Cloudera Data Visualization 8.0.9

Starting Cloudera Data Visualization in Cloudera Base on premises

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

https://docs.cloudera.com/

Legal Notice

Contents

Support matrix **4**

Hardware requirements **4**

Installing Cloudera Data Visualization in Cloudera Base on premises **5**
- Adding Cloudera Data Visualization parcel repository to Cloudera Manager 5
- Downloading and adding Cloudera Data Visualization CSD files 8
- Adding Cloudera Data Visualization as a service 9
- Accessing the Cloudera Data Visualization UI 11

Configuring initial administrator users and groups **12**

Specifying a custom Python interpreter **13**

Enabling Knox SSO for Cloudera Data Visualization **14**

Enabling TLS in Cloudera Data Visualization **17**

Installing additional dependencies for Cloudera Data Visualization in Cloudera Base on premises **18**
- Installing Chromium or Google Chrome for visual export and dashboard previews in Cloudera Data Visualization 18
- Installing Oracle Client libraries for Cloudera Data Visualization 19

Support matrix

Before installation, ensure that your environment meets the following requirements.

Cloudera Data Visualization 8.x supports the following platforms and components:

Cloudera Manager	7.13.1, 7.11.3
Cloudera Base on premises	7.3.1 SP2, 7.3.1 SP1, 7.3.1 7.1.9 SP1, 7.1.9
Python	3.11 (bundled with parcel)
Operating systems	RHEL 9.6, 9.5, 9.4, 9.2, 9.1 RHEL 8.10 FIPS1, 8.10, 8.9, 8.8 FIPS1, 8.8 Ubuntu 22.04, 20.4 SLES 15SP5, 15SP4

Note:

RHEL 8.8 and 8.10 with FIPS enabled are supported only when Cloudera Manager and Cloudera Runtime are deployed in FIPS-compliant mode.

Cloudera Data Visualization 8.x supports the following databases:

OracleDB	23c, 21c, 19.19, 19c
PostgreSQL	17, 16, 15, 14, 13
MySQL	8.4, 8.0
MariaDB	10.11, 10.6, 10.5, 10.4

Related Information

Specifying a custom Python interpreter

Hardware requirements

Use the following guidelines to determine the minimum memory, CPU, and storage requirements needed to run Cloudera Data Visualization in Cloudera Base on premises.

Service	Memory	CPU	Disk
Cloudera Data Visualization	Minimum: 8 GB	Minimum: 1 Core (for Django)	Minimum: 32 GB total – 16 GB for the application database and 16 GB for the /tmp directory

The required disk space may increase with cluster size and user workloads. Cluster size refers to the number of nodes in the cluster and workload means the number of queries run and the number of concurrent unique users accessing the application in a given time period.

Cloudera Data Visualization relies heavily on its database, which can grow rapidly because of retained query history. A minimum of 16 GB disk space is required for the database, but additional space may be necessary depending on usage. To maintain performance and manage storage effectively, regularly clean up unused or outdated queries and documents.

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).

