Machine Learning 1.4.1

Site Administration

Date published: 2020-07-16
Date modified: 2022-07-21

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

Managing Users....4

Configuring Quotas....5

Creating Resource Profiles....5

Onboarding Business Users....6

Adding a Collaborator....7

Monitoring Cloudera Machine Learning Activity....7

Tracked User Events....8

Monitoring User Events....11

Monitoring Active Models Across the Workspace....13

Monitoring and Alerts....14

Application Polling Endpoint....14

Choosing Default Engine....14

Controlling User Access to Features....15

Cloudera Machine Learning Email Notifications....17

Web session timeouts....17

Project Garbage Collection....17

How to make base cluster configuration changes....17

Disable Addons....18

Managing Users

This topic describes how to manage an ML workspace as a site administrator. Site administrators can monitor and manage all user activity across a workspace, add new custom engines, and configure certain security settings.

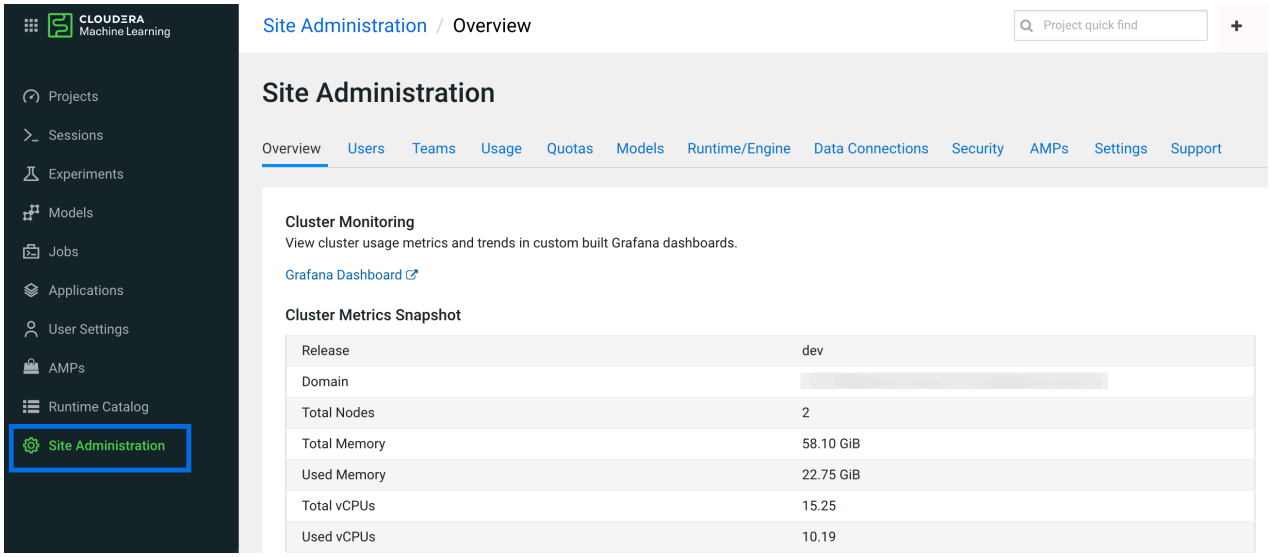
By default, the first user that logs in to a workspace must always be a site administrator. That is, they should have been granted the MLAdmin role by a CDP PowerUser.

Note: On Private Cloud, the corresponding role is EnvironmentAdmin.

Important: Site administrators have complete access to all activity on the deployment. This includes access to all teams and projects on the deployment, even if they have not been explicitly added as team members or collaborators.

Only site administrators have access to a Site Administration dashboard that can be used to manage the workspace. To access the site administrator dashboard:

1. Go to the Cloudera Machine Learning web application and log in as a site administrator.
2. On the left sidebar, click Site Administration. You will see an array of tabs for all the tasks you can perform as a site administrator.



Monitoring Users

The Users tab on the admin dashboard displays the complete list of users. You can see which users are currently active, and when a user last logged in to Cloudera Machine Learning. To modify a user's username, email or permissions, click the Edit button under the Action column.

