

Collaborating on Projects with Cloudera Machine Learning

Date published: 2020-07-16

Date modified: 2022-07-21



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

Collaborating on Projects with Cloudera Machine Learning.....	4
Project Collaborators.....	4
Teams.....	4
ML Business User.....	4
Forking Projects.....	5
Collaborating with Git.....	5
 Sharing Job and Session Console Outputs.....	 5

Collaborating on Projects with Cloudera Machine Learning

This topic discusses all the collaboration strategies available to Cloudera Machine Learning users.

Project Collaborators

If you want to work closely with trusted colleagues on a particular project, you can add them to the project as collaborators. This is recommended for collaboration over projects created under your personal account. Anyone who belongs to your organization can be added as a project collaborator.

Project Visibility Levels: When you create a project in your personal context, Cloudera Machine Learning asks you to assign one of the following visibility levels to the project - Private or Public. Public projects on Cloudera Machine Learning grant read-level access to everyone with access to the Cloudera Machine Learning application. For Private projects, you must explicitly add someone as a project collaborator to grant them access.

Project Collaborator Access Levels: You can grant project collaborators the following levels of access: Viewer, Operator, Contributor, Admin



Note:

Collaborating Securely on Projects

Before adding project collaborators, you must remember that assigning the Contributor or Admin role to a project collaborator is the same as giving them write access to your data in CDH. This is because project contributors and project administrators have write access to all your project code (including any library code that you might not be actively inspecting). For example, a contributor/admin could modify project file(s) to insert code that deletes some data on the cluster. The next time you launch a session and run the same code, it will appear as though you deleted the data yourself.

Additionally, project collaborators also have access to all actively running sessions and jobs by default. This means that a malicious user can easily impersonate you by accessing one of your active sessions. Therefore, it is extremely important to restrict project access to trusted collaborators only. Note that site administrators can restrict this ability by allowing only session creators to run commands within their own active sessions.

For these reasons, Cloudera recommends using Git to collaborate securely on shared projects.

Teams

Users who work together on more than one project and want to facilitate collaboration can create a Team. Teams allow streamlined administration of projects. Team projects are owned by the team, rather than an individual user. Only users that are already part of the team can be added as collaborators to projects created within the team context. Team administrators can add or remove members at any time, assigning each member different access permissions.

Team Member Access Levels: You can grant team members the following levels of access: Viewer, Operator, Contributor, Admin.

ML Business User

The ML Business User role is for a user who only needs to view any applications that are created within Cloudera Machine Learning. This is the ideal role for an employee who is not part of the Data Science team and does not need higher-level access to workspaces and projects, but needs to access the output of a Data Science workflow. MLBusinessUser seats are available for purchase separately.

Forking Projects

You can fork another user's project by clicking Fork on the Project page. Forking creates a new project under your account that contains all the files, libraries, configurations, jobs, and dependencies between jobs from the original project.

Creating sample projects that other users can fork helps to bootstrap new projects and encourage common conventions.

Collaborating with Git

Cloudera Machine Learning provides seamless access to Git projects. Whether you are working independently, or as part of a team, you can leverage all of benefits of version control and collaboration with Git from within Cloudera Machine Learning. Teams that already use Git for collaboration can continue to do so. Each team member will need to create a separate Cloudera Machine Learning project from the central Git repository.

For anything but simple projects, Cloudera recommends using Git for version control. You should work on Cloudera Machine Learning the same way you would work locally, and for most data scientists and developers that means using Git.

Sharing Job and Session Console Outputs

This topic describes how to share the results of your research (that is, output from sessions and jobs) with teammates and project stakeholders.

Cloudera Machine Learning lets you easily share the results of your analysis with one click. Using rich visualizations and documentation comments, you can arrange your console log so that it is a readable record of your analysis and results. This log continues to be available even after the session stops. This method of sharing allows you to show colleagues and collaborators your progress without your having to spend time creating a report.

To share results from an interactive session, click Share at the top of the console page. From here you can generate a link that includes a secret token that gives access to that particular console output. For jobs results, you can either share a link to the latest job result or a particular job run. To share the latest job result, click the Latest Run link for a job on the Overview page. This link will always have the latest job results. To share a particular run, click on a job run in the job's History page and share the corresponding link.

You can share console outputs with one of the following sets of users.

- All anonymous users with the link - By default, Cloudera Machine Learning allows anonymous access to shared consoles. However, site administrators can disable anonymous sharing at any time.

Once anonymous sharing has been disabled, all existing publicly shared console outputs will be updated to be viewable only by authenticated users.

- All authenticated users with the link - This means any user with a Cloudera Machine Learning account will have access to the shared console.
- Specific users and teams - Click Change to search for users and teams to give access to the shared console. You can also come back to the session and revoke access from a user or team the same way.

Sharing Data Visualizations

If you want to share a single data visualization rather than an entire console, you can embed it in another web page. Click the small circular 'link' button located to the left of most rich visualizations to view the HTML snippet that you can use to embed the visualization.