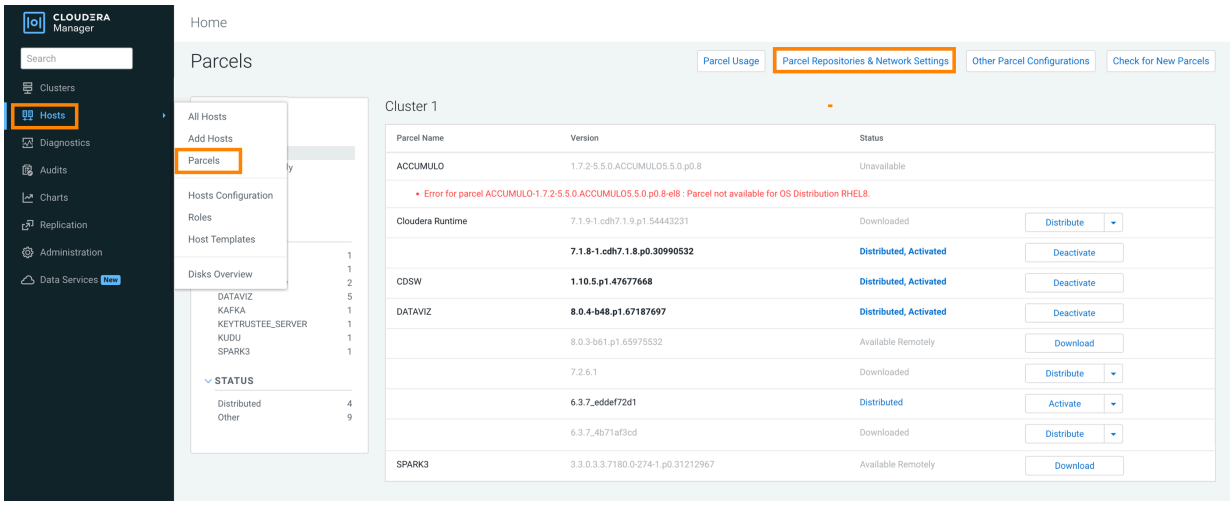
About this task

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

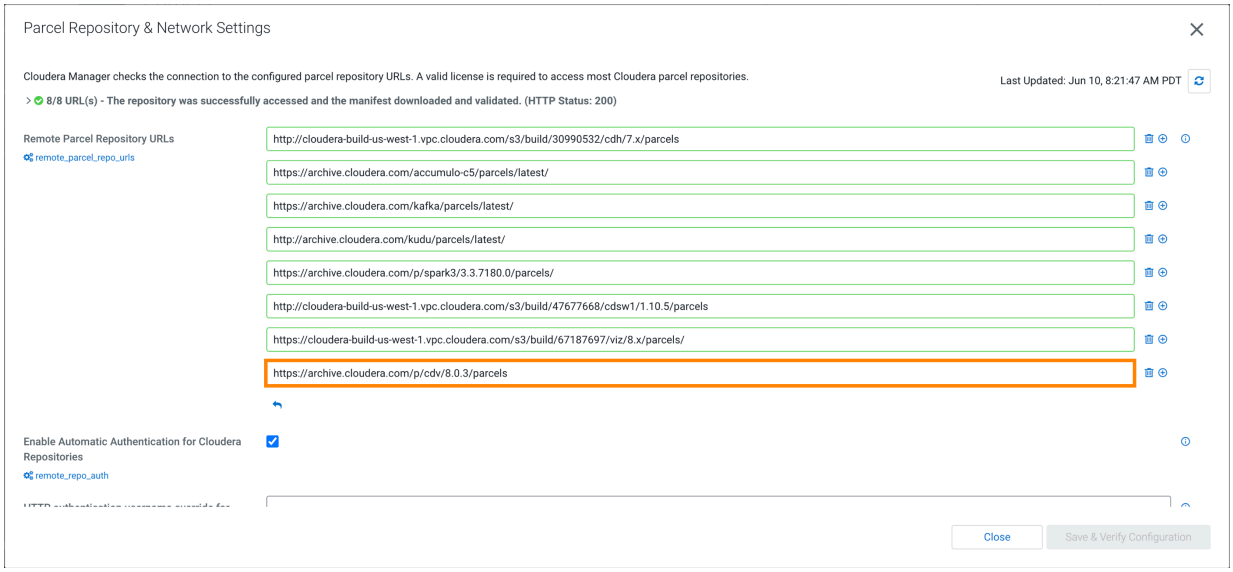
1. Log in to Cloudera Manager.
2. From the left navigation bar, go to Hosts Parcels .

3. Select Parcel Repositories & Network Settings on the Parcels page.



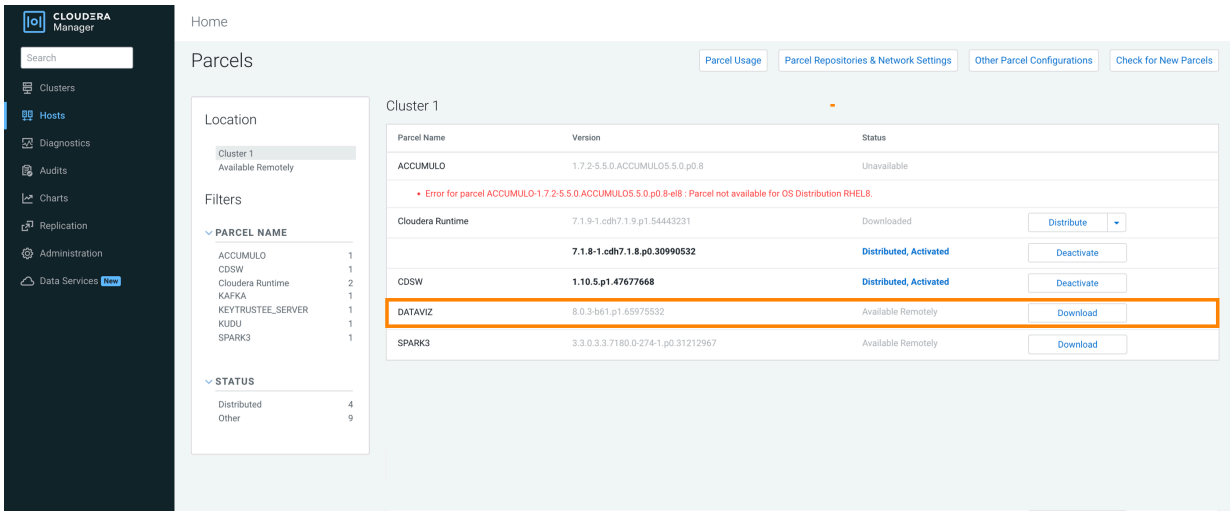
4. On the Parcel Repository & Network Settings page, click + to add a new row in the Remote Parcel Repository URLs list.

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.

