Base on premises

Learn how to deploy Cloudera Data Visualization in Cloudera Base on premises.

Cloudera Data Visualization, when deployed in Cloudera Base on premises, allows you to explore datasets and build interactive visualizations directly within your on-premises Cloudera clusters. One key benefit of this deployment option is that it does not require a container runtime, and can be used without a Kubernetes environment.

This guide provides step-by-step instructions for installing Cloudera Data Visualization on a Cloudera Base on premises cluster.

Limitations

- **OS compatibility:**

  Ubuntu 20, as well as RHEL 8 and 9, are supported. For the detailed list of supported OS versions, see the Support Matrix.
- **Data source connections:**

  Data connection auto-discovery is not supported. You need to manually set up and configure connections to the data sources.

Prerequisites

- Cloudera Manager and Cloudera Base are installed and running.
  - For installation instructions, see:
    - Installing Cloudera Manager
    - Installing Cloudera Runtime
  - For supported versions, see the Support Matrix.
- You have sudo or root access to the relevant cluster nodes.
- You have a MySQL or other relational database configured for metadata storage.

Important:

Install Cloudera Data Visualization on a dedicated host to avoid resource contention with other services. Do not deploy it on nodes already running resource-intensive roles. The Cloudera Manager installation wizard evaluates host hardware to recommend role placements. Ensure that hardware is properly configured before running the wizard to allow appropriate role assignments.

Adding Cloudera Data Visualization parcel repository to Cloudera Manager

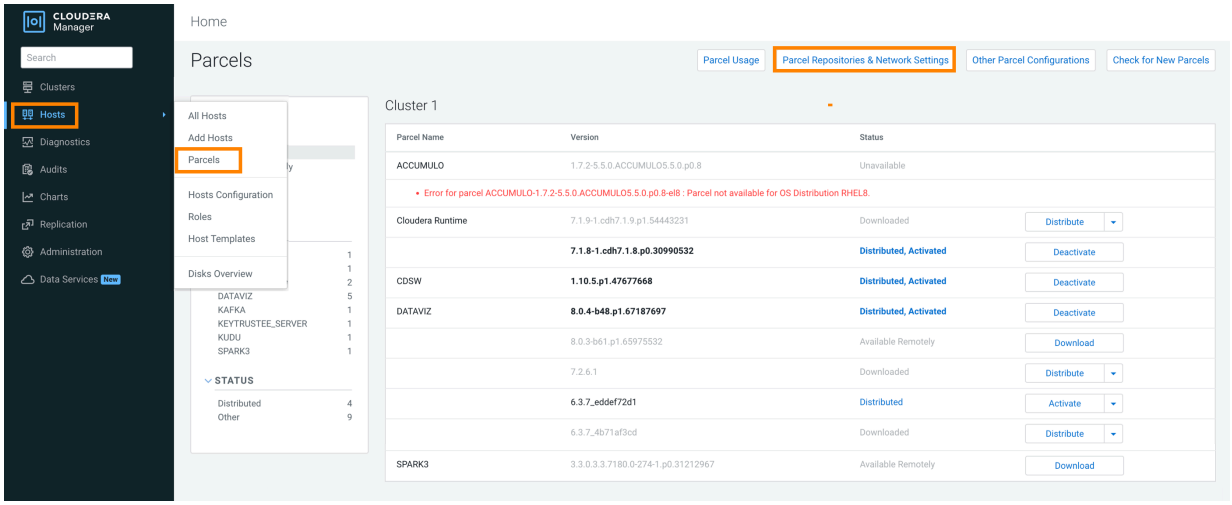
Before installing Cloudera Data Visualization, you must set up two components: a Cloudera Data Visualization parcel and a Custom Service Descriptor (CSD).

About this task

This guide explains how to add the Cloudera Data Visualization parcel to Cloudera Manager. It involves updating the parcel repository URL, then downloading, distributing, and activating the parcel to make it available for installation.

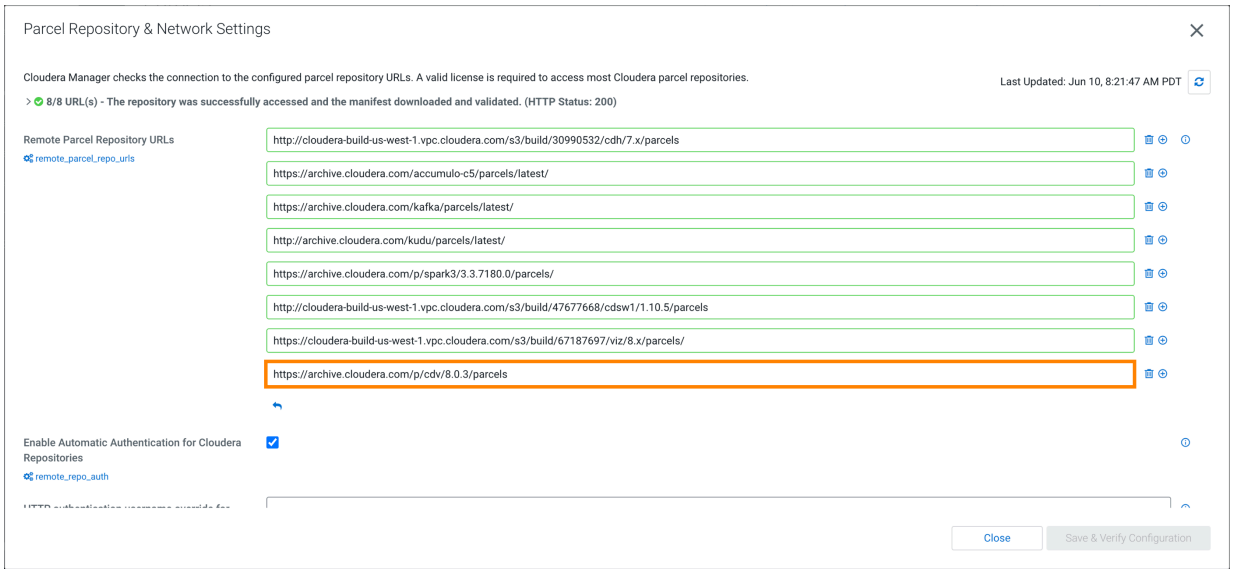
Procedure

1. Log in to Cloudera Manager.
2. From the left navigation bar, go to Hosts Parcels .
3. Select Parcel Repositories & Network Settings on the Parcels page.



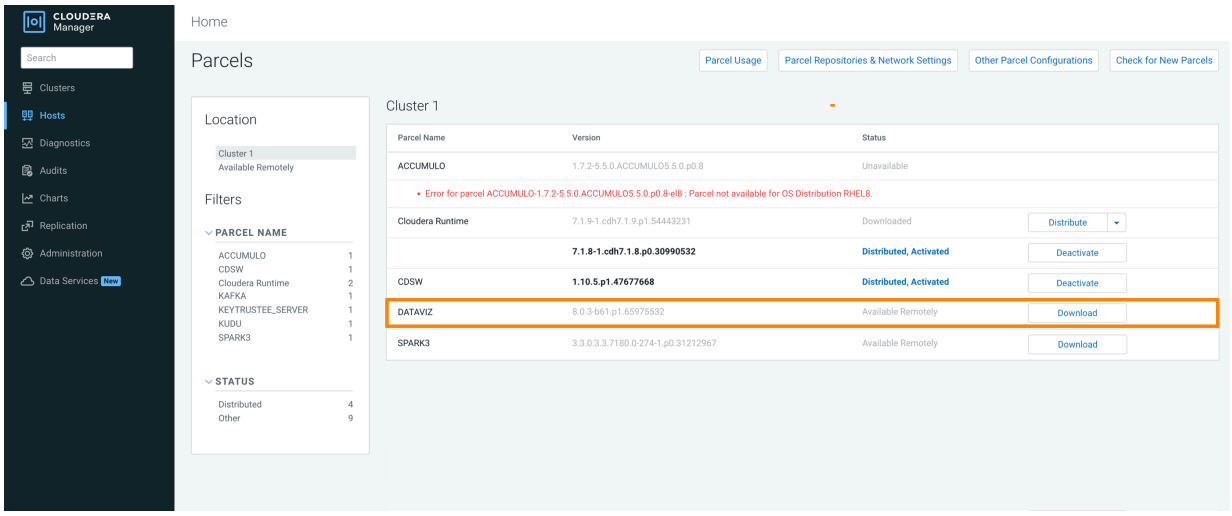
4. On the Parcel Repository & Network Settings page, click + to add a new row in the Remote Parcel Repository URLs list.

