

Cloudera AI

Site Administration

Date published: 2020-07-16

Date modified: 2025-10-31

CLOUDERA

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

Legal Notice

© Cloudera Inc. 2025. 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.....	6
Service Accounts.....	7
Creating a machine user and synchronizing to workbench.....	7
Synchronizing machine users from the Synced team.....	7
Run workloads using a service account.....	8
Configuring Quotas.....	8
Managing multiple CPU and GPU workloads using Resource Groups.....	9
Modifying workloads of resource groups.....	10
Creating Resource Profiles.....	11
Creating Resource profiles.....	14
Disable or deprecate Runtime addons.....	15
Onboarding Business Users.....	16
Adding a collaborator.....	17
User Roles.....	17
Business Users and Cloudera AI.....	18
Managing your Personal Account.....	19
Creating a Team.....	19
Managing a Team Account.....	21
Managing a Synced Team.....	21
Monitoring Cloudera AI Activity.....	22
Tracked User events.....	23
Monitoring User Events.....	26
Monitoring active Models across the Workbench.....	27
Monitoring and alerts.....	28

Application polling endpoint.....	29
Choosing default engine.....	29
Controlling User access to features.....	30
Setting custom Spark configurations at workbench-level.....	32
Configuring Job Retry settings.....	32
Cloudera AI email notifications.....	34
Downloading diagnostic bundles for a workbench.....	34
Web session timeouts.....	35
Project garbage collection.....	35
Ephemeral storage.....	35
Ports used by Cloudera AI.....	37
Export Usage List.....	38
Private cluster support.....	38
Enable a private cluster.....	38
User Defined Routing (UDR).....	39
Embed a Cloudera AI application in an external website.....	41
Setting up Cloudera AI Workbenches for high volume Workloads.....	42
Host name required by Learning Hub.....	43
Configuring external authentication with LDAP and SAML.....	44
Configuring LDAP/Active Directory authentication.....	44
LDAP general settings.....	44

LDAP group settings.....	45
Test LDAP Configuration.....	46
Configuring SAML authentication.....	46
Configuration options.....	47
Configuring HTTP Headers for Cloudera AI.....	48
Enable HTTP security headers.....	49
Enable HTTP Strict Transport Security (HSTS).....	49
Enable Cross-Origin Resource Sharing (CORS).....	50
SSH Keys.....	50
Personal key.....	50
Team key.....	50
Adding an SSH key to GitHub.....	51
Creating an SSH tunnel.....	51
Hadoop authentication for Cloudera AI Workbenches.....	51
Cloudera AI and outbound network access.....	52
Non-transparent proxy and egress trusted list.....	52

Managing Users

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

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



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 workbench. To access the site administrator dashboard:

1. Go to the Cloudera AI 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.

Cluster Metrics Snapshot	
Domain	ml-0f18c5b6-ff9.eng-ml-d.xcu2-8y8x.dev.cldr.work
Total Nodes	1
Total Memory	29.19 GiB
Used Memory	21.16 GiB
Total vCPUs	7.63
Used vCPUs	8.30
Total GPUs	0
Used GPUs	0

Monitoring Users

The Users tab on the **Administrator** dashboard displays the complete list of users. You can see which users are currently active, and when a user last logged in to Cloudera AI. You can search for a user by entering their User ID, Username, or Email in the User quick find box. To modify a user's username, email or permissions, click the Edit button under the **Action** column.



Note: The Disabled checkbox does not have any effect when external authentication is in use.

Synchronizing Users

You can synchronize users within an Cloudera AI Workbench with those users 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 Workbench.

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



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

Synchronizing Groups

Groups of users can be created in the Cloudera management console and imported to Cloudera AI. However, changes made in Cloudera do not automatically update in Cloudera AI. You need to manually trigger an update, using Sync Teams. For more information, see *Creating a Team*.

Related Information

[Cloudera AI email notifications](#)

[Creating a Team](#)

Service Accounts

Service accounts are used by machine users that require a user account, without needing to use an account of an actual user.

Like other users, this machine user can be granted necessary permissions and roles, and be added as a collaborator to projects in order to run workloads. Machine users can also create projects and workloads.

Creating a machine user and synchronizing to workbench

The MLAdmin role is required to create machine users.