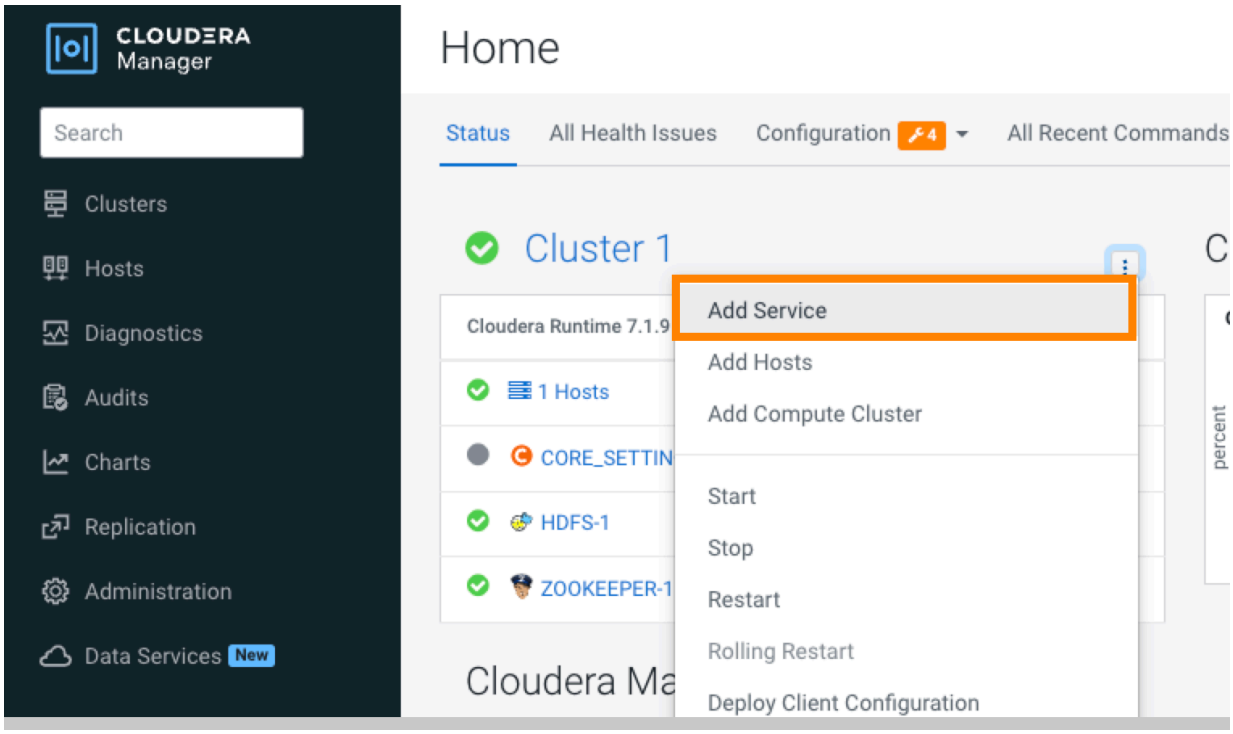
Before you begin

Ensure the following:

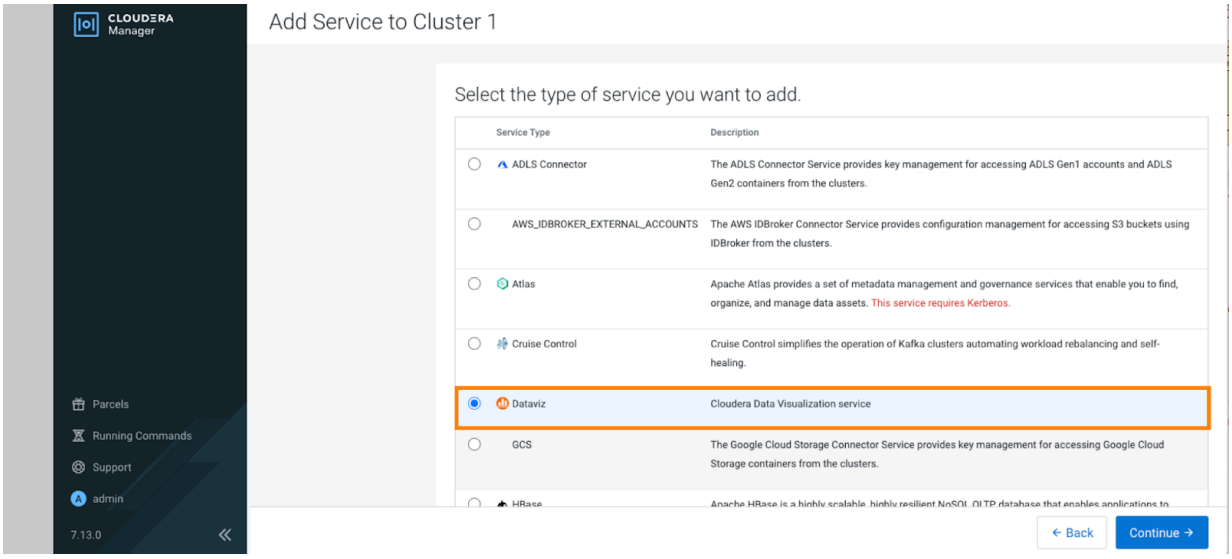
- A Cloudera Base on premises cluster is installed and running.
- You have added the Cloudera Data Visualization CSD files in the /opt/cloudera/csd folder, and Cloudera Manager has been restarted.
- You have added, downloaded, distributed, and activated the Cloudera Data Visualization parcel in Cloudera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

