

..

Accessing Data Visualization in CDSW

Date published: 2020-10-30

Date modified: 2022-09-21

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.

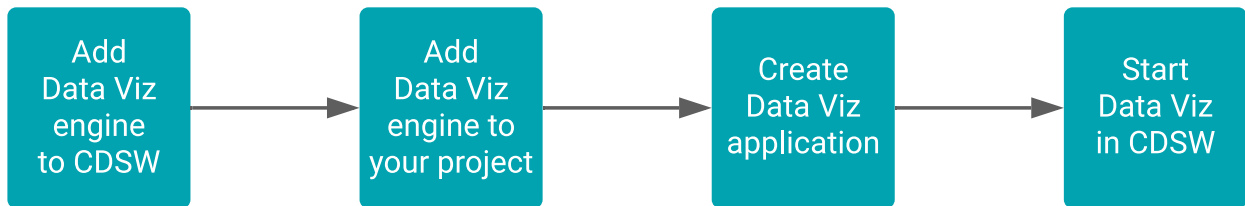
Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES 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

Accessing Data Visualization with ML Runtime.....	4
Adding a custom Data Visualization engine to CDSW.....	4
Adding a Data Visualization engine to your CDSW project.....	5
Creating a Data Visualization application in CDSW.....	6
Starting Data Visualization in CDSW.....	10
 Accessing Data Visualization with ML Runtime.....	 12
Setting ML runtime as default engine in CDSW.....	12
Setting ML Runtime as engine in a CDSW project.....	12
Creating Data Visualization application in CDSW using ML Runtime.....	13
Starting Data Visualization in CDSW.....	14

Accessing Data Visualization with ML Runtime

CDP Data Visualization is integrated with Cloudera Data Science Workbench (CDSW) workflows. You can use it in an on-prem environment by running it as a project application inside of CDSW. Learn how you can add a Data Visualization engine and create your application in the CDSW enterprise data science platform.



Tip: In CDP Public Cloud, you can also work with Data Visualization in:

- [Cloudera Machine Learning service](#)
- [Cloudera Data Warehouse service](#)

Follow these steps to set up your Data Visualization application in CDSW:

Adding a custom Data Visualization engine to CDSW

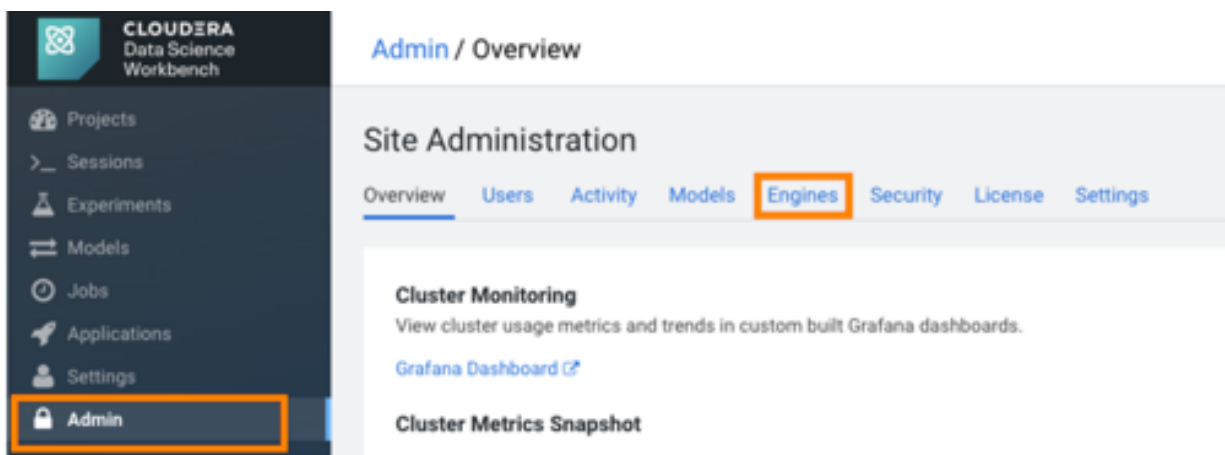
Learn how to add a Data Visualization engine to Cloudera Data Science Workbench.