Synchronizing Users

You can synchronize Users within an ML Workspace with those that have been defined access at the Environment level (through the MLAdmin, MLUser, and MLBusinessUser roles). Doing so for new Users enables you to take administrative actions such as setting Team assignments, defining Project Collaborators, and more, all prior to the new Users' first time logging in to the Workspace.

To synchronize Users, go to Site Administration Users , and click Synchronize Users. This adds any users defined at the Environment level to the workspace, updates any role changes that have been made, and deactivates any users that have been deactivated.

Note: The Administrator should periodically perform user synchronization to ensure that users who are deactivated on the environment level are also deactivated in CML.

Related Information

Cloudera Machine Learning Email Notifications

Configuring Quotas

This topic describes how to configure CPU, GPU, and memory quotas for users of an ML workspace.

Before you begin

Required Role: MLAdmin

Note: On Private Cloud, the corresponding role is EnvironmentAdmin.

Make sure you are assigned the MLAdmin role in CDP. Only users with the MLAdmin role will be logged into ML workspaces with Site Administrator privileges.

There are two types of quota: Default and Custom. Default quotas apply to all users of the workspace. Custom quotas apply to individual users in the workspace, and take precedence over the default quota.

Procedure

1. Log in to the CDP web interface.
2. Click ML Workspaces, then open the Workspace for which you want to set quotas.
3. Click AdminQuotas.
4. Switch the Default Quotas toggle to ON.

   This applies a default quota of 2 vCPU and 8 GB memory to each user in the workspace.

   If your workspace was provisioned with GPUs, a default quota of 0 GPU per user applies. If you want users to have access to GPUs, you must modify the default quotas as described in the next step.
5. If you want to change the default quotas, click on Default (per user) .

   CML displays the Edit default quota dialog box.
6. Enter the CPU, Memory, and GPU quota values that should apply to all users of the workspace.
7. Click Update.
8. To add a custom quota for a specific user, click Add User.
9. Enter the user name, and enter the quotas for CPU, Memory, and GPU.
10. Click Add.

Results

Enabling and modifying quotas will only affect new workloads. If users have already scheduled workloads that exceed the new quota limits, those will continue to run uninterrupted. If a user is over their limit, they will not be able to schedule any more workloads.

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

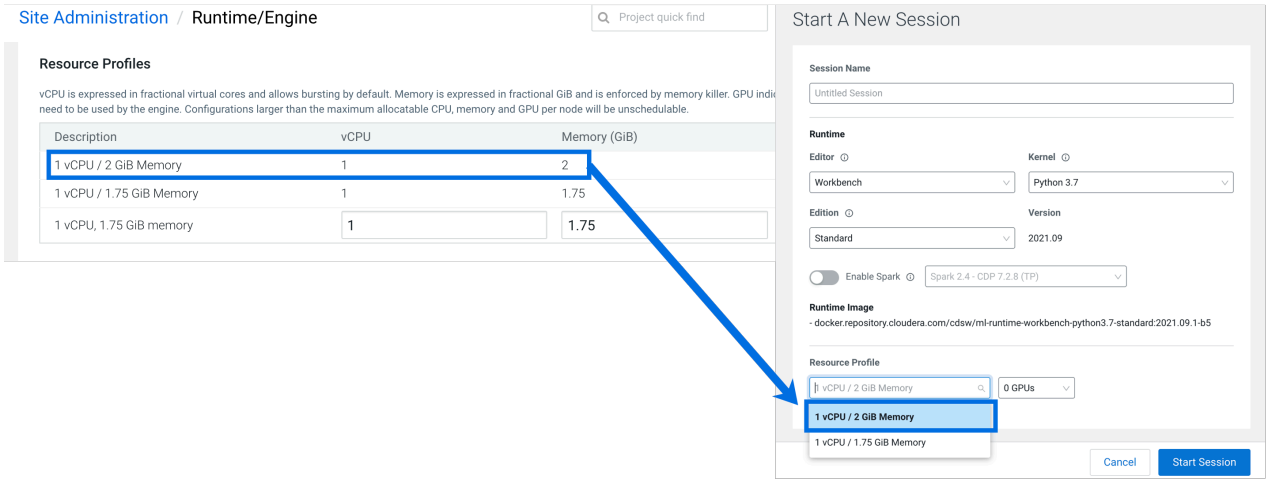
1. To create resource profiles, go to the Site Administration Runtime/Engine page.
2. Add a new profile under Resource Profiles.