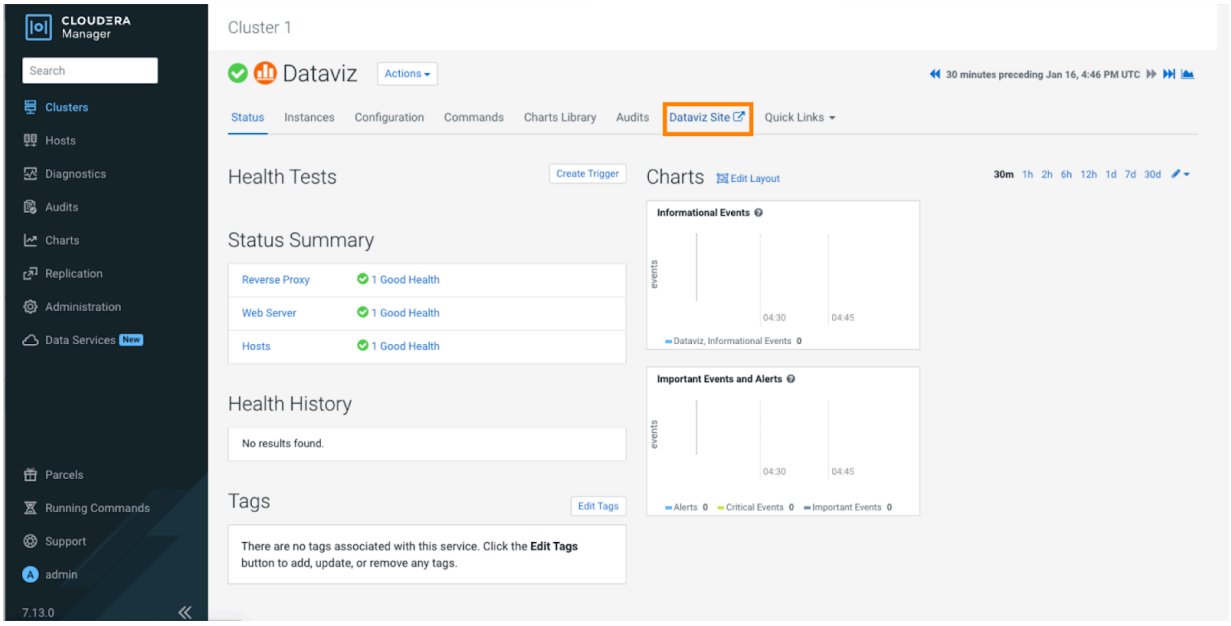
Related Information
Adding a service

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

Related Information
Changing user passwords

Configuring initial administrator users and groups

When you add the Cloudera Data Visualization service in Cloudera Manager, you can configure which users and groups are granted administrator privileges in the application. Administrator access can be assigned either by explicitly specifying admin users and groups, or, when using LDAP, by assigning administrator privileges based on LDAP group membership.

About this task

This section explains how to explicitly configure administrator users and groups using Cloudera Manager. For information about assigning administrator privileges using LDAP user flags, see *Configuring user authentication using LDAP*.

Before you begin

- The Cloudera Data Visualization service is already added to the Cloudera Base on premises cluster.
- Authentication is configured (for example, LDAP or SSO).
- User and group information is available through your configured authentication provider.

Note: Group information is required only if you plan to assign administrator privileges based on groups.

Procedure

1. In Cloudera Manager, go to Configuration .
2. Locate the following properties (use the search field if needed).
   - Admin Users maps to the ADMIN_USERS environment variable
   - Admin Groups maps to the ADMIN_GROUPS environment variable
3. Configure the values as required.
   - **Admin Users**

     Enter a comma#separated list of usernames that should be granted administrator privileges.

     Example: alice,bob
   - **Admin Groups**

     Enter a comma#separated list of group names whose members should be granted administrator privileges.

     Example: cdv-admins,platform-admins
4. Click Save.
5. If prompted, restart the Cloudera Data Visualization service to apply the changes.

Results

At each login to Cloudera Data Visualization:

- If the user's username matches an entry in the list of Admin Users, the user is granted administrator privileges.
- If Admin Groups is configured and the user belongs to one of the listed groups, the user is granted administrator privileges.

Security considerations

- Grant administrator privileges only to users or groups that require full platform access. The Administrator role provides unrestricted control over Cloudera Data Visualization, including access to all datasets, visualizations, and system settings.

- Use Admin Groups instead of individual users wherever possible, as group-based administration simplifies access management and reduces the risk of orphaned admin accounts.
- Removing a user or group from these configuration properties does not automatically revoke existing administrator privileges. To demote an administrator, update the user's role manually in the Cloudera Data Visualization UI or through the appropriate API.

Related Information

Configuring user authentication using LDAP

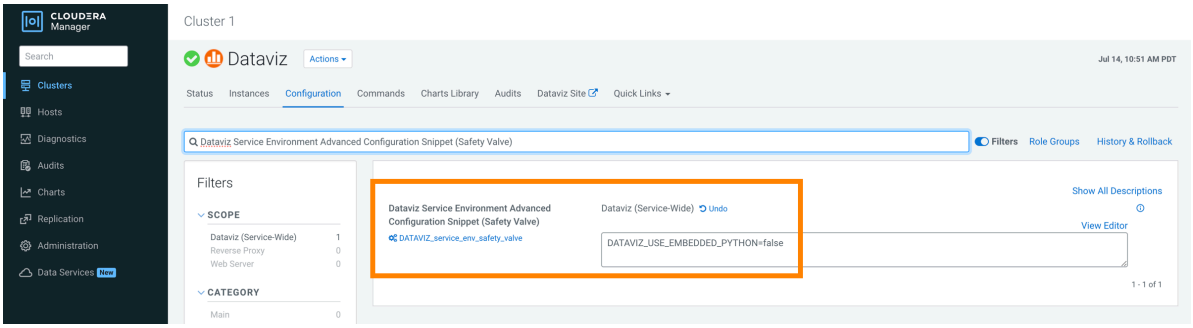
Specifying a custom Python interpreter

Cloudera Data Visualization includes an embedded Python interpreter by default. However, you can configure the system to use the OS-provided Python or a custom Python build instead. This provides greater flexibility and can improve performance when compatible Python versions are available on your system.