Procedure

1. Go to the Cloudera Data Science Workbench web application and log in as a site administrator.
2. On the left sidebar, click Admin.

You will see an array of tabs for all the tasks you can perform as a site administrator.

3. Click the Engines tab.



- Provide a name for the Data Visualization engine that appears in your Engine list along with the full link to the docker image inside the Cloudera Docker Repository.

Use the following engine:

`docker.repository.cloudera.com/cloudera/cdv/cdswdataviz:6.4.2-b13`



Note: This is the docker image path of the latest released version for the 6.x line of Data Visualization. You can find path information for earlier versions in the Release Notes.

The screenshot shows the Cloudera Admin / Engines page. The left sidebar contains navigation links: Projects, Sessions, Experiments, Models, Jobs, Applications, Settings, and Admin. The main content area is titled 'Admin / Engines' and includes a search bar and a user profile 'admin'. Below the title, there are two sections: 'Engines Profiles' and 'Engine Images'.

Engines Profiles

vCPU is expressed in fractional virtual cores and allows bursting. Memory is expressed in fractional GiB and is enforced by memory killer. GPU indicates the number of GPUs that need to be used by the engine. Configurations larger than the maximum allocatable CPU, memory and GPU per node will be unschedulable.

Description	vCPU (burstable)	Memory (GiB)	Actions
1 vCPU / 2 GiB Memory	1	2	Edit Delete
2 vCPU / 4 GiB Memory	2	4	Edit Delete
1 vCPU (burstable), 1.75 GiB memory	1	1.75	Add

Engine Images

Whitelist Docker images for project owners to use in their jobs and sessions. These must be public images in registries that are accessible from the Cloudera Machine Learning hosts.

Description	Repository Tag	Editors	Default	Actions
Base Image v10	docker.repository.cloudera.com/cdsw/engine:10	Jupyter Notebook	☒	Edit Deprecate
cdswengine_viz 6.0.0.0_5a44903ad	docker.repository.cloudera.com/viz/cdswengine_viz 6.0.0.0_5a44903ad		☐	Edit Deprecate
cdswengine_viz 6.0.0.0_5a44903ad	docker.repository.cloudera.com/viz/cdswengine_viz 6.0.0.0_5a44903ad			Add

- Click Add.

What to do next

Add a Data Visualization engine to your CDSW project

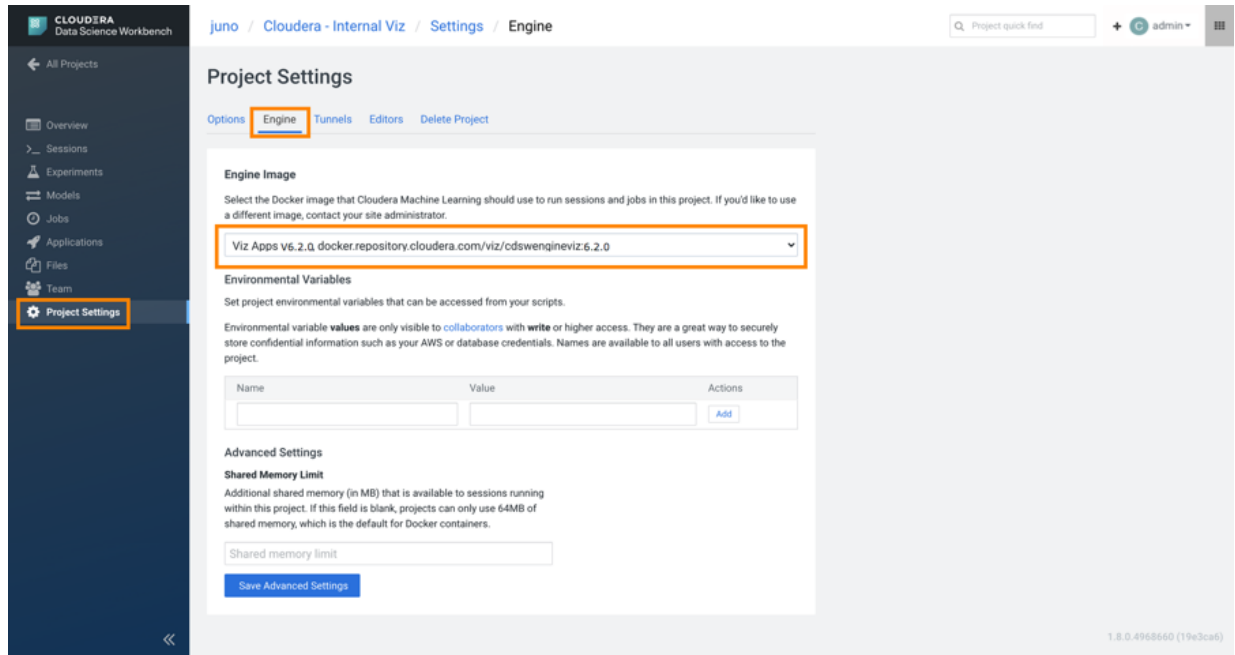
Adding a Data Visualization engine to your CDSW project