5. Enter the URL of the Cloudera Data Visualization parcel repository.



Note: Do not include the specific parcel file name in the URL, only provide the directory path. Cloudera Manager will scan the directory and display all available parcels. For example: https://archive.cloudera.com/p/cdv/8.0.3/parcels/.

6. Click Save & Verify Configuration.
7. Click Close to return to the Parcels page.
8. On the Parcels page, locate the Cloudera Data Visualization parcel in the list.



9. Click Download.
10. Once the download is complete, click Distribute.
11. After distribution, click Activate.

    When prompted, choose whether to restart the service or activate without restarting. Select your preferred option and click OK.

Results

After activation, the parcel becomes available for use when adding the Cloudera Data Visualization service.

What to do next

Once the parcel is activated, you must add the Cloudera Data Visualization CSD files to Cloudera Manager. For instructions, see *Downloading Cloudera Data Visualization CSD files*.

Downloading and adding Cloudera Data Visualization CSD files

A Custom Service Descriptor (CSD) file contains the configuration details for Cloudera Managerto recognize and manage a new service. Follow the steps below to download the required CSD file, apply the correct permissions, and restart Cloudera Manager.

Before you begin

- You must have root or sudo privileges on the Cloudera Manager host.

Procedure

1. Open a terminal and connect to the host where Cloudera Manager is installed.

   ssh *[***USERNAME***]@[***CLOUDERA-MANAGER-HOST***]*

   - Replace *[***USERNAME***]* with a user that has appropriate permissions (for example: root or a sudo-enabled user).
   - Replace *[***CLOUDERA-MANAGER-HOST***]* with the appropriate hostname or IP address of your Cloudera Manager server.

   Example: ssh root@cloudera-manager.example.com
2. Once logged in, verify that the default CSD directory exists.

   ls -lA /opt/cloudera/csd

   Note: If the directory does not exist, create it before proceeding using mkdir /opt/cloudera/csd.
3. Navigate to the CSD directory.

   cd /opt/cloudera/csd
4. Download the Cloudera Data Visualization CSD files to the default CSD directory on the Cloudera Manager host.

   Note: The /opt/cloudera/csd directory requires root privileges. Use `sudo` to ensure the download succeeds when saving the file directly to this location.

   sudo wget https://*[***YOUR-DOWNLOAD-LOCATION***]*/*[***YOUR-CSD-FILE***]*.jar

   - Replace *[***YOUR-DOWNLOAD-LOCATION***]* with the download URL provided by your administrator or Cloudera.

     Note: Accessing https://archive.cloudera.com/p/ requires authentication. You can provide your Cloudera credentials using the --user and --password options.

     ```
     sudo wget --user=<your-username> --password=<your-password> http
     s://...
     ```
   - Replace *[***YOUR-CSD-FILE***]*.jar with the actual filename provided by your administrator or Cloudera.

   For example: https://archive.cloudera.com/p/cdv/8.0.3/redhat8/yum/DATAVIZ-8.0.3-b61.p1.65975532.jar
5. List the directory contents to confirm that the file has been downloaded.

   ls -lA

   You should see the downloaded JAR file listed in the directory.
6. Restart the Cloudera Manager Server to load the new CSD by running the following command on the host where Cloudera Manager is installed.

   sudo service cloudera-scm-server restart
7. Restart Cloudera Management Service in the UI.
   a) Go to the Status tab.
   b) Select Actions Restart to restart the Cloudera Management Service.

What to do next

After restarting, Cloudera Manager will detect the Cloudera Data Visualization service. You can now proceed to add and configure it in your cluster.

Adding Cloudera Data Visualization as a service

Use the Add Service wizard in Cloudera Manager to deploy and configure Cloudera Data Visualization on a Cloudera Base on premises cluster.

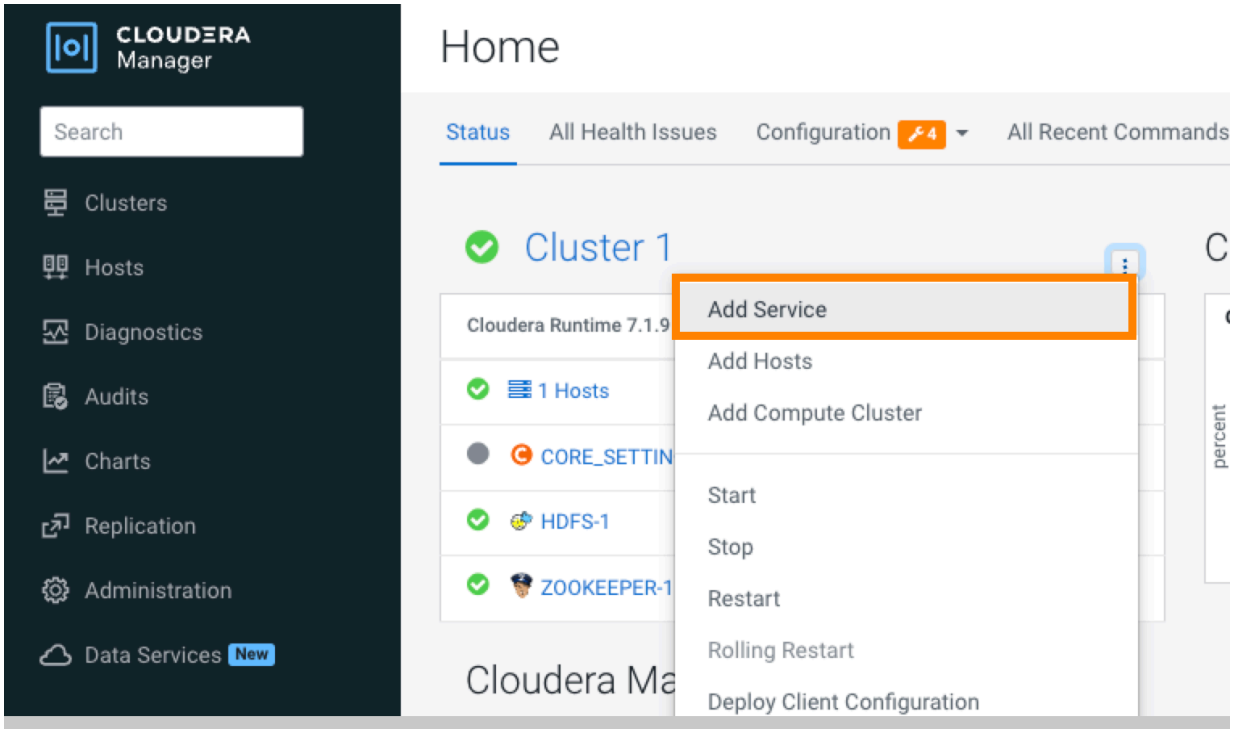
Before you begin

Ensure the following:

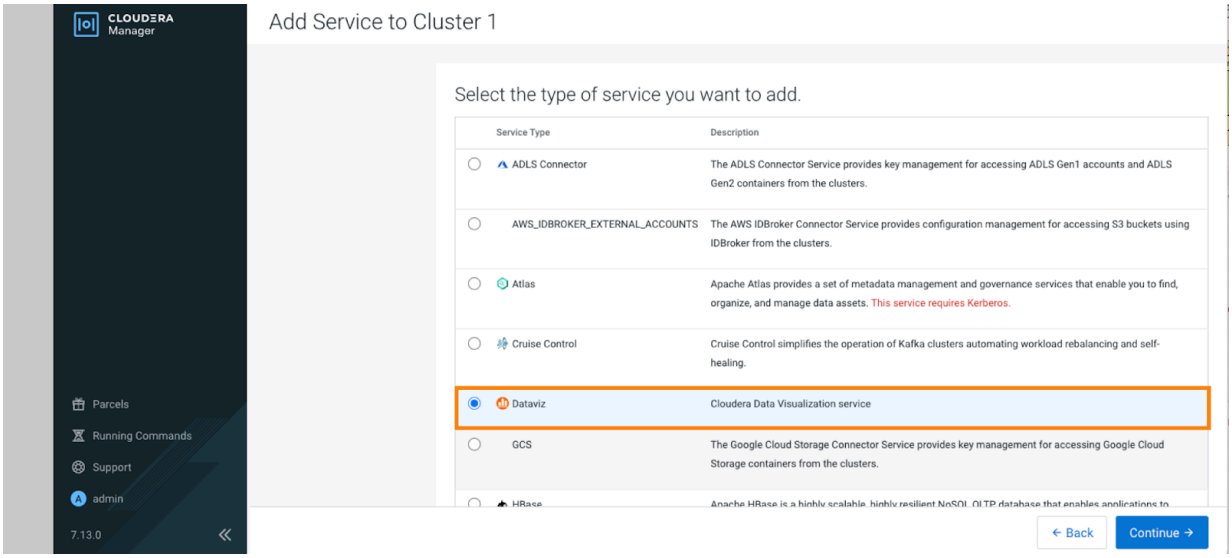
- A Cloudera Base on premises cluster is installed and running.
- You have added the Cloudera Data Visualization CSD files in the /opt/cloudera/csd folder, and Cloudera Manager has been restarted.
- You have added, downloaded, distributed, and activated the Cloudera Data Visualization parcel in Cloudera Manager.