   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 hosts are equipped with GPUs, and you have enabled them for use by setting the relevant properties in cdsw.conf.

Results

Figure 1: Resource profiles available when launching a session



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.

Related Information
Adding a Collaborator

Adding a Collaborator

Project Owners can add Collaborators to a project.

About this task

The Project Owner performs these steps:

Procedure

1. Go to Collaborators, and enter the user id in the Search box.
2. Choose the user id, and click Add. The user is added with their role displayed.

Results
Now, when the Business User logs in, they are able to access the Applications under this project.

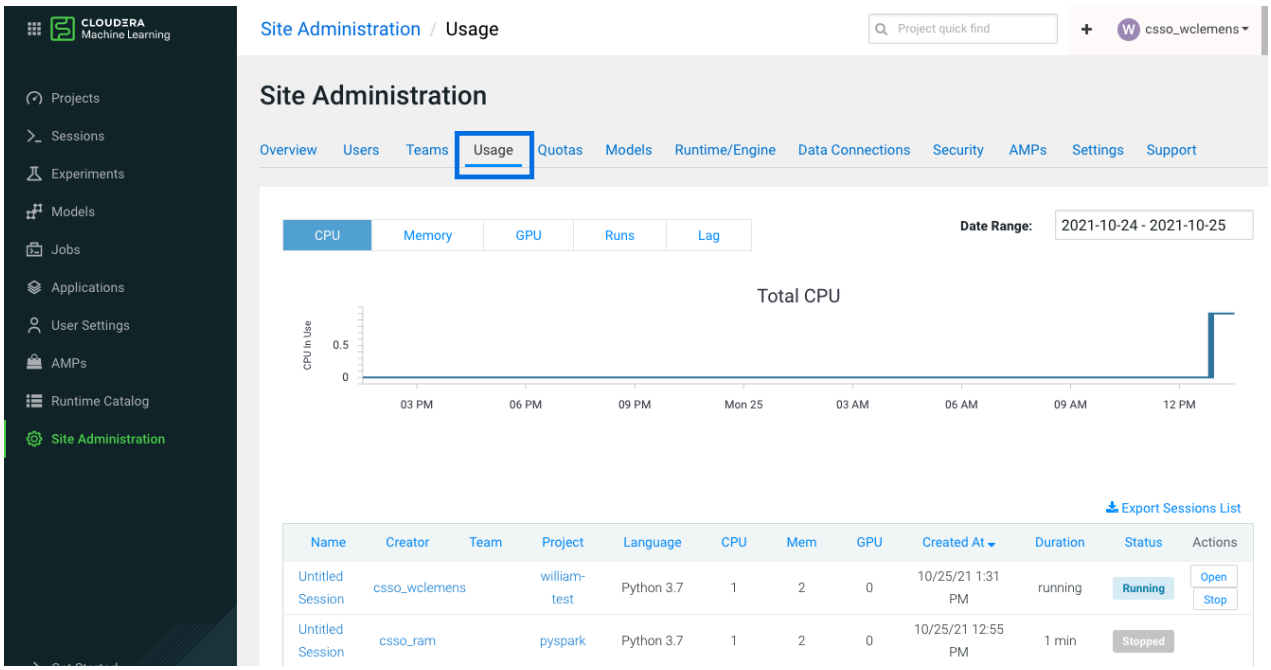
Monitoring Cloudera Machine Learning Activity

This topic describes how to monitor user activity on an ML workspace.

Required Role: Site Administrator

The Admin Overview tab displays basic information about your deployment, such as the number of users signed up, the number of teams and projects created, memory used, and some average job scheduling and run times. You can also see the version of Cloudera Machine Learning you are currently running.

The Admin Activity tab of the dashboard displays the following time series charts. These graphs should help site administrators identify basic usage patterns, understand how cluster resources are being utilized over time, and how they are being distributed among teams and users.



Important: The graphs and numbers on the Admin Activity page do not account for any resources used by active models on the deployment. For that information, go to Admin Models page.