Learn how to add the custom CDP Data Visualization engine to a project in Cloudera Data Science Workbench.

Procedure

- On the left sidebar, click Projects.
- Select the project where you want to add your custom engine.
- Open Project Settings from the left sidebar and click the Runtime/Engine tab.

4. Choose the Data Visualization image from the drop-down menu.



5. Click Save Engine.

What to do next

Create a Data Visualization application in CDSW

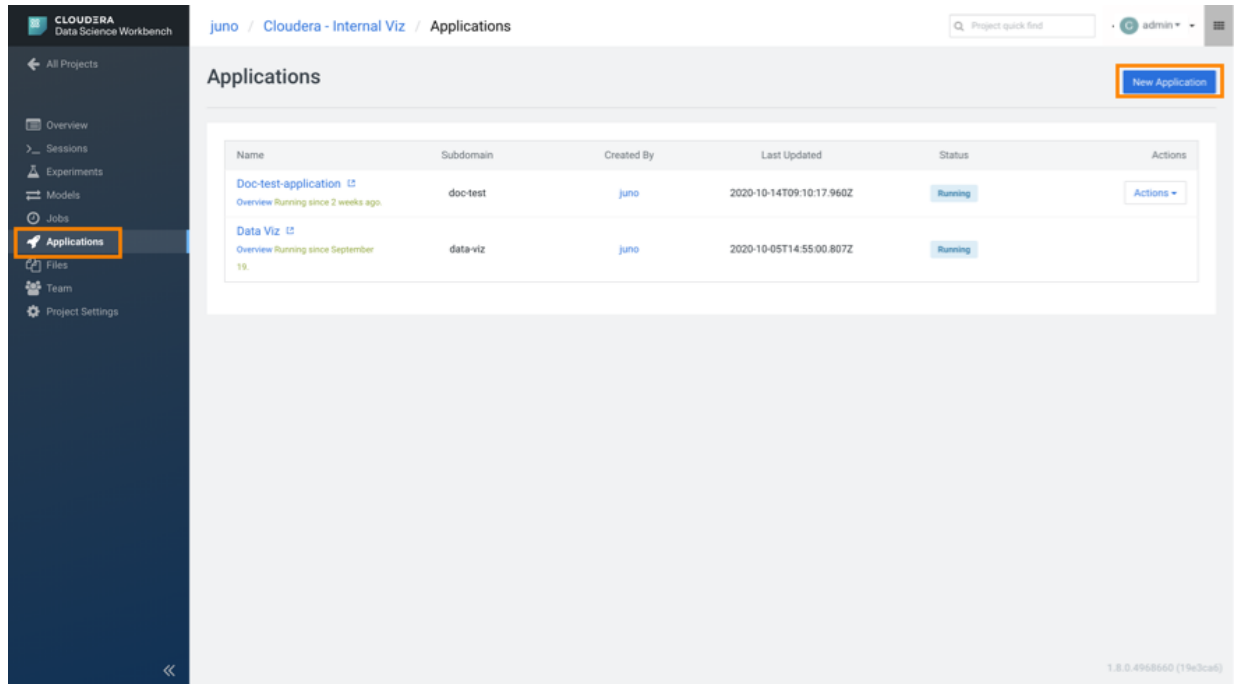
Creating a Data Visualization application in CDSW