Procedure

1. Log in to Cloudera Manager.
2. On the Home screen, open the drop-down menu next to your cluster name and select Add Service.



3. Select Dataviz from the list of available services, then click Continue.



The Add Service wizard is displayed.

4. In the Select Dependencies step, leave the default No Optional Dependencies setting and click Continue.
5. Assign roles to hosts.
   - On single-node clusters, roles are automatically assigned, so click Continue to proceed.
   - On multi-node clusters, manually assign roles to the desired hosts and then click Continue to proceed.
6. Configure a metadata database.
   a) Configure the required database connection details:
      - Type: Select your database type from the dropdown. (For example: MySQL)
      - Database Hostname: Enter the hostname used when creating the database.
      - Database Name: Add the name of the Cloudera Data Visualization database.
      - Username: Add the Cloudera Data Visualization database username.
      - Password: Add the Cloudera Data Visualization database password.
   b) You can click Test Connection to verify the database settings.
   c) Click Continue.
7. Review the service configuration summary, then click Continue to begin the installation.
8. On the Command Details screen, monitor the installation progress.
9. Once the installation is complete, click Continue.
10. On the Summary screen, click Finish to complete the setup.

    The Cloudera Data Visualization service is now added to your cluster. It will appear in the list of cluster services in Cloudera Manager.
11. Restart the Cloudera Management Service to update stale configurations with new monitoring entities.
12. Check the health status of the new service to verify it has started correctly.

    If the Health Status is Good, the service has started successfully.

Results

You have completed your Cloudera Data Visualization installation on Cloudera Base on premises. For further information about adding services to Cloudera Runtime, see *Adding a service*.

What to do next

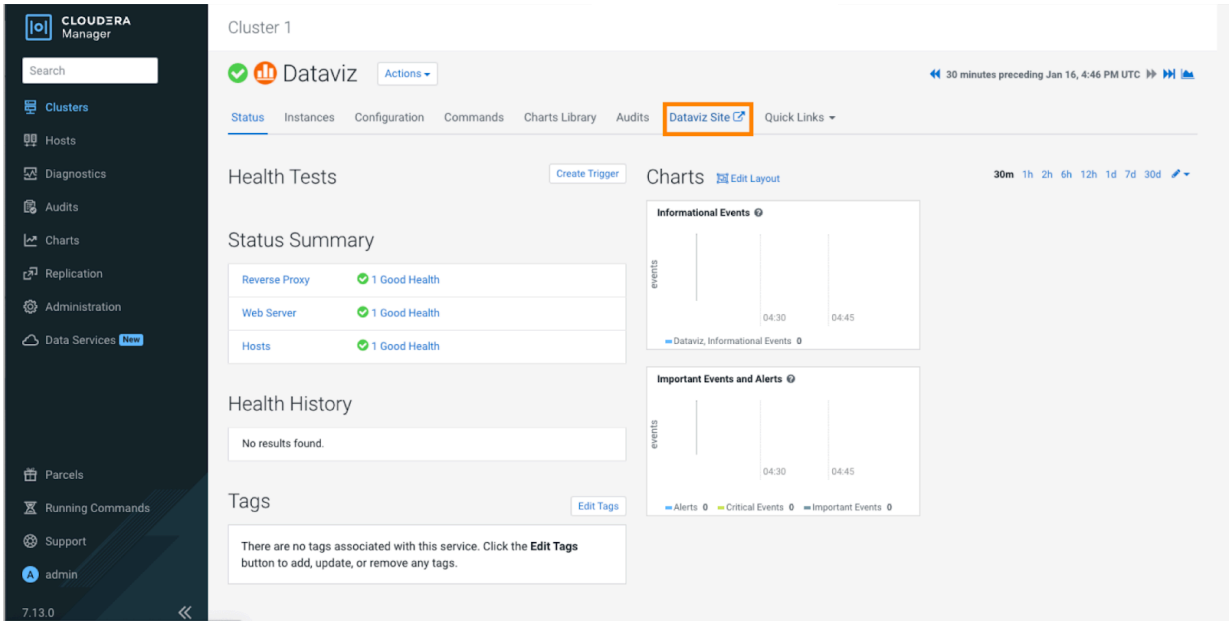
Open the Cloudera Data Visualization web interface to begin creating dashboards.

Accessing the Cloudera Data Visualization UI

After adding the service, follow these steps to access the Cloudera Data Visualization user interface.

Procedure

1. In Cloudera Manager, go to the Dataviz service details page.
2. Click Dataviz Site.



   This opens the Cloudera Data Visualization interface in a new browser tab.
3. Use the default credentials to log in:
   - Username: vizapps_admin
   - Password: vizapps_admin

   Note:

   It is recommended to change the default password after first login to help maintain the account security. For more information, see *Changing user passwords*.

Results

You have now successfully added the Cloudera Data Visualization service and logged into the user interface. You can begin:

- Connecting to data sources
- Creating dashboards and visualizations

Deploying Cloudera Data Visualization in Cloudera Data Warehouse

Learn how to create and launch a Cloudera Data Visualization instance within the Cloudera Data Warehouse data service. You can use Cloudera Data Visualization with Cloudera Data Warehouse in both Cloudera on cloud and Cloudera on premises to explore and visualize data stored in database catalogs providing data-driven insights throughout the entire data lifecycle.

About this task

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

Cloudera 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 Cloudera 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 Cloudera Data Visualization instance is not deleted.

Before you begin

- You are logged into the Cloudera web interface and you have opened the Cloudera Data Warehouse service.
- You have DWAdmin role in Cloudera Data Warehouse.
- You have activated your environment.
- You have a Hive/Impala warehouse in running state.
- If you are using Cloudera Data Visualization with Cloudera Data Warehouse in Cloudera on premises:

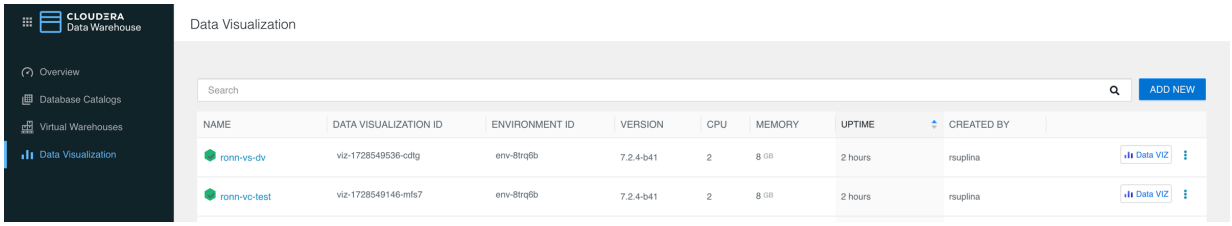
  You have an admin group created in Management Console User Management . To log in using LDAP, select the Sync Groups on Login option from Management Console Administration Authentication , so that the associated groups can be imported when you log in to Cloudera Data Visualization.

Procedure

1. In Cloudera Data Warehouse, click Data Visualization in the left navigation panel.

   A list of existing Cloudera Data Visualization instances appears, if there are any.
2. Click ADD NEW to create a new instance.

   If you are creating the first Cloudera Data Visualization instance in Cloudera Data Warehouse, click CREATE.