Procedure

1. In Management Console, go to User Management.
2. In Actions, click Create Machine User.
3. Enter a name for the machine user and click Create.
4. In the Cloudera AI Workbenches UI, find your workbench and in Actions, click Manage Access.
5. Search for the machine user name you just created, and in Update Resource Roles, assign the MLWorkspaceAdmin or MLWorkspaceUser role. Click Update Roles.



Note: Machine users can alternatively be assigned to environments.

6. Return to the workbench in Cloudera AI, and in Site Administration Users, click Run Sync Now to manually synchronize the users for the workbench.
7. In Site Administration, search for the machine user name.

Synchronizing machine users from the Synced team

You can synchronize machine users that are part of a synced team to your project.

Procedure

1. In Management Console User Management Groups, click Create Group.
2. Enter the name for the group, and click Create.
3. Click Add Members to search for and add group members, including machine users.
4. To add the team (group) to your environment, go to Environments Actions Manage Access.

5. Click Update Role to update the role as follows, and click Update Roles.
 - Environment User: Only users who have read access to the environment are synced. Alternatively, you can assign the Environment User role to the machine user.
 - MLAdmin or MLUser role: only users with either role are synced to Cloudera AI Workbench.
6. Click Synchronize Users and wait for synchronization to complete. Then return to your Cloudera AI Workbench.
7. In Site Administration Teams , select Sync Teams and then choose the group to synchronize.
8. Click Create Team, and the team is created in Cloudera AI.

What to do next

To add members to a synced team, add them in the control plane and synchronize them to Cloudera AI via the Site Administration Teams Sync Teams option. You cannot add users to a group manually in Cloudera AI.

To add the service user as a collaborator of the project, see the instructions in [Adding a collaborator](#) and further details in [Adding project collaborators](#).

Run workloads using a service account

You can run various types of workloads using a service account. First, make sure the service account is available in your project.

1. Create a project, or enter an existing project.
2. In Collaborators, add the service account. Specify the Operator or Admin role and click Add.

Run a job with a service account

1. In Jobs, click New Job.
2. For Run Job as, select Service Account and choose the account from the list.
3. Make other settings as needed, and click Create Job.

Run an application with a service account

1. Click New Application.
2. For Run Job as, select Service Account and choose the account from the list.
3. Make other settings as needed, and click Create Application.

Run a model with a service account

1. In Models, click New Model.
2. For Deploy Model as, select Service Account and choose the account from the list.
3. Make other settings as needed, and click Deploy Model.

Configuring Quotas

This topic describes how to configure CPU, GPU, and memory quotas for users of an Cloudera AI Workbench.

Before you begin

Required Role: MLAdmin

Make sure you are assigned the MLAdmin role in Cloudera. Only users with the MLAdmin role will be logged into Cloudera AI Workbenches with Site Administrator privileges.

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

Procedure

1. Log in to the web interface.
2. Click Cloudera AI Workbenches, then open the workbench 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 workbench.

If your workbench 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) .

Cloudera AI displays the Edit default quota dialog box.

6. Enter the CPU, Memory, and GPU quota values that should apply to all users of the workbench.
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.

What to do next

To specify the maximum number of replicas in a model deployment, go to Site Administration Settings Model Deployment Settings . The default is 9 replicas, and up to 199 can be set.

Managing multiple CPU and GPU workloads using Resource Groups

Learn how to use Resource Groups and Resource Profiles to provision and manage multiple CPU and GPU environments within your Cloudera AI workbench. This provides enhanced control over workload scheduling and allows for the segregation of workloads based on instance types.

This feature eliminates the previous limitation of having only one CPU and a maximum of one GPU instance group per workspace. You can now provision multiple groups of both CPU and GPU instances as needed.

Resource Group

Resource Group serves as a parent entity that defines the specific type of node (for example, m5.xlarge) where your workloads will execute. Resource Groups are provisioned using the Cloudera AI Control Plane, and every workbench requires at least one CPU resource group, which is identifiable by a unique name specified when creating the resource group. While multiple resource groups can share the same instance type, a single resource group maps to only one type, and its name or instance type cannot be edited from within the Cloudera AI Workbench.

Resource Profile

Resource Profiles are an existing Cloudera AI feature that has been enhanced and is now segregated into CPU and GPU profiles. A resource profile defines the allowed resource combination (CPU cores, memory, and GPU count) that a workload will consume on a node.