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

Procedure

1. Go to a project's Overview page.
2. On the left sidebar, click Applications.

3. Click New Application.



The screenshot shows the Cloudera Data Science Workbench interface. The left sidebar contains a navigation menu with items: All Projects, Overview, Sessions, Experiments, Models, Jobs, Applications (highlighted with an orange box), Files, Team, and Project Settings. The main content area is titled 'Applications' and features a 'New Application' button in the top right corner, also highlighted with an orange box. Below the button is a table listing applications.

Name	Subdomain	Created By	Last Updated	Status	Actions
Doc-test-application Overview Running since 2 weeks ago.	doc-test	juno	2020-10-14T09:10:17.960Z	Running	Actions
Data Viz Overview Running since September 19.	data-viz	juno	2020-10-05T14:55:00.807Z	Running	

1.8.0-4968660 (19e3ca6)

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`.

- Engine Kernel – Select the kernel needed for this application.



Note: Use Python 3 to run the script as the application.

- Engine Profile – Select the computing resources needed for this application.



Tip: Choose an engine profile with 2 vCPU and 4 GiB RAM. A smaller engine profile also works but might have worse performance.

Create an Application

General

Name *

Subdomain *

Description

Script *

i Choose a script that hosts a web application on either CDSW_READONLY_PORT or CDSW_APP_PORT. Both ports acts the same as CDSW_READONLY_PORT in a Session because in an Application container, CDSW_APP_PORT is the same as CDSW_READONLY_PORT. This means that a user with at least read access to the project can view the web application. However, please note that it is upto you to make the application really read-only in terms of entities belonging to the project. In other words, CDSW does not prevent you from running an application that allows a viewer to modify files that belongs to the project.

Engine Kernel

☐ Python 2

☒ Python 3

☐ R

☐ Scala

5. Click Create Application.

Results

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

You can Stop, Restart, or Delete a CDSW application from the Actions drop-down menu on the Applications page.

If you want to make changes to an existing application, click Overview under the application name. Then go to the Settings tab to make any changes and update the application.

What to do next[Start Data Visualization in CDSW](#)

