

Cloudera Data Science Workbench

Managing Cloudera Data Science Workbench

Date published: 2020-02-28

Date modified:

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with a stylized 'E' that has three horizontal bars.

<https://docs.cloudera.com/>

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

Cloudera Data Science Workbench Email Notifications.....	4
Managing License Keys for Cloudera Data Science Workbench.....	4
Trial License.....	4
Cloudera Enterprise License.....	4
Uploading License Keys.....	4
Creating Resource Profiles.....	5
User Access to Features.....	6
Onboarding Business Users.....	7
Adding CDSW Session Metadata Information.....	7
Web session timeouts.....	9

Cloudera Data Science Workbench Email Notifications

By default, all emails are sent from `noreply@your-cdsw-domain`. However, if your SMTP domain is different from the Cloudera Data Science Workbench domain, or it does not allow spoofing, you will need to explicitly specify the email address in the No Reply Email field, otherwise email sending may fail.

Required Role: Site Administrator

To specify an email address for outbound invitations and job notifications, go to the `Admin Settings` tab.

Cloudera Data Science Workbench sends email notifications when you add collaborators to a project, share a project with a colleague, and for job status updates (email recipients are configured per job). Emails are not sent when you create a new project. Email preferences cannot currently be configured at an individual user level.

Managing License Keys for Cloudera Data Science Workbench

Cloudera Data Science Workbench requires a Cloudera Enterprise license. To obtain a Cloudera Enterprise license, call 866-843-7207. Note that only one license key can be used at a time.

After an initial trial period of 60 days, you must upload a license key to continue to use Cloudera Data Science Workbench.

Trial License

Cloudera Data Science Workbench is fully functional during a 60-day, non-renewable trial period. The trial period starts when you create your first user.

If 60 days or fewer remain on the license, a badge in the lower left corner of the dashboard displays the number of days remaining. The initial trial period is 60 days, so the remaining days are always displayed during the trial period.

When a trial license expires, functionality is limited as follows:

- A warning banner notifies you that the license has expired and suggests you contact the site administrator or upload a license key.
- You cannot create new users, projects, sessions, or jobs.
- Existing users can log in and view their projects and files.
- You cannot run existing jobs.

At this point you can obtain a Cloudera Enterprise license and upload it to Cloudera Data Science Workbench using the steps described below. Cloudera Data Science Workbench will then go back to being fully functional.

Cloudera Enterprise License

When an Enterprise license expires, a warning banner displays, but all product features remain fully functional.

Contact [Cloudera Support](#) to receive an updated license.

Uploading License Keys

You can upload license keys.

Procedure

1. Go to Admin License .
2. Click Upload License.
3. Select the license file to be uploaded and click Upload.

Creating Resource Profiles

Resource profiles define how many vCPUs and how much memory the product will reserve for a particular workload (for example, session, job, model).

About this task

As a site administrator you can create several different vCPU, GPU, and memory configurations which will be available when launching a session/job. When launching a new session, users will be able to select one of the available resource profiles depending on their project's requirements.

Procedure

1. To create resource profiles, go to the Site Administration Runtime/Engine page.
2. Add a new profile under Resource Profiles.

Cloudera recommends that all profiles include at least 2 GB of RAM to avoid out of memory errors for common user operations.

You will see the option to add GPUs to the resource profiles only if your Cloudera Data Science Workbench hosts are equipped with GPUs, and you have enabled them for use by setting the relevant properties in `cdsw.conf`.

Results

If there are two worker nodes and 10 vCPU available overall, if one user tries to establish a session with 8 vCPU, CDSW will not allow it. The memory and CPU must be contiguous (adjacent to each other). When a user spins a session, the pod triggers on a single node and resources on the same node are utilized. This is expected behavior for Kubernetes.

Figure 1: Resource profiles available when launching a session

The screenshot displays the 'Site Administration / Runtime/Engine' page. On the left, a table lists 'Resource Profiles' with columns for Description, vCPU, and Memory (GiB). The first row, '1 vCPU / 2 GiB Memory', is highlighted with a blue box. A blue arrow points from this row to the 'Start A New Session' dialog box on the right. In the dialog, the 'Resource Profile' dropdown menu is open, showing the same '1 vCPU / 2 GiB Memory' profile selected.

Description	vCPU	Memory (GiB)
1 vCPU / 2 GiB Memory	1	2
1 vCPU / 1.75 GiB Memory	1	1.75
1 vCPU, 1.75 GiB memory	1	1.75

Start A New Session

Session Name:

Runtime: Editor: Workbench, Kernel: Python 3.7

Edition: Standard, Version: 2021.09

Enable Spark: Spark 2.4 - CDP 7.2.8 (TP)

Runtime Image: - docker.repository.cloudera.com/cdsw/ml-runtime-workbench-python3.7-standard.2021.09.1-b5

Resource Profile: 1 vCPU / 2 GiB Memory (selected)

Buttons: Cancel, Start Session

User Access to Features

Cloudera Data Science Workbench provides Site Administrators with the ability to restrict or control specific functionality that non-Site Administrator users have access to. This is done through the Security and Settings tabs on the Admin page.