These profiles are uniquely configured for one Resource Group (establishing a one-to-many relationship where one Resource Group can have many Resource Groups). While the Resource Group dictates the type of node, the Resource Profile dictates the specific resource combination used, and its defined resource capacity must never exceed the maximum capacity of its associated Resource Group.



Note: When using API, a Resource Profile definition (for example, 2 CPU / 4GB) can be reused across multiple Resource Groups. However, when scheduling a workload using the API, you must use the specific, unique Profile ID associated with the desired Resource Group.

Workload Scheduling

The Resource Group feature provides enhanced control over workload scheduling by ensuring that a specific workload lands on the desired type of node. When creating any workload, such as a Session, Job, Application, or Model, the user now selects two key fields, the Resource Group, which defines the instance type (node type) to be used, and the Resource Profile, which dictates the specific CPU, memory, and GPU combination the workload will consume on that chosen node.

Resource Group feature transition and legacy profiles

When you upgrade your environment from a version where the Resource Group feature was disabled to a version where it is enabled, the behavior of your existing Resource Profiles will change, as they must now be linked to a specific Resource Group.

This transition involves the following key changes:

- **Legacy profile visibility**

All existing legacy Resource Profiles that were created without an associated Resource Group will no longer be visible in the UI after the upgrade.

Consequently, when a user attempts to launch a new Session, Job, Application, or Model, they will not see the unassociated legacy profile listed. These profiles are effectively inactive until they are manually linked to a Resource Group by an Administrator.

- **Automatic default profile creation**

During the upgrade process, the system automatically creates the minimum required Resource Group and its associated profiles to ensure basic functionality:

- **Default CPU Resource Group:** The system creates one default CPU Resource Group.
- **Default Profiles:** The system automatically attaches two default resource profiles to this new CPU Resource Group.

These two newly created default profiles are the only Resource Profiles visible immediately after the upgrade, until administrators manually restore or create others. This ensures that users have at least a basic set of profiles available to start workloads.

Modifying workloads of resource groups

When provisioning a workbench, administrators can add multiple CPU and GPU Resource Groups beyond the minimum requirement. Each Resource Group is automatically configured with two Resource Profiles. Specifically, a CPU Resource Group will contain CPU Resource Profiles, and a GPU Resource Group will contain GPU Resource Profiles.

About this task

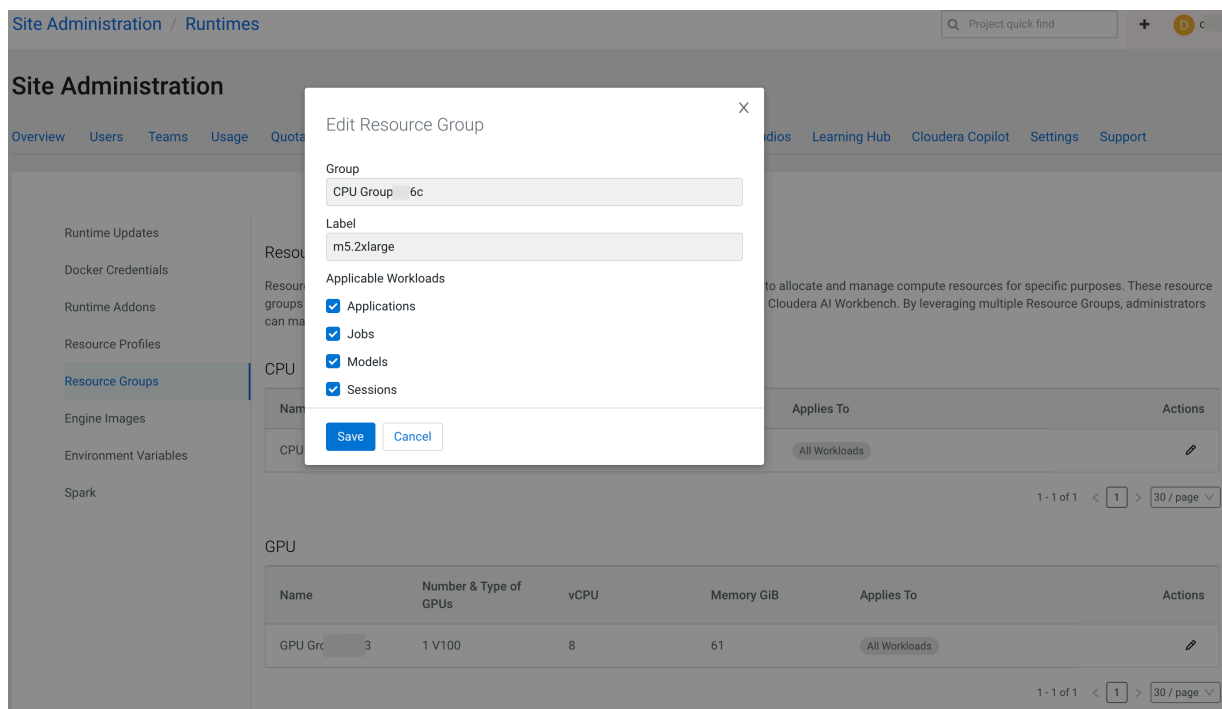
Administrators can see a new Resource Groups field in the Site Administration UI, to tailor the resource usage to specific workload requirements.