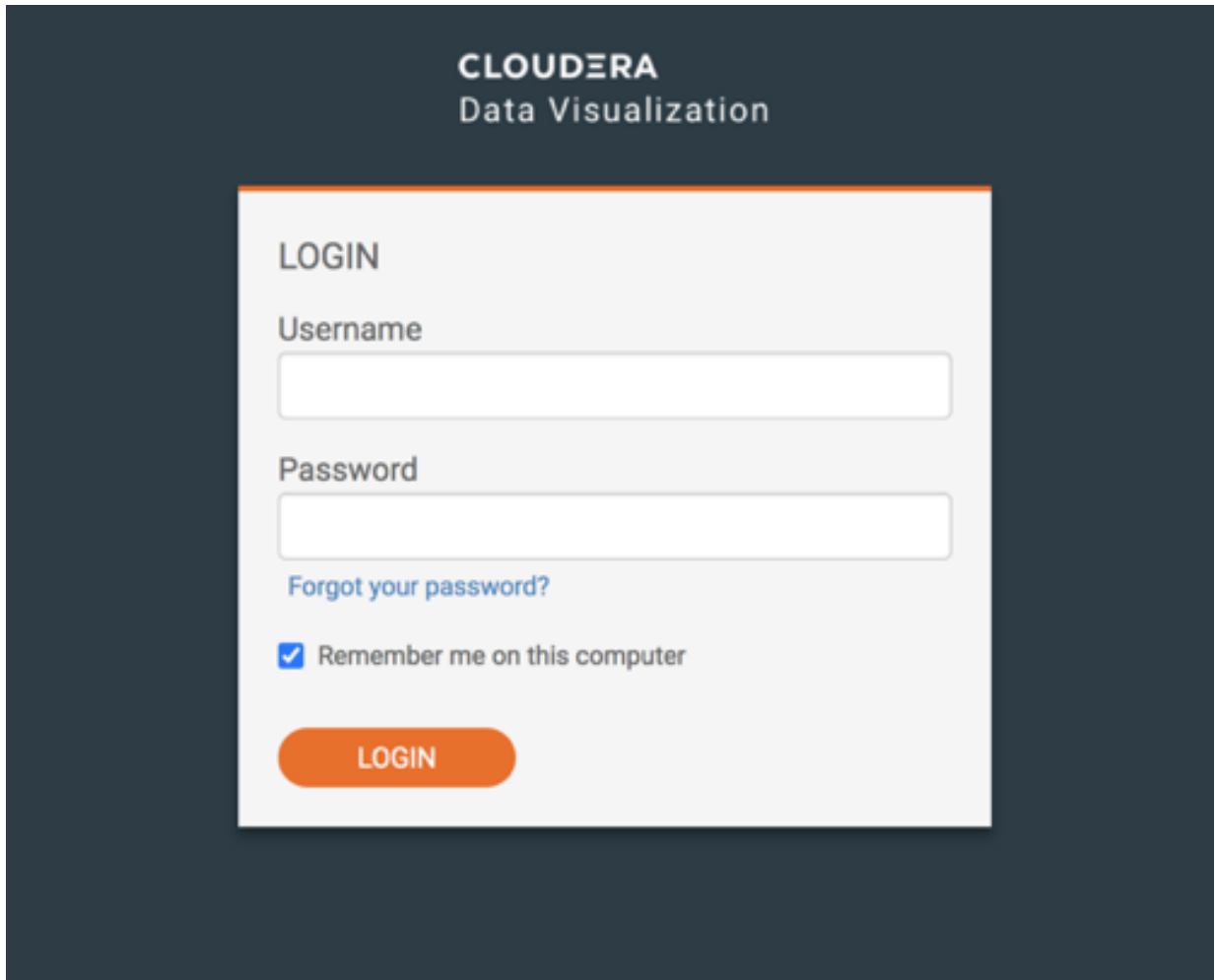
Starting Data Visualization in CDSW

Learn how to start your Data Visualization application in Cloudera Data Science Workbench (CDSW).

Procedure

1. On the Applications page, click the name of your Data Visualization application to access the web interface.
The Data Visualization application opens in a new browser tab.

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

The image shows the Cloudera Data Visualization login interface. At the top, the text "CLUDERA Data Visualization" is displayed in white on a dark blue background. Below this, a white login form is centered. The form has a title "LOGIN" in bold. It contains two input fields: "Username" and "Password". Below the password field is a link "Forgot your password?". There is a checkbox labeled "Remember me on this computer" which is checked. At the bottom of the form is an orange "LOGIN" button.

The default credentials are:

- 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 [Setting authentication parameters](#) 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.

Related Information

[Setting authentication parameters](#)

[Data Visualization in CDP](#)

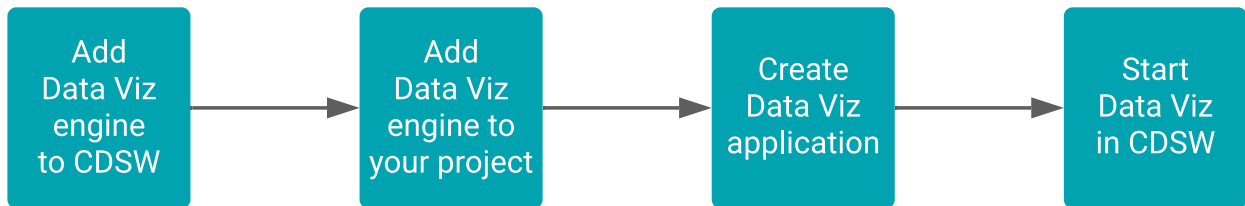
[User interface overview](#)

