Cloudera Data Visualization 7.2.5

Starting Data Visualization in CML

Date published: 2020-10-30
Date modified: 2024-06-28

CLOUD≡RA

https://docs.cloudera.com/

Legal Notice

Contents

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.

Using Exploratory Data Science and Visualization in CML

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

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 CML user interface and provides various features for Exploratory Data Analysis.

From the Data tab in CML 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 CML, see the Exploratory Data Science and Visualization documentation

Upgrading Cloudera Data Visualization in CML/CDSW

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

About this task
This guide walks you through the considerations for performing an upgrade, the available options, and the steps of the upgrade process.

Note:

If you are using an external database like MySQL, MariaDB, Postgres, or Oracle, you need to make a backup of the database and restore it as a new database that the new instance of CDV can connect to.

Before you begin

Before initiating the upgrade, review the Known issues and limitations in Data Visualization in the Release Notes for both the currently running version and the version to which you are planning to upgrade. This step helps you understand any potential challenges, limitations, or specific actions required for a successful upgrade.

Procedure

1. Stop the application.

On the Applications page in CML/CDSW, click Actions Stop to stop the selected CDV application.

2. Back up the SQLite metadata database.

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

    b) When your session is running, click >_ Terminal Access above the session log pane.

       A terminal window opens, and its default working directory is /home/cdsw, where all your project files reside.

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

       This directory contains the CDV logs, files, settings, and the arcviz.db database.

    d) Run the following command to make a copy of your current CDV metadata.

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

       You can use this backup in case you need to revert to a previous CDV version.

3. Navigate to the Application Settings page.

4. Locate the runtime configuration option.

    This is where you can select the desired runtime for your application.

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

    You may have different runtime options to select from.

6. Click Update Application to save the change.

    The application will be restarted automatically with the newly selected runtime.

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

Restoring a Cloudera Data Visualization version in CML/CDSW

Restoring a previous version means reverting your Cloudera Data Visualization (CDV) application to a previous version. You can restore your system to a known good state when you encounter any problems during upgrade and the upgrade process fails, or when there are major issues with the new software version and you need to return to the previous CDV version to maintain system functionality.

About this task

To restore a previous CDV version, you should either be connecting to an external database like MySQL or you should have created a backup of the SQLite metadata database.

> **Note:**
>
> If you are using an external database like MySQL, MariaDB, Postgres, or Oracle, you need to make a backup of the database and restore it as a new database that the new instance of CDV can connect to.

Procedure

1. Stop the application.

    On the Applications page in CML/CDSW, click Actions Stop to stop the selected CDV application.

2. Back up the SQLite metadata database.

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

    b) When your session is running, click >_ Terminal Access above the session log pane.

       A terminal window opens, and its default working directory is /home/cdsw, where all your project files reside.

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

       This directory contains the CDV logs, files, settings, and the arcviz.db database.

    d) Run the following command to make a copy of your current CDV metadata.

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

       You can use this backup in case you need to revert to a previous CDV version.

3. Create a new CDV application.

4. Copy the SQLite database to /home/cdsw/.arc/ in your new project.

5. Start the new application.