Before you begin

The required Cloudera AI version is *2.0.52-B60 OR ABOVE*.

Procedure

1. In the Cloudera console, click the Cloudera AI tile.
The **Cloudera AI Workbenches** page displays.
2. Click on the name of the workbench.
The workbench **Home** page displays.
3. Click Site Administration in the left navigation pane.
4. Select Runtimes tab.
5. On the Runtimes page, click the Resource Groups option.
6. Under the Actions menu, click the Edit button to specify the applicable workloads for the specific resource group.
Only the workloads you select will be allowed to schedule on this resource group.



7. Click Save.

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

Every Resource Group contains its respective Resource Profiles. By default, every Resource Group is configured with a minimum set of Resource Profiles:

- CPU Resource Group: Contains a minimum of two default CPU Resource Profiles.
- GPU Resource Group: Contains a number of default GPU Resource Profiles determined by the number of GPUs in that group. For instance, if the group is configured with 4 GPUs, it will include two default profiles for GPU=1, two for GPU=2, and two for GPU=4.

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, application, job, or a model, users will be able to select one of the available resource profiles depending on their project's requirements.

Administrators can add, edit, and delete CPU or GPU Resource Profiles from the Resource Profiles page.

Procedure

1. To create resource profiles, go to the [Site Administration Runtime](#) page.
2. Click Add CPU Resource Profile.

X

Add CPU Resource Profile

* Resource Group

CPU Group e76c

▼

* CPU

4

* Memory


20

Save

Cancel


3. Select the Resource Group name, and specify the CPU and Memory.
4. Click Save.

5. Click Add GPU Resource Profile.



Add GPU Resource Profile

* Resource Group

GPU Group dd83

* GPU

1

* CPU

2

* Memory

23

Save

Cancel

6. Select the Resource Group name, and specify the GPU, CPU, and Memory.
7. Click Save.

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

By default, the Enable CPU Bursting option is selected for the Resource Profiles to use burstable CPU settings to help better resource utilization. To use the resource profile as a hard limit on vCPU consumption, disable this CPU bursting option.

Results

If there are two worker nodes and 10 vCPU available overall, if one user tries to establish a session with 8 vCPU, Cloudera AI 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' interface. On the left, a sidebar lists navigation options: Overview, Users, Teams, Usage, Quotas, Models, **Runtimes**, Data Connections, Security, AMPs, AI Studios, Learning Hub, Cloudera Copilot, Settings, and Support. The 'Runtimes' section is active, showing 'Resource Profiles'. Under 'CPU Resource Profiles', there is a table with two entries:

Resource Group	Description	vCpu	Memory	Actions
Default CPU Group	2 vCPU / 4 GiB Memory	2	4	[Edit] [Delete]
Default CPU Group	4 vCPU / 8 GiB Memory	4	8	[Edit] [Delete]

The 'Memory' column header and the value '4' in the first row are highlighted with a red box. An orange arrow points from this box to the 'Start A New Session' dialog. In this dialog, the 'Resource Group' dropdown is open, showing three options: 'Default CPU Group', '2 vCPU / 4 GiB', and '4 vCPU / 8 GiB'. The '2 vCPU / 4 GiB' option is selected. The dialog also shows settings for Editor (JupyterLab), Kernel (Python 3.10), Edition (Nvidia GPU), Version (2025.09), and Runtime Image.