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 Cloudera Data Visualization to be connected to.
   c) User Groups – Add user groups to allow user access to Cloudera Data Visualization for selected groups. If no group is added, all Cloudera users will have non-admin access.
   d) Admin Groups* – Add admin groups to allow configuration access to Cloudera Data Visualization for selected groups.

      For more information on Cloudera 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.

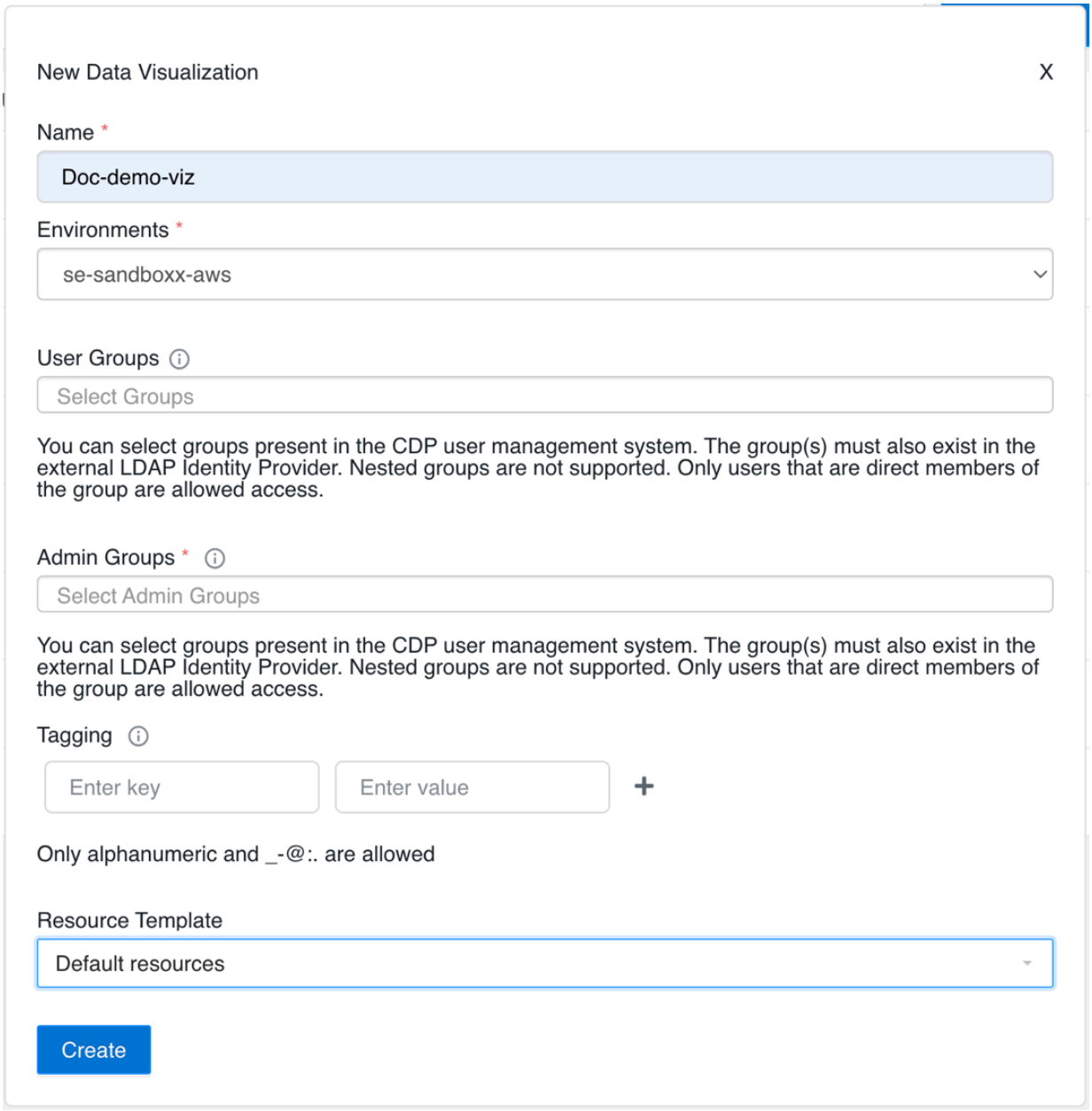
      Note: This field is not available in Cloudera Data Warehouse on Cloudera on premises.

   f) Resource Template – Select the Cloudera Data Warehouse resource template from the drop-down menu.

      You can use a predefined template or duplicate it to create a custom template to allocate Kubernetes resources based your workload requirements and node sizes. Each template has different overall CPU and memory allocations.

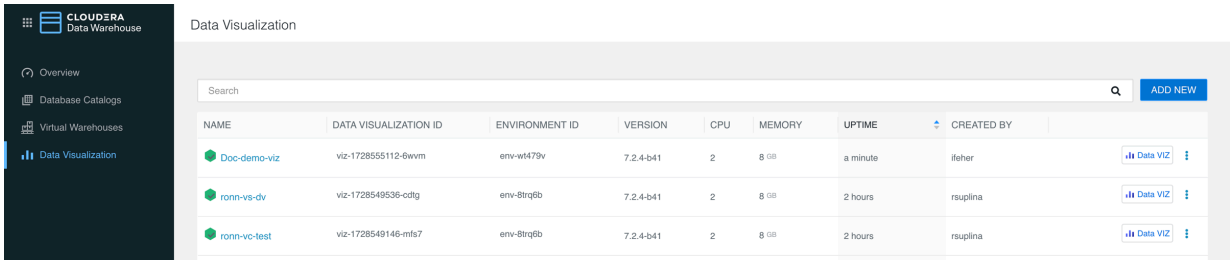
Resource template	Description	CPU	Memory
Default resources	Default resources for components inside Cloudera Data Visualization	2	8192 MB
Medium resources	Double the resources for Cloudera Data Visualization compared to default.	4	16384 MB
Large resources	Triple the resources for Cloudera Data Visualization compared to default	6	24576 MB

Resource template	Description	CPU	Memory
Reduced resources	Reduced resources for Cloudera Data Visualization, recommended only for initial exploration.	0.5	8192 MB



4. Click CREATE.
   Instance creation starts. Wait until the Cloudera Data Visualization instance is in running state.

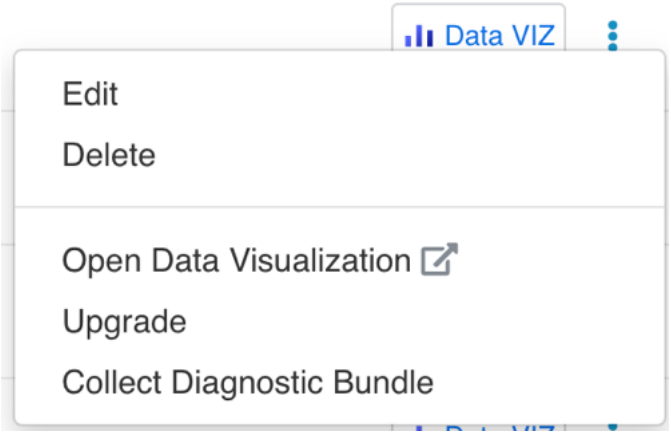
5. You can find the list of Cloudera Data Visualization instances and environments appears under the Data Visualization menu that you can open from the left navigation panel.



6. Select one from the list of running Cloudera Data Visualization instances and click 

to start Cloudera 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 Cloudera on cloud and Cloudera on premises. 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. Cloudera Data Visualization opens in a new browser tab and you land on the Cloudera 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 standalone Cloudera Data Visualization application in Cloudera AI

Learn how to deploy Cloudera Data Visualization in Cloudera AI.

Creating a Cloudera AI project with Cloudera Data Visualization Runtime

Learn how to create a Cloudera AI project with Cloudera Data Visualization Runtime as the default runtime.

About this task

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

Procedure

1. Click Projects on the left sidebar of Cloudera AI Workbench.
2. Click New Project.
3. Enter a Project Name.
4. You can add a description for your project.
5. Set the visibility of the project.
   - Private (default) – Only you can access the project.
   - Public – All authenticated users can view the project.
6. Under Initial Setup, select how you want to create your project.