[Data Visualization quickstart](#)

[Connecting to a data source](#)

Accessing Data Visualization with ML Runtime

Cloudera Data Visualization (CDV) is integrated with Cloudera Data Science Workbench (CDSW) workflows. You can use it in an on-prem environment by running it as a project application inside of CDSW. One option for accessing Data Visualization is to set ML Runtime as the default engine.



For more information about ML Runtimes, see [Managing ML Runtimes](#).



Tip: In CDP Public Cloud, you can also work with Data Visualization in:

- [Cloudera Machine Learning service](#)
- [Cloudera Data Warehouse service](#)

Follow these steps to set up your Data Visualization application in CDSW:

Setting ML runtime as default engine in CDSW

Learn how to set Machine Learning (ML) Runtime as the default engine in Cloudera Data Science Workbench (CDSW).

Procedure

1. Go to the Cloudera Data Science Workbench web application and log in as a site administrator.
2. On the left sidebar, click Admin.

You will see an array of tabs for all the tasks you can perform as a site administrator.

3. Click the Runtime/Engine tab.
4. Select ML Runtime as the Default Engine.

Results

The default engine type of the workspace is updated. Now you can create a Data Visualization application. If you want to use ML Runtime only in selected projects, you can set it as a project-level engine in the Project Settings.

What to do next

[Add a Data Visualization engine to your CDSW project](#)

Setting ML Runtime as engine in a CDSW project

Learn how to define ML Runtime as the default engine in your Cloudera Data Science Workbench (CDSW) project.

About this task



Note: You only need to set ML Runtime as your project engine if the CDSW workspace of your project is not set to use ML Runtime by default.

Procedure

1. On the left navigation bar, click Projects.
2. Click the project where you want to set ML Runtime as the default engine type.
3. Open Project Settings from the left navigation bar and click the Runtime/Engine tab.
4. Select ML Runtime.

What to do next

Create a Data Visualization application in CDSW using ML Runtime

Creating Data Visualization application in CDSW using ML Runtime

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

Procedure

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

The screenshot shows the Cloudera Data Science Workbench (CDSW) interface. On the left sidebar, the 'Applications' menu item is highlighted. The main content area is titled 'Applications' and shows a table of existing applications. A 'New Application' button is located in the top right corner of the main content area.

Name	Subdomain	Created By	Last Updated	Status	Actions
Doc-test-application Overview	doc-test	juno	2020-10-14T09:10:17.960Z	Running	Actions
Data Viz Overview	data-viz	juno	2020-10-05T14:55:00.807Z	Running	

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`.

- 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.

5. Click Create Application.

Results

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

You can Stop, Restart, or Delete a CDSW application from the Actions drop-down menu on the Applications page.

If you want to make changes to an existing application, open Application Details and go to the Settings tab to make any changes and update the application.

What to do next

[Start Data Visualization in CDSW](#)