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 AI 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 2: Resource profiles available when launching a session

The screenshot shows the 'Site Administration' interface with the 'Runtimes' tab selected. On the left, the 'Resource Profiles' section is expanded. The 'CPU Resource Profiles' table lists two profiles:

Resource Group	Description	vCpu	Memory	Actions
Default CPU Group	2 vCPU / 4 GiB Memory	2	4	[Edit] [Delete]
Default CPU Group	4 vCPU / 8 GiB Memory	4	8	[Edit] [Delete]

An arrow points from the 'Memory' column (specifically the value '4' in the first row) to the 'vCPU/Memory GiB' dropdown in the 'Start A New Session' dialog. The dialog shows the 'Resource Group' dropdown set to 'Default CPU Group' and the 'vCPU/Memory GiB' dropdown with three options: '2 vCPU / 4 GiB', '2 vCPU / 4 GiB', and '4 vCPU / 8 GiB'.

Disable or deprecate Runtime addons

Disable or deprecate a Spark Runtime addon.

About this task

You can disable or deprecate any Spark Runtime addon from the Runtime/Engine tab of Site Administration.

Procedure

1. Select Site Administration in the left Navigation bar.
2. Select the Runtime/Engine tab.

3. Select Disabled or Deprecated from Actions next to any *SPARK* addon.

Site Administration / Runtime/Engine

Runtime Updates

☒ Enable Runtime Updates

New Runtime variants and versions are automatically downloaded and made available on clusters with Internet access. Unchecking this checkbox can disable this feature.

Hadoop CLI Version: Hadoop CLI - CDP 7.2.8 - HOTFIX...

Runtime Addons

Status	Name	ID	Component	Created At	Reason	Actions
Available	Hadoop CLI - CDP 7.2.10 - HOTFIX-1 JAVA 8U342	1	HadoopCLI	11/07/2022 12:54 PM		
Available	Hadoop CLI - CDP 7.2.11 - HOTFIX-4 JAVA 8U342	5	HadoopCLI	11/07/2022 12:54 PM		
Available	Hadoop CLI - CDP 7.2.14 - JAVA 8U342	4	HadoopCLI	11/07/2022 12:54 PM		
Available	Hadoop CLI - CDP 7.2.8 - HOTFIX-1 JAVA 8U342	6	HadoopCLI	11/07/2022 12:54 PM		
Deprecated	Spark 2.4.8 - CDE 1.15 - HOTFIX-1	2	Spark	11/07/2022 12:54 PM	PodCompleted:	
Disabled	Spark 3.2.0 - CDE 1.15 - HOTFIX-2	3	Spark	11/07/2022 12:54 PM	PodCompleted:	

< 1 >



Note: You can also return the status to Available using Actions.

Onboarding Business Users

There are two procedures required for adding Business Users to Cloudera AI. First, an Administrator 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 Cloudera AI Workbench 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 Cloudera AI Workbench, go to **Site Administration > Users**, and click **Synchronize Users**. This adds the necessary Users defined at the Environment level to the workbench, 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

Complete the following steps as a Project owner:

Procedure

1. In the Cloudera console, click the Cloudera AI tile.
The Home page displays.
2. Select the required Workbench.
The Cloudera AI Workbench page displays.
3. Click **Projects** in the left navigation pane and select the required project.
4. Go to **Collaborators**, and enter the user ID in the Search box.
5. Choose the User ID, and click **Add**. The user or team is added with their role displayed.

Results

When the Business User logs in, they can access the Applications within this project.

User Roles

Users in Cloudera AI are assigned one or more of the following roles.



Important: Cloudera on cloud allows customers to maintain full ownership and control of their data and workloads and is designed to operate in some of the most restricted on cloud environments. Since Cloudera on cloud runs in a customer's cloud account, Security and Compliance is a shared responsibility between Cloudera and its on cloud customers. User roles form the first layer of security for securing the Cloudera AI workloads.

It is your responsibility to diligently allocate the permissions to the users. For more information, see *Cloudera's Shared Responsibility Model*.

There are two categories of roles: environment resource roles, which apply to a given Cloudera environment, and workbench resource roles, which apply to a single workbench. To use workbench resource roles, you may need to upgrade the workbench or create a new workbench.