- CPU - Total number of CPUs requested by sessions running at this time.

  Note that code running inside an n-CPU session, job, experiment or model replica can access at least n CPUs worth of CPU time. Each user pod can utilize all of its host's CPU resources except the amount requested by other user workloads or Cloudera Machine Learning application components. For example, a 1-core Python session can use more than 1 core if other cores have not been requested by other user workloads or CML application components.
- Memory - Total memory (in GiB) requested by sessions running at this time.
- GPU - Total number of GPUs requested by sessions running at this time.
- Runs - Total number of sessions and jobs running at this time.
- Lag - Depicts session scheduling and startup times.
  - Scheduling Duration: The amount of time it took for a session pod to be scheduled on the cluster.
  - Starting Duration: The amount of time it took for a session to be ready for user input. This is the amount of time since a pod was scheduled on the cluster until code could be executed.

The Export Sessions List provides a CSV export file of the columns listed in the table. It is important to note that the exported duration column is in seconds for a more detailed output.

Tracked User Events

The tables on this page describe the user events that are logged by Cloudera Machine Learning.

Table 1: Database Columns

When you query the user_events table, the following information can be returned:

Information	Description
id	The ID assigned to the event.
user_id	The UUID of the user who triggered the event.

Information	Description
ipaddr	The IP address of the user or component that triggered the event. 127.0.0.1 indicates an internal component.
user agent	The user agent for this action, such as the web browser. For example: `Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36`
event_name	The event that was logged. The tables on this page list possible events.
description	This field contains the model name and ID, the user type (NORMAL or ADMIN), and the username.
created_at	The date (YYYY-MM-DD format) and time (24-hour clock) the event occurred .

Table 2: Events Related to Engines

Event	Description
engine environment vars updated	-
engine mount created	-
engine mount deleted	-
engine mount updated	-
engine profile created	-
engine profile deleted	-
engine profile updated	-

Table 3: Events Related to Experiments

Event	Description
experiment run created	-
experiment run repeated	-
experiment run cancelled	-

Table 4: Events Related to Files

Event	Description
file downloaded	-
file updated	-
file deleted	-
file copied	-
file renamed	-
file linked	The logged event indicates when a symlink is created for a file or directory.
directory uploaded	-

Table 5: Events Related to Models

Event	Description
model created	-
model deleted	-

Table 6: Events Related to Jobs

Event	Description
job created	-
job started	-
stopped all runs for job	-
job shared with user	-
job unshared with user	-
job sharing updated	The logged event indicates when the sharing status for a job is changed from one of the following options to another: • All anonymous users with the link • All authenticated users with the link • Specific users and teams

Table 7: Events Related to Licenses

Event	Description
license created	-
license deleted	-

Table 8: Events Related to Projects

Event	Description
project created	-
project updated	-
project deleted	-
collaborator added	-
collaborator removed	-
collaborator invited	-

Table 9: Events Related to Sessions

Event	Description
session launched	-
session terminated	-
session stopped	-
session shared with user	-
session unshared with user	-

Event	Description
update session sharing status	The logged event indicates when the sharing status for a session is changed from one of the following options to another: • All anonymous users with the link • All authenticated users with the link • Specific users and teams

Table 10: Events Related to Admin Settings

site config updated	The logged event indicates when a setting on the Admin Settings page is changed.

Table 11: Events Related to Teams

Event	Description
add member to team	-
delete team member	-
update team member	-

Table 12: Events Related to Users

Event	Description
forgot password	-
password reset	-
update user	If the logged event shows that a user is banned, that means that the user account has been deactivated and does not count toward the license.
user signup	-
user login	The logged event includes the authorization method, LDAP/SAML or local.
user logout	-
ldap/saml user creation	The logged event indicates when a user is created with LDAP or SAML.

Monitoring User Events

This topic shows you how to query the PostgresSQL database that is embedded within the Cloudera Machine Learning deployment to monitor or audit user events.

About this task

Querying the PostgresSQL database that is embedded within the Cloudera Machine Learning deployment requires root access to the Cloudera Machine Learning Master host.