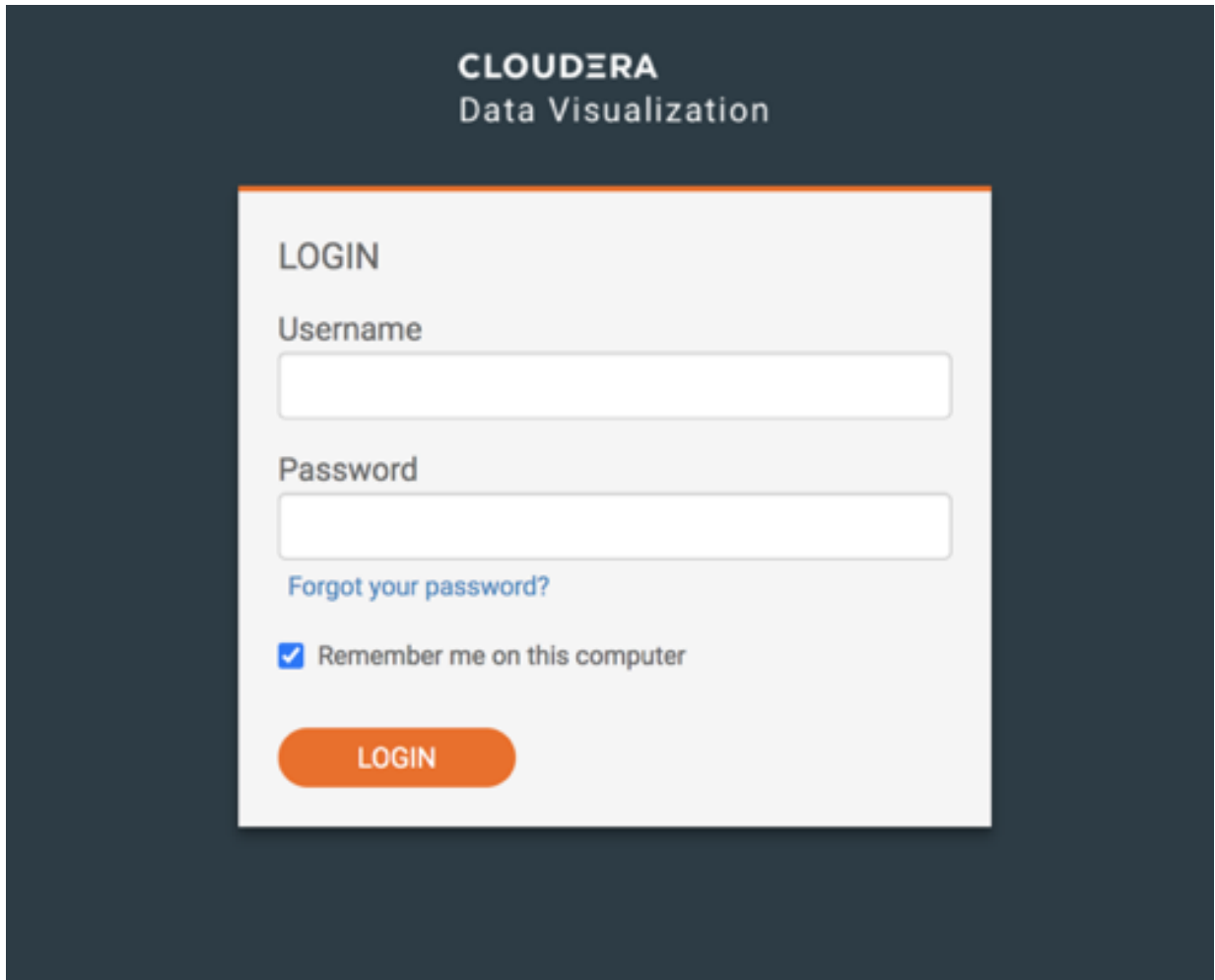
Starting Data Visualization in CDSW

Learn how to start your Data Visualization application in Cloudera Data Science Workbench (CDSW).

Procedure

1. On the Applications page, click the name of your Data Visualization application to access the web interface.
The Data Visualization application opens in a new browser tab.

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

The image shows the Cloudera Data Visualization login interface. At the top, the text "CLUDERA Data Visualization" is displayed in white on a dark blue background. Below this, a white login form is centered. The form has a title "LOGIN" in bold. It contains two input fields: "Username" and "Password". Below the password field is a link "Forgot your password?". There is a checkbox labeled "Remember me on this computer" which is checked. At the bottom of the form is an orange "LOGIN" button.

The default credentials are:

- 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 [Setting authentication parameters](#) 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.

Related Information

[Setting authentication parameters](#)

[Data Visualization in CDP](#)

[User interface overview](#)

[Data Visualization quickstart](#)

[Connecting to a data source](#)