Procedure

1. To use a non-embedded Python interpreter:
   a) In Cloudera Manager, go to Clusters and select the Dataviz service.
   b) Click the Configuration tab.
   c) Set the following environment variable in the Dataviz Service Environment Advanced Configuration Snippet (Safety Valve) configuration to disable the embedded Python.

      DATAVIZ_USE_EMBEDDED_PYTHON=false

      This enables Cloudera Data Visualization to use the system's default, OS-provided Python interpreter.
   d) Restart the Dataviz service to apply the changes.
2. To use a custom Python build:

   If you want to use a specific Python installation (for example one you compiled from source or installed with pyenv), follow these additional steps:

   a) Set the PYTHONHOME environment variable to point to the root directory of the Python installation you want to use.
   b) Configure the library paths by making sure that the following directories are accessible:

      - python/lib
      - sqlite/lib (only if you compiled SQLite with this Python build)

      You can do this by:

      - Adding them to the LD_LIBRARY_PATH environment variable, **or**
      - Registering them with ldconfig if you are using dynamically linked executables.

Enabling Knox SSO for Cloudera Data Visualization

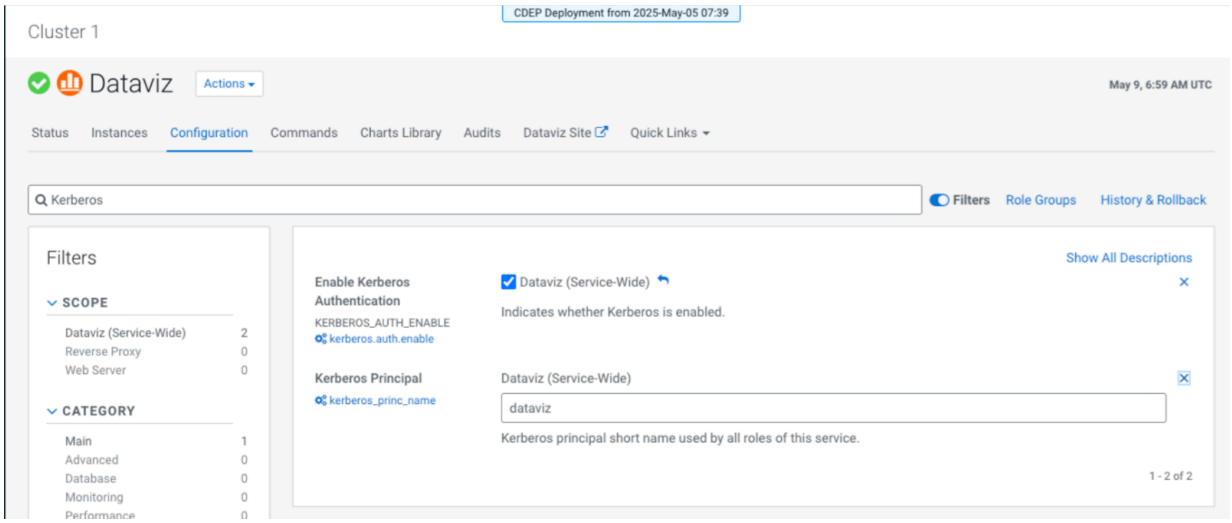
You can configure Cloudera Data Visualization to use Knox Single Sign-On (SSO) if Knox is available in your cluster. The setup requires a Kerberized environment and includes configuring both Cloudera Data Visualization and Knox.

Before you begin

- The cluster must be Kerberized.
- Cloudera Data Visualization must be installed and Knox must be available on your cluster.
- Kerberos authentication must be enabled for Cloudera Data Visualization.

  If not already configured, enable Kerberos authentication for Cloudera Data Visualization.