   You can either create a blank project, or select a source for your project files.
   - Blank – Start with an empty project (no templates, files, or Git sources).
   - Templates – Pre-built example projects in R, Python, PySpark, or Scala to help you get started.
   - APMs – Use Accelerators for ML Projects to to include jobs, models, and experiments.
   - Local Files – Upload an existing project from a compressed file or folder.
   - Git – Clone a Git repository for version control and collaboration.
7. Configure which Runtime(s) will be available for this particular project.

   Projects are configured with the latest Python and R ML Runtimes. You can change this configuration in the Advanced Options view, where you can add ML Runtimes based on more detailed Editor, Kernel, Edition, and Version criteria.

   a) Enable Advanced Options to customize.
   b) For Cloudera Data Visualization 8.0.0 and higher:
      1. Set the Editor to PBJ Workbench.
      2. Select Cloudera Data Visualization from the Kernel dropdown.
      3. Edition and Version will be automatically set for Cloudera Data Visualization.
   c) For Cloudera Data Visualization versions lower than 8.0.0, set the Editor to Workbench.

      Kernel, Edition, and Version will be automatically set for Cloudera Data Visualization.
   d) Click Add Runtime.
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 as the only Runtime during setup, it will be the default Runtime when creating applications.

Adding Cloudera Data Visualization Runtime to an exisiting Cloudera AI project

Learn how to add a Cloudera Data Visualization Runtime to an existing Cloudera AI project.

About this task

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

If you want upgrade your Cloudera Data Visualization Runtime image to a newer version in an air-gapped deployment, you need to manually load the specific Runtime image into the cluster. For general guidance on air-gapped installation see *Updating ML Runtime images on Cloudera AI installations* in the Cloudera AI documentation.

Procedure

1. Click Projects on the left sidebar of Cloudera AI Workbench.
2. Select the project where you want to add Cloudera Data Visualization Runtime.

3. Open Project Settings from the left navigation bar.
4. Click the Runtime tab.
5. Click Add Runtime.

   The Add new runtime to project modal window opens.

   a) For Cloudera Data Visualization 8.0.0 and higher:

      1. Set the Editor to PBJ Workbench.
      2. Select Cloudera Data Visualization from the Kernel dropdown.
      3. Edition and Version will be automatically set for Cloudera Data Visualization.

   b) For Cloudera Data Visualization versions lower than 8.0.0, set the Editor to Workbench.

      Kernel, Edition, and Version will be automatically set for Cloudera Data Visualization.
6. Click Submit.

Results

Now that Cloudera Data Visualization Runtime is added, you can select it when creating a Cloudera Data Visualization application.

What to do next

To proceed, create a Cloudera Data Visualization application.

Creating a Cloudera Data Visualization application in Cloudera AI

Learn how to create a Cloudera Data Visualization application in Cloudera AI to help you visualize and interact with your data insights. This integration allows for seamless visualization of ML Model outputs, data exploration, and reporting within the same platform.

About this task

Important:

Each Cloudera AI project can host only one standalone Cloudera Data Visualization application. Since applications share the same metadata and logs, any changes made by one application may overwrite those made by another.

Before you begin

Ensure that a Cloudera Data Visualization Runtime is available in the Cloudera AI project where you plan to create the Cloudera Data Visualization application.

For more information about ML Runtimes, see Managing ML Runtimes and Using Runtime Catalog.

Procedure

1. Navigate to the Overview page of your Cloudera AI project.
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.
   - Run Application as – If the application is to run in a service account, select Service Account and pick an account from the dropdown.
   - Subdomain – Enter a subdomain that will be used to construct the URL for the web application. Use only URL-friendly characters.
   - Description – Add a description of the application.
   - Script – Use the script located at: /opt/vizapps/tools/arcviz/startup_app.py
   - Runtime – If only one Runtime is available in your project, the fields will be prepopulated. If multiple Runtimes are available, you can select which Runtime to use.
     For Cloudera Data Visualization 8.0.0 and higher
     - Editor – Select PBJ Workbench.
     - Kernel – Select Cloudera Data Visualization.
     - Edition and Version will autopopulate.

     For Cloudera Data Visualization versions lower than 8.0.0
     - Editor – Select Workbench.
     - Kernel, Edition, and Version will autopopulate.
5. Click Create Application.

Results

After a few minutes, the application status will change from Starting to Running on the Applications page. Your Cloudera Data Visualization application is now ready for use.

You can restart, stop, or delete the application using the options in the supplemental menu. If you want to make changes to the application, navigate to Application Details Settings .

What to do next

Start Cloudera Data Visualization.

Creating a data connection in Cloudera Data Visualization

Cloudera Data Visualization allows you to connect to various external data sources to enhance your data analysis and visualization capabilities.

About this task

Cloudera Data Visualization currently supports the following connection types:

- Hive
- Impala
- MariaDB
- MySQL
- PostgreSQL
- Phoenix [Technical Preview]
- Solr [Technical Preview]
- Spark SQL
- SQLite (not supported in Cloudera Data Warehouse)
- Snowflake [Technical Preview]
- Trino [Technical Preview]

These supported connection types provide flexibility and versatility for integrating multiple sources into your data analysis workflows.

In Cloudera Data Warehouse, the connection to the database catalog is automatically set up when you enable Cloudera Data Visualization in a Virtual Data Warehouse. You can also create your own connections to other data warehouses, but it is not supported.

Database catalogs and virtual data warehouses automatically inherit the same security restrictions that are applicable to your Cloudera environment. There is no need to specify the security setup again for each database catalog or virtual warehouse. If you cannot see the data in the connected database catalog after you log in, check and adjust data access permissions or your environment and data warehouse user permissions.

In Cloudera AI, you can set up several connection types. For example, you can connect Cloudera Data Visualization to an Impala or Hive data warehouse.

Before you begin

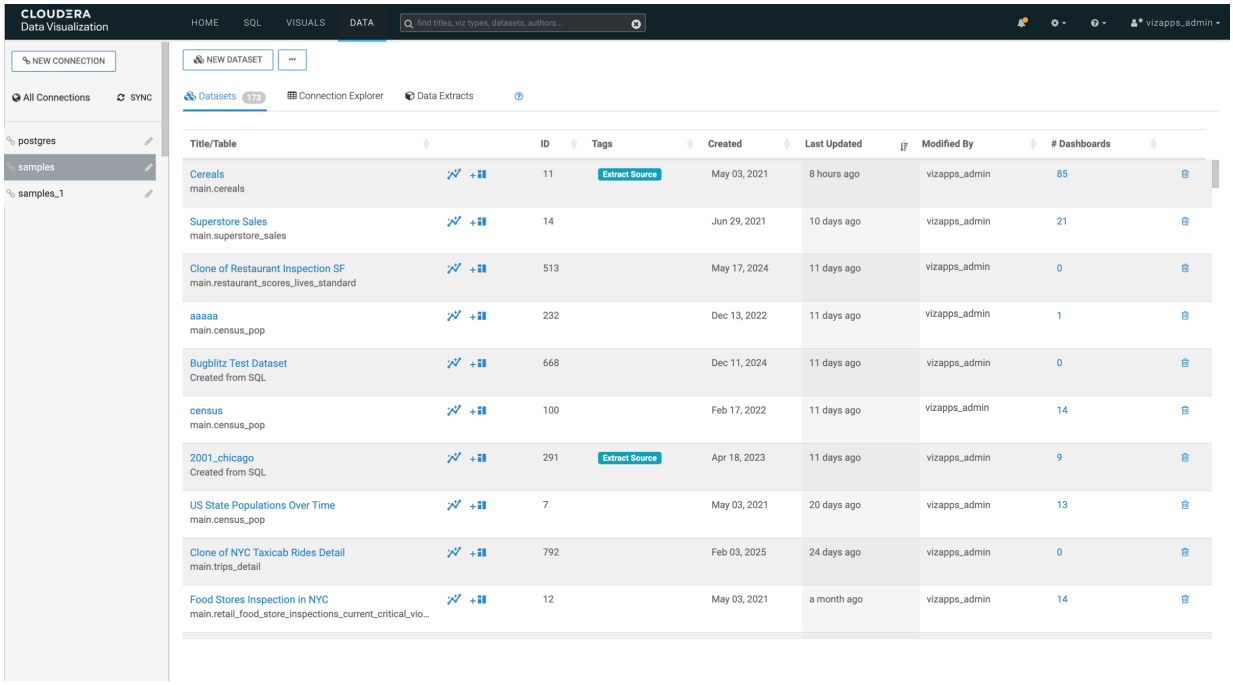
- You must have the Manage data connections privilege or be an administrator to create new connections.
- You can create a connection by:
  - Manually entering connection details.

