Data Discovery and Exploration (Preview)

Date published: 2022-04-21
Date modified: 2022-04-21
Legal Notice

© Cloudera Inc. 2022. 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.

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.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.
This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.
Exploring data in CML

The Data Discovery and Exploration experience enables all features for Exploratory Data Science. From the Data UI, 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
- Quickly create visualizations of the data to understand its properties
- Create dashboards that you can share with your team

Prerequisites

- There must be auto-discovered data connections, or your Administrator must create data connections to virtual warehouses or other data sources in your CDP environment.
- As a data scientist, you must create and set a Workload password. See https://docs.cloudera.com/machine-learning/cloud/mlde/topics/ml-mlde-using-data-connections.html for more information.

Starting Data Discovery and Visualization

1. Create or open a project.
2. In the project, click Data in the left navigation pane.
3. If the Data Visualization app was previously started, it appears in the UI.
4. If not, click Launch Data Application to start the app. The app takes a few minutes to start the first time.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.
The app opens on the Home tab. When the Application starts, you can see four tabs at the top of the UI. From here, you can follow the steps in Get Started, or try some of the following tasks.
Here are some things you can try:

1. Run a SQL query

On the SQL tab, select a Data Connection from the dropdown box.

Note: If you do not see the data connections that you have configured in the Project, refer to the Troubleshooting section for more information.

Once you selected a Data Connection you can explore the available databases and tables. You can start writing a SQL query and click the table names to insert them into the query string. For more information, see Creating datasets (https://docs.cloudera.com/data-visualization/cloud/work-with-data/topics/viz-create-dataset.html)
2. Build a dashboard

Click **New Dashboard** to access the Dashboard Designer. You can select from a wide variety of graphs, click to add them to your dashboard, and select the columns from your dataset to use in the visual. For more information, see: Create a Dashboard (https://docs.cloudera.com/data-visualization/cloud/quickstart/topics/viz-qs-create-dashboard.html)
Environmental Variables

Environmental Variables can be set in the Settings UI. Select Applications, and in the options menu of the Data Discovery and Visualization app, select Application Details.

Next, select the Settings tab, and at the bottom of the UI, select Set Environment Variables. You can enter a name and value pair, then click Add to add the environment variable.

NOTE: Do not delete the CDV_SUB_MODE and MLExplore name:value pair. This variable is necessary for correct operation of the application.
Troubleshooting

You launch the application and see the spinner, but then the empty data page returns.

On the applications tab, click Data Discovery and Visualization application, and check the logs. Check if another user stopped or restarted the application. On the applications tab, click on Settings tab > Resource Profile, and check if the available resources are sufficient. You can increase the resources if necessary, then restart the application.

Data Application doesn’t show all data connections

See the error message to fix the workload password. To set the workload password:

1. View your profile, and set a workload password using the "Set Workload Password" link
2. In your ML Workspace, go to "User Settings" page
3. Under "Environmental Variables" tab, set your "WORKLOAD_PASSWORD" variable to the same password from your Cloudera Management Console profile

Alternatively, you may need to click Sync under the Data Connections to get all of the data connections available at the project level.

HTTP code 401: Unauthorized when accessing Hive

Check that the Hive data connection is working.

1. In the control plane, go to Data Warehouses.
2. Click Virtual Warehouses, and select the data warehouse for the connection.
3. Check the status of that data warehouse, and make sure that it is running or in a good state.