Security Tab

Table 1: Security Tab

Property	Description
Allow remote editing	Disable this property to prevent users from connecting to the Cloudera Data Science Workbench deployment with <code>cdswctl</code> and using local IDEs, such as PyCharm.
Allow only session creators to run commands on active sessions	By default, a user's permission to active sessions in a project is the same as the user's permission to that project, which is determined by the combination of the user's permission as a project collaborator, the user's permission in the team if this is a team project, and whether the user is a Site Administrator. By checking this checkbox, only the user that created the active session will be able to run commands in that session. No other users, regardless of their permissions in the team or as project collaborators, will be able to run commands on active sessions that are not created by them. Even Site Administrators will not be able to run commands in other users' active sessions.
Allow console output sharing	Disable this property to remove the Share button from the project workspace and workbench UI as well as disable access to all shared console outputs across the deployment. Note that re-enabling this property does not automatically grant access to previously shared consoles. You will need to manually share each console again.
Allow anonymous access to shared console outputs	Disable this property to require users to be logged in to access shared console outputs.
Allow file upload/download through UI	Use this checkbox to show/hide file upload/download UI in the project workspace. When disabled, Cloudera Data Science Workbench API will forbid request of downloading file(s) as attachment. Note that the backend API to upload/edit/read the project files are intact regardless of this change in order to support basic Cloudera Data Science Workbench functionality such as file edit/read.

Settings Tab

Table 2: Settings Tab

Property	Description
Allow users to create public projects	Disable this property to restrict users from creating new public projects. Site Administrators will have to create any new public projects.
Allow users to create projects	Disable this property to restrict users from creating new projects. Site Administrators will have to create any new projects.
Allow users to create teams	Disable this property to restrict users from creating new teams. Site Administrators will have to create any new teams.
Allow users to run experiments	Disable this property to hide the Experiments feature in the UI. Note that this property does not affect any active experiments.
Allow users to create models	Disable this property to hide the Models feature in the UI. Note that this property does not affect any active models. In particular, if you do not stop active models before hiding the Models feature, they continue to serve requests and consume computing resources in the background.

Onboarding Business Users

There are two procedures required for adding Business Users to CML. First, an Admin ensures the Business User has the correct permissions, and second, a Project Owner adds the Business User as a Collaborator.

Before you begin

Make sure the user is already assigned in your external identity provider, such as LDAP.

About this task

The Admin user performs these steps:

Procedure

1. In Environments, select the correct environment where the ML workspace is hosted.
2. In Manage Access, search for the user, and add the ML Business User role. Make sure the user does not already have a higher-level permission, such as ML Admin or ML User, either through a direct role assignment or a group membership.
3. Click Update Roles.
4. Inside the ML Workspace, go to Site Administration > Users , and click Synchronize Users. This adds the necessary Users defined at the Environment level to the Workspace, and updates any role changes that have been made.

What to do next

Add the ML Business User as a Collaborator to a Project.

Adding CDSW Session Metadata Information

Administrators can make it mandatory for the users to enter additional metadata before starting a CDSW session. Administrators can configure up to three fields to collect such additional metadata information for compliance purposes and can make these mandatory to be filled in. This metadata is captured and stored in the user_events table and in the audit log as key-value pairs.

Procedure

1. Sign in to the CDSW web UI as the admin user.
2. Go to Admin and click Settings.

3. Turn on the Session Metadata clicking the toggle button.

The Session Metadata table is enabled for editing, as shown in the following image:

Session Metadata

ON

Label	Actions
Organization	<input type="button" value="Edit"/> <input type="button" value="Remove"/>
<input type="text" value="Session metadata label"/>	<input type="button" value="Add"/>

4. Add the session metadata fields in the table.
You can add up to three fields.

5. Test the feature by starting a new session with any non-admin user.

The previously added fields should appear on the **Start A New Session** form, as shown in the following image:

Start A New Session

⚠ Not authenticated to Hadoop
Before you can connect to your secure Hadoop cluster, you must enter your credentials under [Settings > Hadoop Authentication](#)

* **Session Name**
Untitled Session

Engine

Editor ⓘ Workbench

Kernel ⓘ Python 3

Engine Image - Configure
Base Image v12 - docker.repository.cloudera.com/cdsw/engine:12

Resource Profile
1 vCPU / 2 GiB Memory

Session Metadata ⓘ

* **Organization**
Marketing

Cancel Start Session

Web session timeouts

You can set web sessions to time out and require the user to log in again. This time limit is not based on activity, it is the maximum time allowed for a web session.

You can set timeout limits for Users and Admin Users in [Site Administration Security](#).

- **User Web Browser Timeout (minutes)** - This timeout sets the default maximum length of time that a web browser session for an Admin user can remain inactive. You remain logged in if you are actively using the session. If you are not active, then after a 5-minute warning, you are automatically logged out. Any changes to the setting take effect for any subsequent Admin user logins.
- **Admin User Web Browser Timeout (minutes)** - This timeout sets the default maximum length of a web browser session for an Admin user. This timeout is not based on inactivity. When this time limit is reached, the Admin user is asked to refresh the session. Any changes to the setting take effect for any subsequent Admin user logins.