    If you create a connection, you automatically have the privileges to create and manage datasets on this connection, and also build dashboards and visuals in these datasets. For more information on user privileges, see RBAC permissions.
  - Importing connection details from the JSON configuration of another connection. For instructions on how to obtain the JSON configuration, see Using connection details shared in JSON format.

Procedure

1. On the main navigation bar, click DATA.

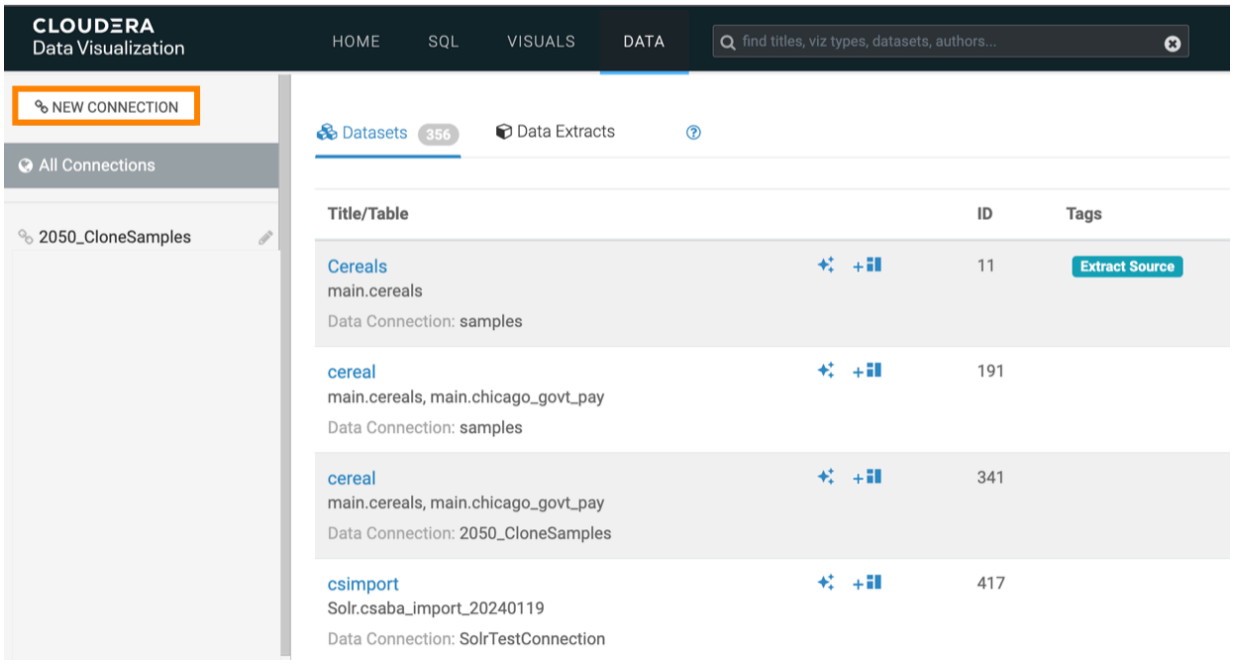
   The DATA interface opens, displaying the Datasets tab.



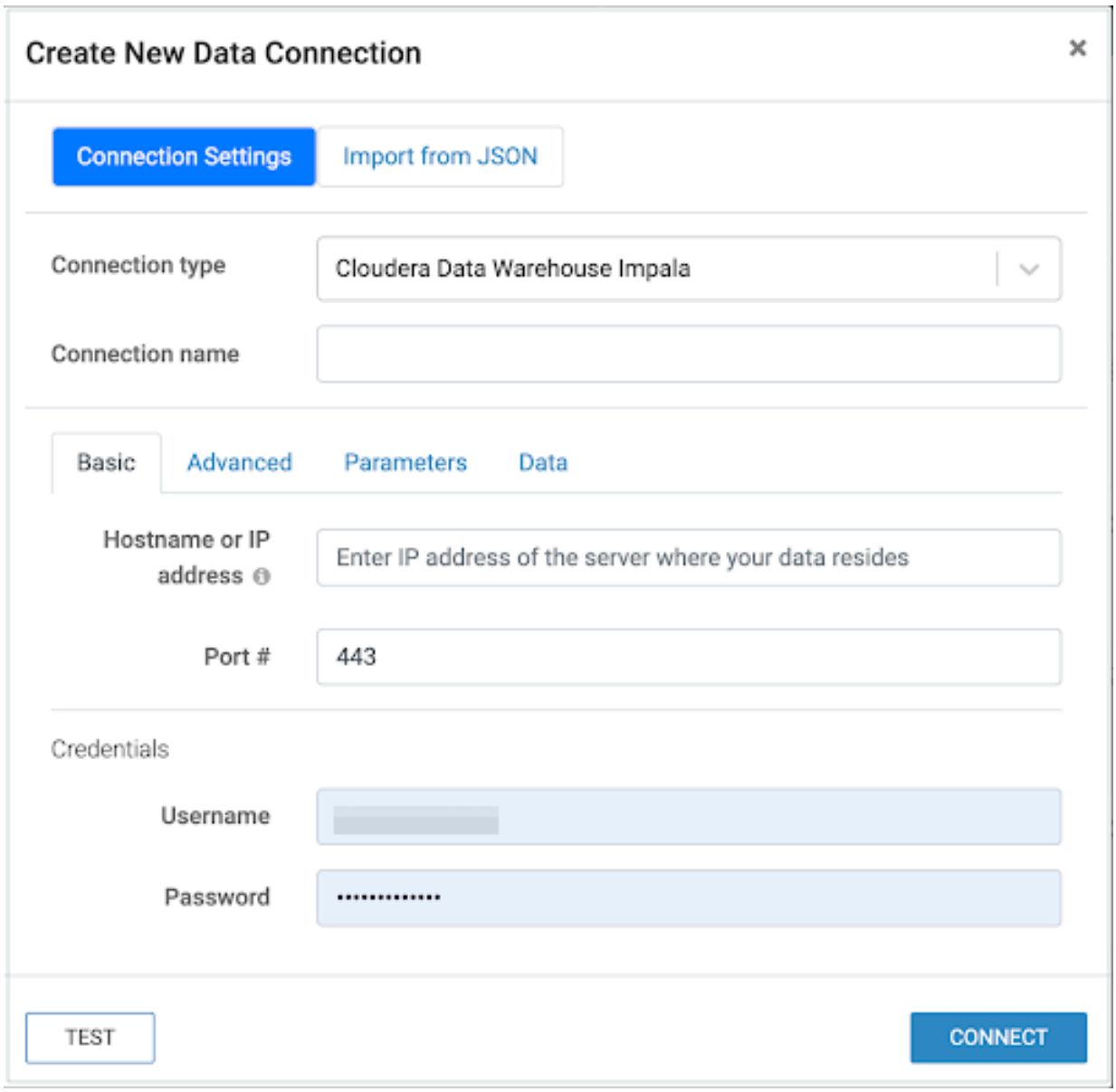
2. On the side menu bar, click NEW CONNECTION.

Note:

The NEW CONNECTION button is only visible to administrators or users with Manage data connections privilege.



The Create New Data Connection modal is displayed.