Procedure

1. SSH to the Cloudera Machine Learning Master host and log in as root.

   For example, the following command connects to cdsw-master-host. as root:

   ```
   ssh root@cdsw-master-host.yourcdswdomain.com
   ```

2. Get the name of the database pod:

```
kubectl get pods -l role=db
```

The command returns information similar to the following example:

```
NAME                  READY   STATUS    RESTARTS   AGE
db-86bbb69b54-d5q88   1/1     Running   0          4h46m
```

3. Enter the following command to log into the database as the sense user:

```
 kubectl exec <database pod> -ti -- psql -U sense
```

For example, the following command logs in to the database on pod db-86bbb69b54-d5q88:

```
kubectl exec db-86bbb69b54-d5q88 -ti -- psql -U sense
```

You are logged into the database as the sense user.

4. Run queries against the user_events table.

For example, run the following query to view the most recent user event:

```
select * from user_events order by created_at DESC LIMIT 1
```

The command returns information similar to the following:

```
id          | 3658
user_id     | 273
ipaddr      | ::ffff:127.0.0.1
user_agent  | node-superagent/2.3.0
event_name  | model created
description | {"model":"Simple Model 1559154287-ex5yn","modelId":"50","
userType":"NORMAL","username":"DonaldBatz"}
created_at  | 2019-05-29 18:24:47.65449
```

5. Optionally, you can export the user events to a CSV file for further analysis:
   a) Copy the user_events table to a CSV file:

      ```
      copy user_events to '/tmp/user_events.csv' DELIMITER ',' CSV HEADER;
      ```

   b) Find the container that the database runs on:

      ```
      docker ps | grep db-86bbb
      ```

      The command returns output similar to the following:

      ```
      c56d04bbd58 c230b2f564da "docker-entrypoint..." 7 days ago Up 7 days k8s
      _db_db-86bbb69b54-fcfm6_default_8b2dd23d-88b9-11e9-bc34-0245eb679f96_0
      ```

      The first entry is the container ID.
   c) Copy the user_events.csv file out of the container into a temporary directory on the Master host:

      ```
      docker cp <container ID>:/tmp/user_events.csv /tmp/user_events.csv
      ```

      For example:

      ```
      docker cp 8c56d04bbd58:/tmp/user_events.csv /tmp/user_events.csv
      ```

   d) Use SCP to copy /tmp/user_events.csv from the Cloudera Machine Learning Master host to a destination of your choice.

      For example, run the following command on your local machine to copy user_events.csv to a local directory named events:

      ```
      scp root@cdsw-master-host.yourcdswdomain.com:/tmp/user_events.csv /local
      /directory/events/
      ```

What to do next

For information about the different user events, see *Tracked User Events*.

Related Information

Tracked User Events

Monitoring Active Models Across the Workspace

This topic describes how to monitor all active models currently deployed on your workspace.

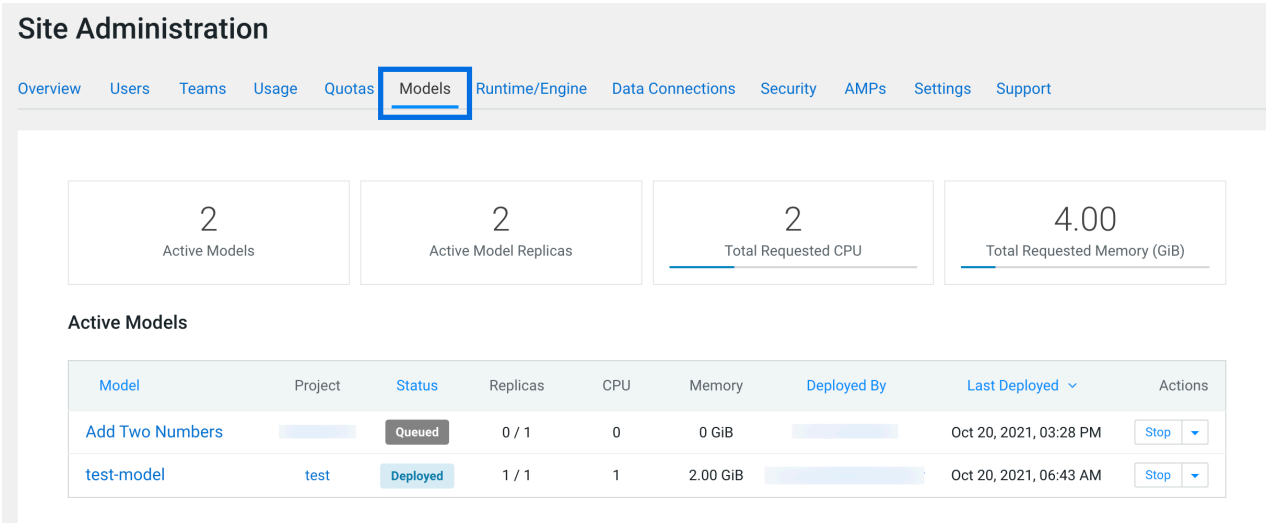
What is an Active Model?

