

Starting Cloudera Data Visualization in Cloudera Machine Learning

Date published: 2020-10-30

Date modified: 2024-10-30



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 a Cloudera Data Visualization application in Cloudera Machine

Learning.....	4
Creating a Cloudera Machine Learning project with Cloudera Data Visualization Runtime.....	4
Adding Cloudera Data Visualization Runtime to a Cloudera Machine Learning project.....	4
Creating a Cloudera Data Visualization application in Cloudera Machine Learning.....	5

Using Exploratory Data Science and Visualization in Cloudera Machine

Learning.....	6
----------------------	----------

Upgrading Cloudera Data Visualization in Cloudera Machine

Learning/Cloudera Data Science Workbench.....	7
--	----------

Restoring a Cloudera Data Visualization version in Cloudera Machine

Learning/Cloudera Data Science Workbench.....	8
--	----------

Deploying a Cloudera Data Visualization application in Cloudera Machine Learning

Learn how to deploy Cloudera Data Visualization in Cloudera Machine Learning.

Creating a Cloudera Machine Learning project with Cloudera Data Visualization Runtime

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

About this task

If you know that your project is going 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 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 a 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 Cloudera Data Visualization Runtime to a Cloudera Machine Learning project

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

About this task

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

Procedure

1. Click Projects on the left sidebar of the 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 Cloudera Data Visualization 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 Cloudera Data Visualization application.

What to do next

Create a Cloudera Data Visualization application.

Creating a Cloudera Data Visualization application in Cloudera Machine Learning

Learn how to create a Cloudera Data Visualization application in Cloudera Machine Learning 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

The following steps will guide you on how to create a Cloudera Data Visualization application in a Cloudera Machine Learning project.



Important:

Each Cloudera Machine Learning project can host only one standalone Cloudera Data Visualization application.

Before you begin

Ensure that a Cloudera Data Visualization Runtime is available in the Cloudera Machine Learning project where you want 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 Machine Learning 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.
- Subdomain – Enter a subdomain that will be used to construct the URL for the web application. Ensure it only contains URL-friendly characters.
- Description – Add a description for the application.
- Script – Select the path to the startup script.



Note: Use the script located at: `/opt/vizapps/tools/arcviz/startup_app.py`

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



Note: The selected edition determines the version of the Runtime variant.

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 fully operational.

You can restart, stop, or delete a Cloudera Machine Learning application from 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.

Using Exploratory Data Science and Visualization in Cloudera Machine Learning

Learn how to deploy Cloudera Data Visualization in Cloudera Machine Learning.

Exploratory Data Science and Visualization makes it simple for data scientists to get started on a data science project. This experience automatically discovers the data sources available to you from within the standard Cloudera Machine Learning user interface and provides various features for Exploratory Data Analysis.

From the Data tab in Cloudera Machine Learning you can:

- Connect to data sources that are available in your project
- Explore the data with SQL to understand its basic shape and characteristics
- Create named datasets that you can reference later
- Create visualizations of the data to understand its properties
- Create dashboards that you can share with your team

For prerequisites and information on how to work with data discovery and visualization in Cloudera Machine Learning, see the [Exploratory Data Science and Visualization documentation](#)

Upgrading Cloudera Data Visualization in Cloudera Machine Learning/Cloudera Data Science Workbench

Cloudera recommends following the Cloudera Data Visualization release cadence by upgrading to the latest version as it becomes available. This practice ensures that your Cloudera Data Visualization instances benefit from the latest new features, security enhancements, and bug fixes.

About this task

This guide outlines the available options performing an upgrade and the steps of the upgrade process. If you need to downgrade, follow the [Restoring Cloudera Data Visualization in Cloudera Machine Learning/Cloudera Data Science Workbench](#) procedure using the backup you created earlier.

Before you begin

- Before initiating the upgrade, review the [Known issues and limitations in Cloudera Data Visualization](#) in the Release Notes for both the currently running version and the version to which you plan to upgrade to. This helps you identify potential challenges, limitations, or specific actions required for a successful upgrade.
- If you are using an external database like MySQL, MariaDB, Postgres, or Oracle, ensure you back up the database and restore it as a new database that the new Cloudera Data Visualization instance can connect to.

Procedure

1. Stop the application.

On the Applications page in Cloudera Machine Learning/CDSW, click **Actions Stop** to stop the selected Cloudera Data Visualization application.

2. Back up the SQLite metadata database.

a) Navigate to your project's Overview page and click **New Session**.

b) When your session is running, click **>_ Terminal Access** above the session log pane to open a terminal window.

The default working directory is `/home/cdsw`, where all your project files reside.

c) Navigate to `/home/cdsw/.arc`.

This directory contains the Cloudera Data Visualization logs, files, settings, and the `arcviz.db` database.

d) Run the following command to create a copy of your current Cloudera Data Visualization metadata.

```
sqlite3 arcviz.db ".backup    arcviz.db-BACKUP-['***version***']"
```

You can use this backup in case you need to revert to a previous Cloudera Data Visualization version.

3. Navigate to the Application Settings page.

4. Locate the runtime configuration option, where you can select the desired runtime for your application.

5. Select the runtime you want to set for your application from the available options.

6. Click Update Application to save the change.

The application will restart automatically with the newly selected runtime.



Note: Allow some time for the restart process to complete.

Restoring a Cloudera Data Visualization version in Cloudera Machine Learning/Cloudera Data Science Workbench

Restoring a previous version means reverting your Cloudera Data Visualization application to an earlier version. You can restore your system to a known stable state, especially when issues arise during an upgrade or if the new version introduces major problems that impact functionality, and you need to return to the previous Cloudera Data Visualization version to maintain system functionality. If you decide to downgrade, follow the steps outlined below.

Before you begin

To restore a previous version, you should have either connected your system to an external database (for example: MySQL, MariaDB, Postgres, Oracle) or created a backup of the SQLite metadata database.



Note:

If you are using an external database like MySQL, MariaDB, Postgres, or Oracle, you must back up the database before the upgrade and restore it as a new database for the Cloudera Data Visualization instance to connect to after the downgrade.

Procedure

1. Stop the application.

On the Applications page in Cloudera Machine Learning/Cloudera Data Science Workbench, click **Actions Stop** to stop the selected Cloudera Data Visualization application.

2. Back up the SQLite metadata database.

- Navigate to your project's Overview page and click **New Session**.
- When your session is running, click **>_ Terminal Access** above the session log pane to open a terminal window.

The default working directory is `/home/cdsw`, where all your project files reside.

- Navigate to `/home/cdsw/.arc`.

This directory contains the Cloudera Data Visualization logs, files, settings, and the `arcviz.db` database.

- Run the following command to create a copy of your current Cloudera Data Visualization metadata.
`sqlite3 arcviz.db ".backup arcviz.db-BACKUP-[***version***]"`

You can use this backup when reverting to a previous Cloudera Data Visualization version.

- Create a new Cloudera Data Visualization application in your Cloudera Machine Learning/Cloudera Data Science Workbench project.
- Copy the backed-up SQLite database to `/home/cdsw/.arc/` in the new project.
- Once the backup is restored, start the new Cloudera Data Visualization application.