Create New Data Connection
Connection Settings
Import from JSON
Connection type
Cloudera Data Warehouse Impala
Connection name
Basic
Advanced
Parameters
Data
Hostname or IP address
Enter IP address of the server where your data resides
Port #
443
Credentials
Username
Password
TEST
CONNECT

3. Choose how you want to create the data connection.

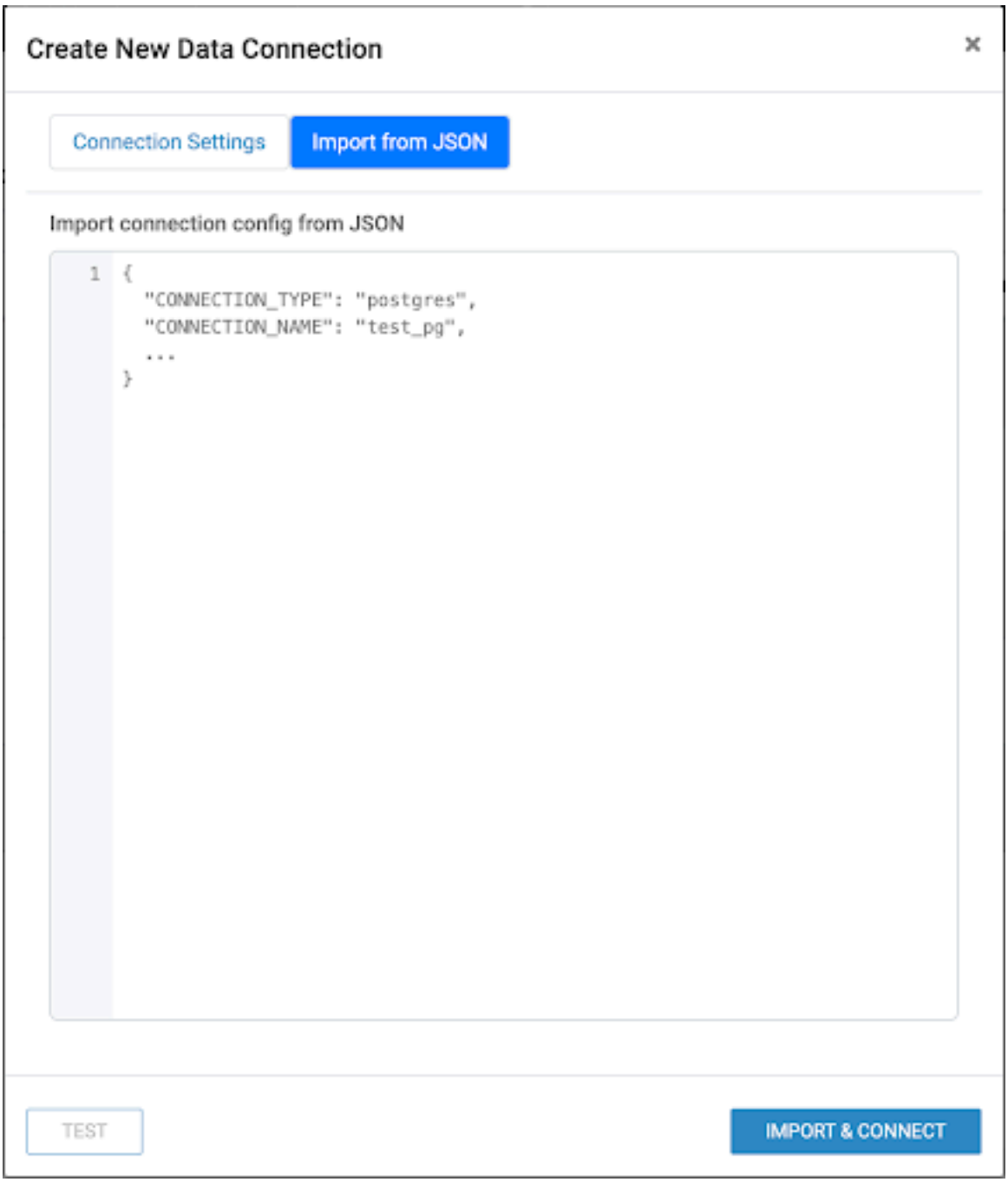
   Option 1: Manually create a connection

   a. On the Connection Settings tab, select the preferred connection type from the drop-down list
   b. Enter a name for the connection.
   c. Fill in the required connection details on all tabs.
   d. Provide your username and password.
   e. Click TEST to verify the connection.

      If any details are missing or invalid, an error message appears. Review and update the fields, and test the connection again.
   f. If the test is successful, click CONNECT to establish the connection.

   Option 2: Import a connection from JSON

   a. Switch to the Import from JSON tab.



b. Paste the connection configuration in JSON format into the text field.

Tip: You can copy the details of an existing data connection from the Data Connection Information modal. For instructions, see *Using connection details shared in JSON format*.

c. Click IMPORT & CONNECT to auto-populate the connection details from the JSON file.

If any configuration details are invalid, an error message appears. You can manually correct the configuration.

Related Information

Creating a Cloudera AI data connection to Impala
Creating a Cloudera AI data connection to a Hive data warehouse
Creating a Cloudera Data Warehouse data connection

Creating a visual

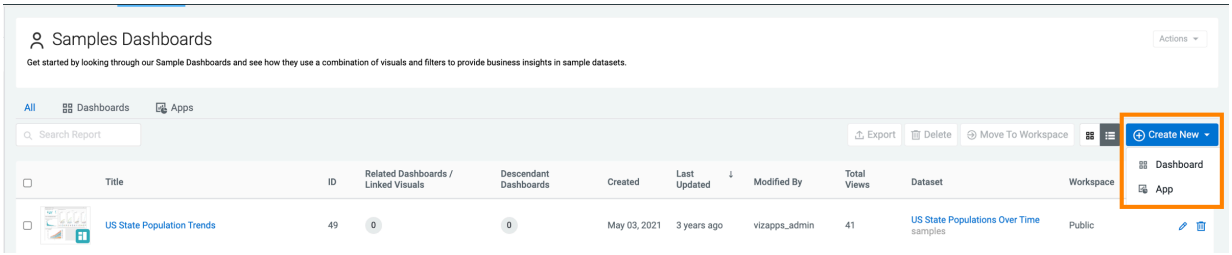
You can easily create new visual representations for your data in Cloudera Data Visualization.

About this task

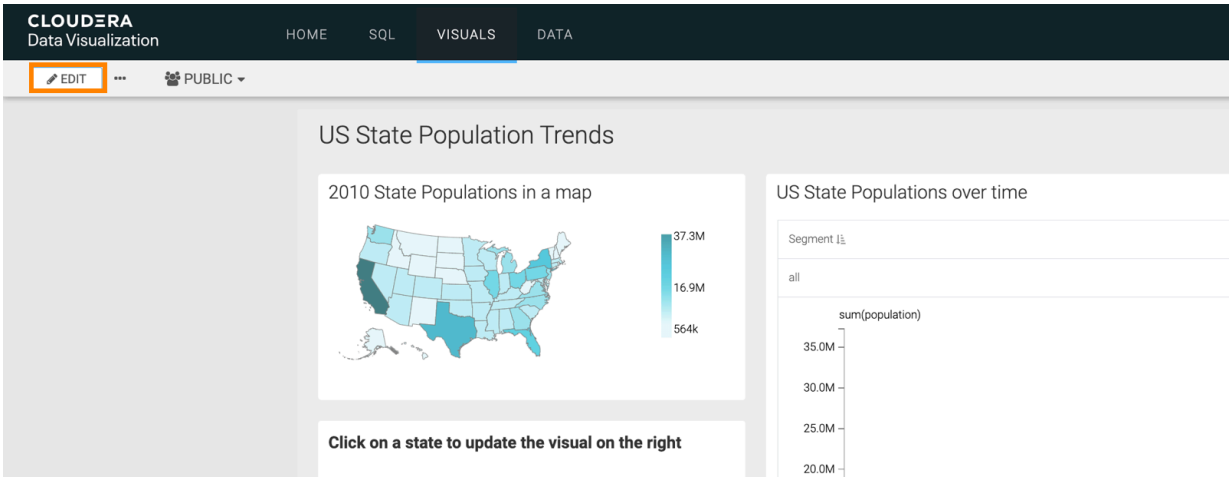
The following steps demonstrate how to create a new visual.