A model that is in the Deploying, Deployed, or Stopping stages is referred to as an active model.

Monitoring All Active Models Across the Workspace

Required Role: Site Administrator

To see a complete list of all the models that have been deployed on a deployment, and review resource usage across the deployment by models alone, go to Admin Models . On this page, site administrators can also Stop/Restart/ Rebuild any of the currently deployed models.



Monitoring and Alerts

Cloudera Machine Learning leverages CDP Monitoring based on Prometheus and Grafana to provide dashboards that allow you to monitor how CPU, memory, storage, and other resources are being consumed by your ML workspaces.

Prometheus is an internal data source that is auto-populated with resource consumption data for each deployment. Grafana is the monitoring dashboard that allows you to create visualizations for resource consumption data from Prometheus. By default, CML provides three Grafana dashboards: K8 Cluster, K8s Containers, and K8s Node. You can extend these dashboards or create more panels for other metrics. For more information, see the Grafana documentation.

Related Information

Grafana documentation

Application Polling Endpoint

The CML server periodically polls applications for their status. The default polling endpoint is the root endpoint (/), but a custom polling endpoint can be specified if the server or other application has difficulty with the default endpoint.

When creating or modifying an application, you can specify a new value for the CDSW_APP_POLLING_END POINT environmental variable. Just replace the default value / that is shown. For more information, see *Analytical Applications*.

You can also set the environmental value in Project Settings Advanced . In this case, any setting made here can be overridden by settings in a given application. However, settings made in Project Settings Advanced also apply when polling sessions.

Related Information

Analytical Applications

Choosing Default Engine

This topic describes how to choose a default engine for creating projects.

Before you begin

Required Role: MLAdmin

Note: On Private Cloud, the corresponding role is EnvironmentAdmin.

Make sure you are assigned the MLAdmin role in CDP. Only users with the MLAdmin role will be logged into ML workspaces with Site Administrator privileges.

There are two types of default engines: ML Runtime and Legacy Engines. However, legacy engines are deprecated in the current release and project settings default to ML Runtime.

Legacy engines Engines contain the machinery necessary to run sessions using all four interpreter options that CML currently supports (Python 2, Python 3, R and Scala) and other support utilities (C and Fortran compilers, LaTeX, etc.). ML Runtimes are thinner and more lightweight than legacy engines. Rather than supporting multiple programming languages in a single engine, each Runtime variant supports a single interpreter version and a subset of utilities and libraries to run the user's code in Sessions, Jobs, Experiments, Models, or Applications.

Procedure

1. Log in to the CDP web interface.
2. Click ML Workspaces, then open the Workspace for which you want to set Default Engine.
3. Click Admin Runtime/Engine .
4. Choose the Default Engine you would like to use as the default for all newly created projects in this workspace.

   Note: Legacy Engines are deprecated in this release and Cloudera recommends using Runtime.

5. Modify the remaining information on the page:
   - Resource Profiles listed in the table are selectable resource options for both legacy Engines and ML Runtimes (for example, when starting a Session or Job)
   - The remaining information on the page applies to site-level settings specific for legacy Engines.

Related Information

ML Runtimes versus Legacy Engines

Controlling User Access to Features

Cloudera Machine Learning provides Site Administrators with the ability to restrict or hide specific functionality that non-Site Administrator users have access to in the UI. For example, a site administrator can hide the models and experiments features from the ML workspace UI.