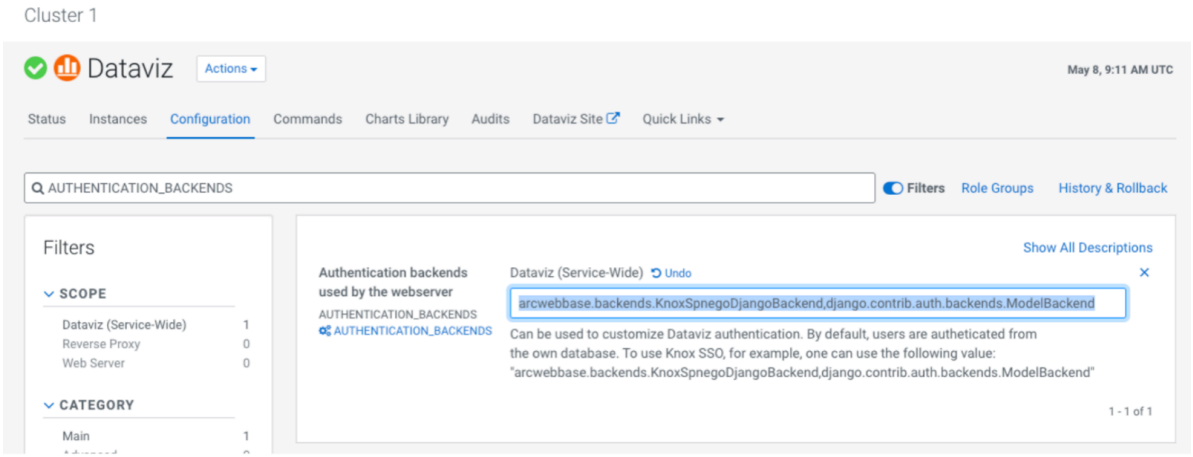
  1. In Cloudera Manager, go to Clusters and select the Dataviz service.
  2. Click the Configuration tab.
  3. In the Search bar, type kerberos to filter the relevant settings.
  4. Find the Enable Kerberos Authentication property.
  5. Check the Dataviz (Service-Wide) box next to Enable Kerberos Authentication.
  6. For Kerberos Principal, enter the Kerberos username (principal short name) the Cloudera Data Visualization service should use.
  7. Click Save Changes.
  8. Restart the Dataviz service for the changes to take effect.



Kerberos is now enabled for Cloudera Data Visualization, which is a prerequisite for enabling Knox SSO.

Procedure

1. Configure Cloudera Data Visualization for Knox SSO.
   a) In Cloudera Manager, go to Clusters and select the Dataviz service.
   b) Click the Configuration tab.
   c) Search for the Authentication backends used by the webserver property and set its value to:
   arcwebbase.backends.KnoxSpnegoDjangoBackend,django.contrib.auth.backends.ModelBackend



Note:

This configuration enables Knox SSO while still allowing fallback login at /arc/apps/login. To enforce Knox authentication only, remove the fallback by setting AUTHENTICATION_BACKENDS to arcwebbase.backends.KnoxSpnegoDjangoBackend

   d) Restart the Dataviz service to apply the changes.
2. Make the Dataviz service definition available to Knox.
   a) Locate DATAVIZ-KNOX-*[***VERSION***]*.tgz on theCloudera archive site and download it to the node where the Knox service is running.

   This file is co-located with the Cloudera Data Visualization CSD on the archive site and must be manually downloaded to each Knox host.

   For example, to download DATAVIZ-KNOX-8.0.4-b47.p1.67141340.tgz to /tmp, run the following command:

```
cd /tmp;
wget https://archive.cloudera.com/p/cdv/8.0.4/redhat8/yum/DATAVIZ-KNOX-
8.0.4-b47.p1.67141340.tgz
```

   b) SSH to the Knox node.

```
ssh [***KNOX-NODE***]
```

   Replace *[***KNOX-NODE***]* with the hostname or IP address of your Knox node.
   c) Navigate to the temp directory and extract the archive to the Knox services directory.

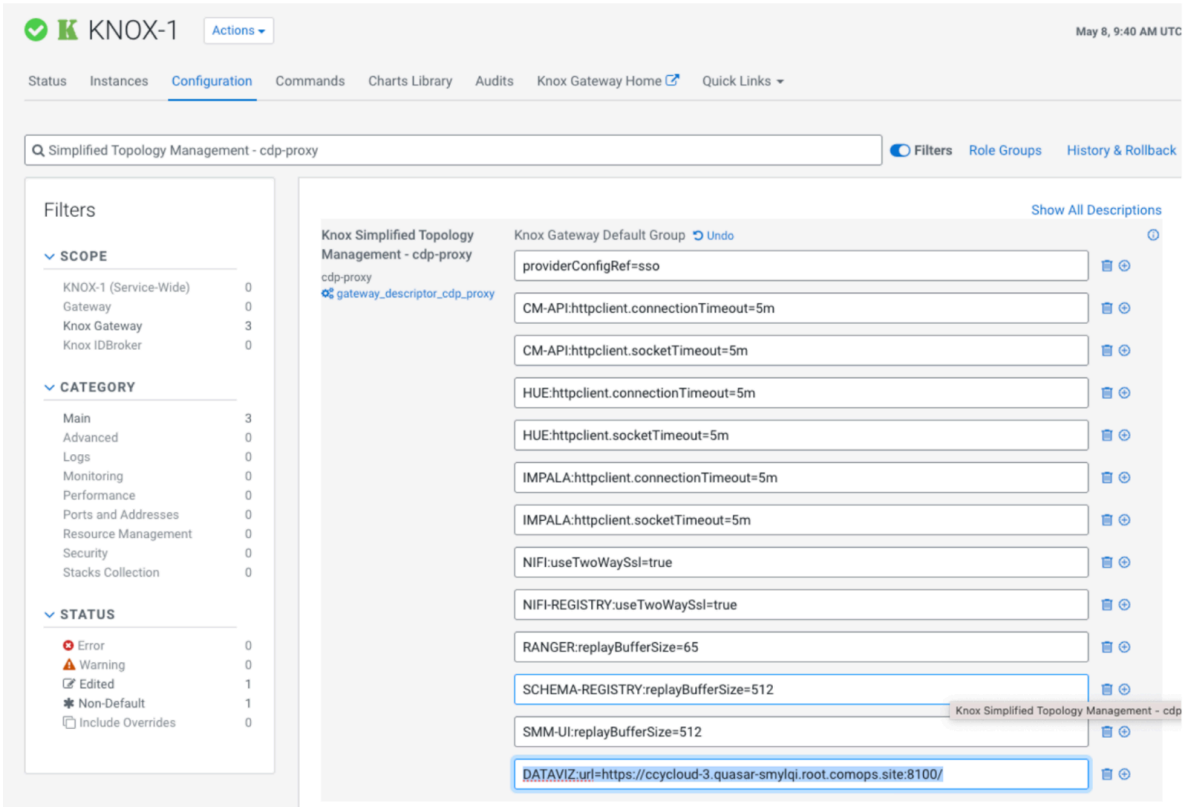
```
cd /tmp
tar xzf /tmp/DATAVIZ-KNOX-[***VERSION***].tgz -C /opt/cloudera/parcels/
CDH/lib/knox/data/services
```