Procedure

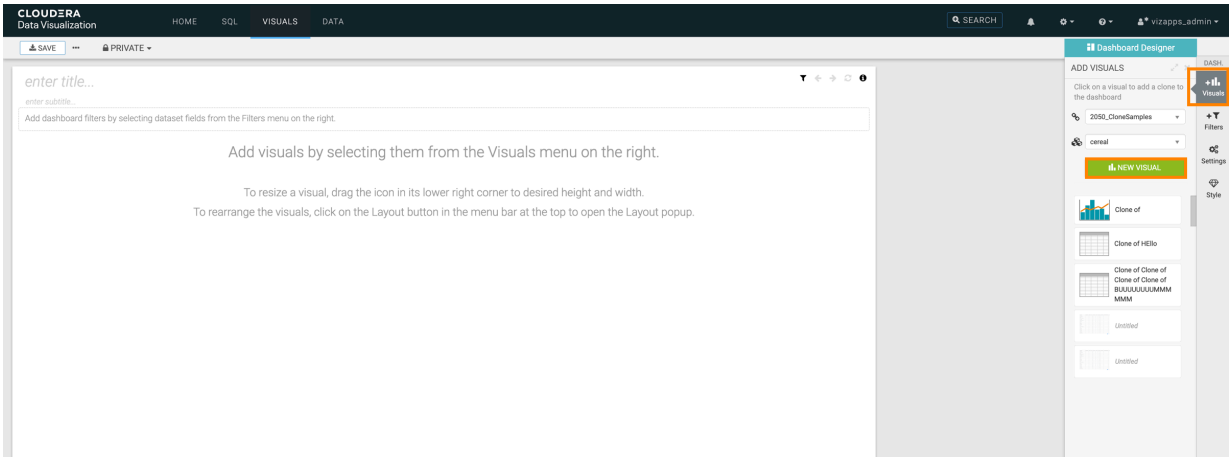
1. On the main navigation bar, click VISUALS.
2. Click Create New Dashboard to open a new, blank dashboard.



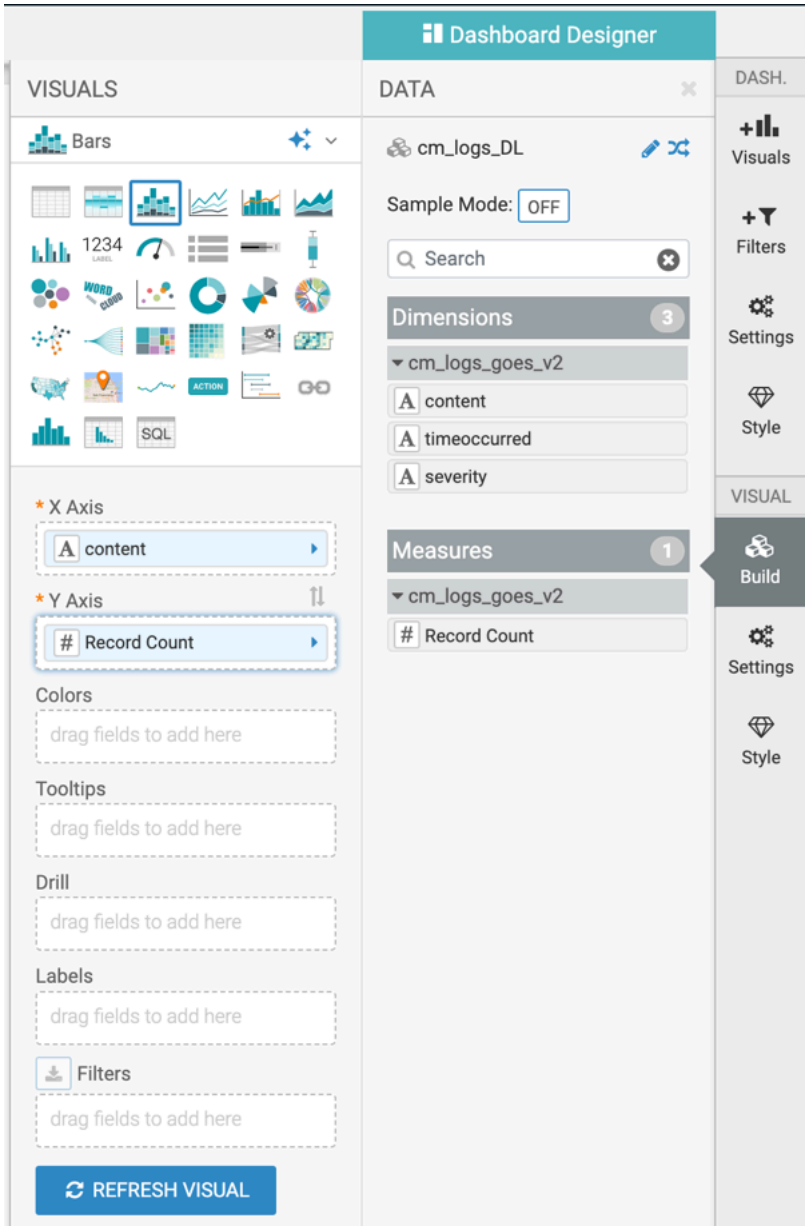
If you already have an existing dashboard and you just want to add a new visual, open the dashboard and click EDIT to make changes to it.



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



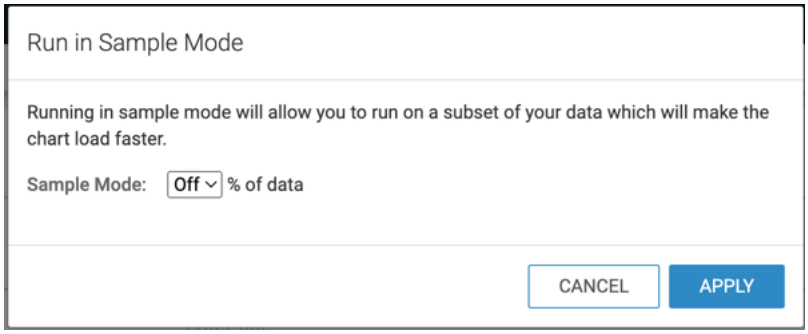
4. The Visual Builder appears.
   By default, a table visual is selected. For more information, see the *Visual Builder* overview.
5. Choose the visual type you want to build.
   In this example, the Bars visual has been selected.



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.



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

   In this example, the following fields have been added:

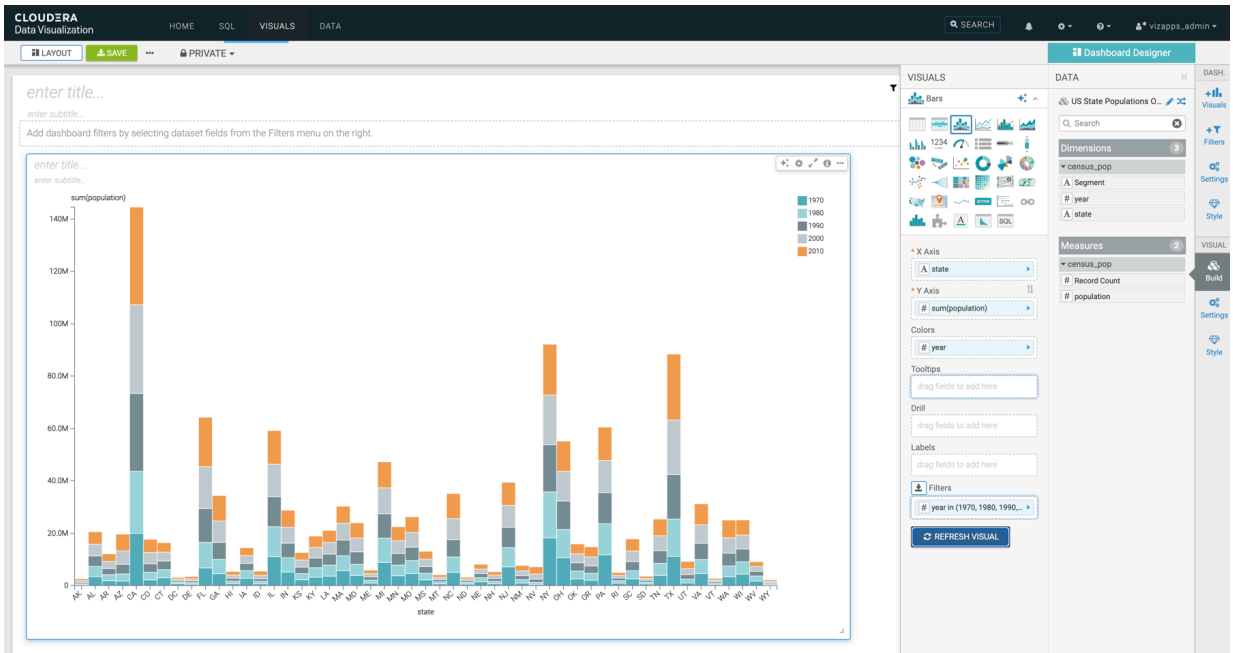
   - state dimension on the X Axis shelf
   - population measure on the Y Axis shelf

     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.

Related Information

Visual type catalog
Creating a dashboard
Creating an application