The settings on this page can be configured through the Security and Settings tabs on the Admin page.

Table 13: Security Tab

Property	Description
Allow remote editing	Disable this property to prevent users from connecting to the Cloudera Machine Learning deployment with cdswctl and using local IDEs, such as PyCharm.

Property	Description
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 Machine Learning 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 Machine Learning functionality such as file edit/read.

Table 14: Settings Tab

Property	Description
Require invitation to sign up	Enable this property to send email invitations to users when you add them to a group. To send email, an SMTP server must first be configured in Settings Email .
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 Legacy Engine users to use the Python 2 kernel	Enable this property to allow Legacy Engine users to select the Python 2 kernel when creating a job. Python 2 is disabled by default.
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. It will also not stop any experiments that have already been queued for execution.
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.
Allow users to create applications	Disable this property to hide the Applications feature in the UI. Note that this property does not affect any active applications. In particular, if you do not stop active applications before hiding the feature, they continue to serve requests and consume computing resources in the background.

Cloudera Machine Learning Email Notifications

Cloudera Machine Learning allows you to send 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). This topic shows you how to specify email address for such outbound communications.

Note that email notifications are not currently enabled by default. Emails are not sent when you create a new project. Email preferences cannot currently be configured at an individual user level.

Option 1: If your existing corporate SMTP server is accessible from the VPC where your ML workspace is running, you can continue to use that server. Go to the AdminSettings tab to specify an email address for outbound invitations and job notifications.

Option 2: If your existing SMTP solution cannot be used, consider using an email service provided by your cloud provider service. For example, Amazon provides Amazon Simple Email Service (Amazon SES).

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 AdministrationSecurity.

- User Web Browser Timeout (minutes) - This timeout sets the default maximum length of time that a web browser session 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 user logins.
- Admin 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.

Project Garbage Collection

Marks orpaned files for deletion from a project and cleans up projects that are marked for deletion.

Procedure

1. Click Site Administration Settings .
2. Scroll to Project Garbage Collection.

   Click Garbage Collect Projects to permanently delete projects marked for deletion.

   Click Clean Up Orphaned Projects to mark orphaned projects for deletion.

Results

Orphaned project files are marked for deletion. All files marked for deletion are permanently deleted when you click Garbage Collect Projects.

How to make base cluster configuration changes

When you make base cluster configuration changes, you need to restart the base cluster to propagate those changes.

In general, as Administrator you perform the following steps:

1. Make the necessary configuration changes in the base cluster.
2. Restart the base cluster.
3. In the Private Cloud compute cluster, perform the specific Kubernetes commands below to restart the ds-cdh-client pods for CML.

For ECS:

1. Access Cloudera Manager.
2. Navigate to the Containerized Cluster ECS Web UI: Clusters Your embedded Cluster ECS Web UI ECS Web UI
3. Select the namespace of your ML Workspace on the top left dropdown.
4. Navigate to Workloads Deployments .
5. Locate ds-cdh-client in the list and perform Restart from the breadcrumbs on the right.

For OCP:

Access the openshift cluster with oc or kubectl, and scale the deployment of ds-cdh-client down and back up. Use the following commands.

1. oc scale deployment/ds-cdh-client --namespace <ml-namespace> --replicas 0
2. oc scale deployment/ds-cdh-client --namespace <ml-namespace> --replicas 1

Disable Addons

As a Private Cloud Administrator, you should ensure that Spark Runtime Addons used on your site that are compatible with the base cluster version. In practice, this means you should disable the incompatible versions that may be installed.

Note: Changing the default Hadoop CLI Runtime Addon causes jobs, models, and application workloads to be unable to start up. Please see the Release Notes for more information.

1. Go to the CDP management console, and determine the base cluster version.
2. In each CML workspace, in Site Administration Runtime ,
   - Set the Hadoop addon as default that has the base cluster version in its name.
   - Keep those Spark addons enabled that have the base cluster version in their name and disable other Spark Addons.
   - If some workloads have been configured to use a disabled Spark Addon, the affected workloads must be reconfigured to use an enabled Spark Addon. This can happen in the event a workspace is upgraded.