   Replace *[***VERSION***]* with the version number of the DATAVIZ-KNOX archive you downloaded.
   d) Verify that the Cloudera Data Visualization service was extracted correctly.

```
ls -la /opt/cloudera/parcels/CDH/lib/knox/data/services | grep dataviz
```

3. Update the Knox cdp-proxy topology.
   a) In Cloudera Manager, go to Clusters and select the Knox service.
   b) Click the Configuration tab.
   c) Search for the Knox Simplified Topology Management - cdp-proxy property.
   d) Edit the cdp-proxy topology by adding a new service entry for Cloudera Data Visualization.

```
DATAVIZ:url=[***DATAVIZ-SERVICE-URL***]
```

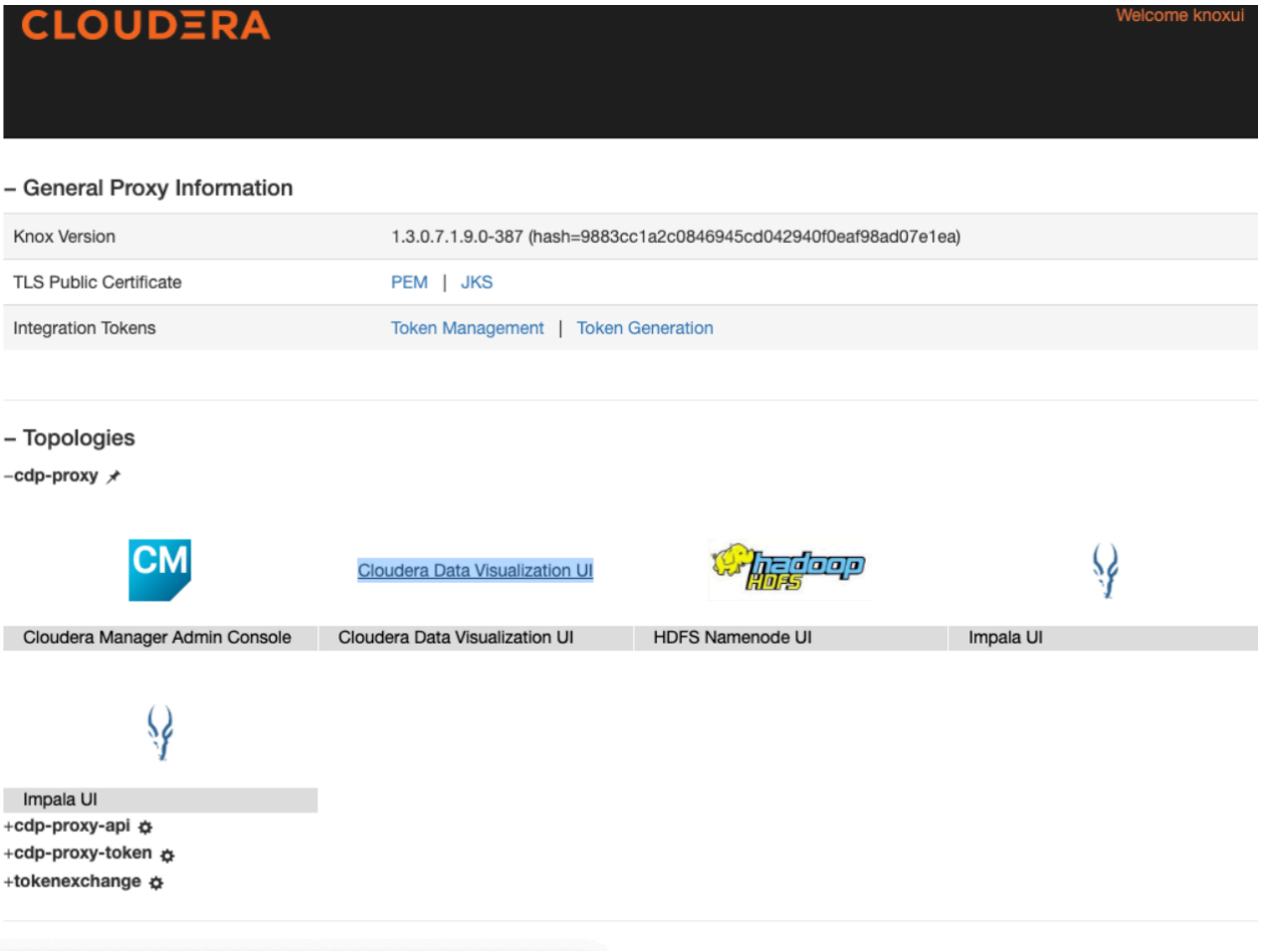


   e) In Cloudera Manager, locate Knox Simplified Topology Management .
   f) Save the topology changes.
   g) Restart the Knox service.

Results

Once configuration is complete, Cloudera Data Visualization will be accessible through Knox at: https://*[***KNOX-HOST**]*:*[***PORT***]*/gateway/cdp-proxy/dataviz/

Alternatively, you can also access Cloudera Data Visualization using the Knox Gateway home page.



Access will be granted based on the identity authenticated by Knox.

Enabling TLS in Cloudera Data Visualization

Cloudera Data Visualization supports both Auto-TLS and manual TLS configuration for securing communication with the Cloudera Data Visualization web server and reverse proxy. Transport Layer Security (TLS) is a cryptographic protocol that ensures secure data transmission between components.

TLS versions

- TLS 1.2
- TLS 1.3

TLS configuration options

Auto-TLS

When Cloudera Manager Auto-TLS is enabled in your Cloudera Base on premises environment, Cloudera Data Visualization automatically leverages the Auto-TLS certificates and keys for secure communication. No additional manual configuration is required

Manual TLS

You can configure TLS manually by specifying certificates and keys through dedicated Cloudera Manager parameters.

- The location of key and certificate files can be set in the Cloudera Data Visualization service configuration.
- Password-protected private keys are supported.

Managing TLS protocols and ciphers

Administrators can control the TLS protocols and (for TLS 1.2) the ciphers used by Cloudera Data Visualization's underlying Caddy web server.

Setting	Description	Example
DATAVIZ_TLS_PROTOCOLS	Minimum and maximum TLS versions supported By default, Caddy supports TLS 1.2 (minimum) and TLS 1.3 (maximum).	DATAVIZ_TLS_PROTOCOLS="tls1.2 tls1.2" (forces TLS 1.2 only)
DATAVIZ_TLS_CIPHERS	List of allowed cipher names for TLS 1.2 connections Note: TLS 1.3 ciphers cannot be configured.	DATAVIZ_TLS_CIPHERS="TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384" (a space-separated list of cipher names as recognized by the Go standard library)

This configuration enables compliance with organizational or regulatory requirements for TLS versions and cipher strength.

For more information on TLS protocols and ciphers, see the Caddy documentation.

Limitations

- Custom cipher management is supported only for TLS 1.2. TLS 1.3 cipher configuration is not supported.
- At least one supported TLS protocol must be enabled for Cloudera Data Visualization startup to succeed.

Installing additional dependencies for Cloudera Data Visualization in Cloudera Base on premises

To enable certain features and integration, Cloudera Data Visualization in Cloudera Base on premises requires additional system dependencies to be installed manually on the hosts running Cloudera Data Visualization.

This section describes how to install and configure the following dependencies:

- Chromium – required for exporting visual content and displaying dashboard thumbnails.
- Oracle Client Libraries – required if you plan to use Oracle as the external metadata store.

Installing Chromium or Google Chrome for visual export and dashboard previews in Cloudera Data Visualization

Cloudera Data Visualization relies on Chromium to render and export visual outputs.

This enables:

- Downloading dashboards as PNG or PDF files
- Displaying thumbnail previews of dashboards

By default Cloudera Data Visualization expects the Chromium binary to be located at: /usr/lib64/chromium-browser/headless_shell

If Chromium is available at this location, no additional configuration is required. If not, you need to install Chromium on the hosts running Cloudera Data Visualization.

Installing Chromium on Red Hat–like systems

Use the following command to install Chromium on RHEL, CentOS, or similar systems: sudo yum install chromium-headless

This installs Chromium at the expected default location used by Cloudera Data Visualization.

Installing Google Chrome on Ubuntu

You can install Google Chrome from the official Google repository by following Google's installation guide.

After installation, the binary is located at: /usr/bin/google-chrome

Using a custom Chromium path

If Chromium is installed at a different location, configure Cloudera Data Visualization to use that path by setting the CHROMIUM_PATH environment variable.

1. In Cloudera Manager, go to Clusters and select the Dataviz service.
2. Go to the Configuration tab.
3. Search for Dataviz Service Environment Advanced Configuration Snippet (Safety Valve).
4. Switch to View as Text and add the following line.

   CHROMIUM_PATH=/path/to/your/custom/headless_shell
5. Save the changes and restart the Dataviz service for the changes to take effect.

Installing Oracle Client libraries for Cloudera Data Visualization

Cloudera Data Visualization supports the use of external databases as a metastore. When using Oracle as the metastore, additional setup steps are required beyond the ones needed for MySQL/MariaDB or PostgreSQL, as Cloudera Data Visualization does not include an Oracle database connector.

About this task

To enable connectivity to an Oracle metastore, you must install the Oracle client libraries on each host running the Cloudera Data Visualization webserver.

Procedure

1. Download the appropriate Oracle Instant Client libraries for your operating system from the official Oracle website.
2. Follow the installation instructions provided on the Installing Oracle Instant Client page.