If a user has more than one role, then the role with the highest level of permissions takes precedence. If a user is a member of a group, it may gain additional roles through that membership.

Environment resource roles

- **MLAdmin:** Grants a Cloudera user the ability to create and delete Cloudera AI Workbenches within a given Cloudera environment. MLAdmins also have Administrator level access to all the workbenches provisioned within this environment. They can run workloads, monitor, and manage all user activity on these workbenches. They can also add the MLUser and MLBusinessUser roles to their assigned environment. This user also needs the account-level role of IAMViewer, in order to access the environment Manage Access page. To create or delete workbenches, this user also needs the EnvironmentAdmin role.
- **MLUser:** Grants a Cloudera user the ability to view Cloudera AI Workbenches provisioned within a given Cloudera environment. MLUsers are also able to run workloads on all the workbenches provisioned within this environment.
- **MLBusinessUser:** Grants permission to list Cloudera AI Workbench for a given Cloudera environment. MLBusinessUsers are able to only view applications deployed under the projects that they have been added to as a Business User.

Workbench resource roles

Workbench roles are for users who are granted access to only a single workbench.

- **MLWorkspaceAdmin:** Grants permission to manage all Cloudera AI workloads and settings inside a specific workbench. To perform resource role assignment, the IAMViewer role is also needed. A user with this role can administer the workbench, but is not able to utilize Cloudera APIs that modify a workbench (for example, creating or upgrading workbenches).
- **MLWorkspaceBusinessUser:** Grants permission to view shared Cloudera AI applications inside a specific workbench.
- **MLWorkspaceUser:** Grants permission to run Cloudera AI workloads inside a specific workbench.

Using the workbench resource roles

A power user or account administrator must assign the first MLWorkspaceAdmin to a workbench. Subsequently, if the MLWorkspaceAdmin also has the IAMViewer role, they can assign resource roles to the workbench.

An MLAdmin (an environment resource role) is not automatically able assign workbench resource roles on the Manage access page. A role such as MLWorkspaceAdmin is needed to do this.

You can check the permissions for a given resource role by clicking the Information icon by each resource role shown in User Management, on the Resources tab for a user, or in a Cloudera user profile.



Note: Any user that lists users or assigns resource roles also needs the account-level role of IAMViewer.

Business Users and Cloudera AI

A user is treated as a Business User inside of Cloudera AI if they are granted the MLBusinessUser role on the Environment of the given Cloudera AI Workbench. Inside the workbench, a Business User is able to access and view applications, but does not have privileges to access any other workloads in the workbench.

Logging in as a Business User

When you log in as a Business User, the only page you see is the Applications page. The page shows any applications associated with any projects that you have been added to as a Collaborator, even though you do not have rights to access the other assets associated with those projects.

In order for applications to appear in your view, contact the Project Owner to add you as a Collaborator to the project. If you have not been added to any projects, or none of the projects that you have been added to have applications, the Applications page displays the message, You currently don't have any applications.

Managing your Personal Account

You can edit personal account settings such as email, SSH keys and Hadoop credentials.

About this task

You can also access your personal account settings by clicking Account settings in the upper right-hand corner drop-down menu. This option will always take you to your personal settings page, irrespective of the context you are currently in.

Procedure

1. Sign in to Cloudera AI.
2. From the upper right drop-down menu, switch context to your personal account.
3. Click Settings.

Profile

You can modify your name, email, and bio on this page.

Teams

This page lists the teams you are a part of and the role assigned to you for each team.

SSH Keys

Your public SSH key resides here. SSH keys provide a useful way to access to external resources such as databases or remote Git repositories. For instructions, see *SSH Keys*.

Related Information

[SSH Keys](#)

Creating a Team

Users who work together on more than one project and want to facilitate collaboration can create a Team. Teams enable you to efficiently manage the users assigned to projects.

About this task

Team projects are owned by the team, rather than an individual user. Team administrators can add or remove members at any time, assigning each member different permissions. A team cannot be deleted and at least one member must be there in the team.

Site Administration

Overview Users **Teams** Usage Quotas Models Runtime Data Connections Security AMPs Learning Hub Settings Support

Create Team

* Name

Description

Team Type
☐ Local ☒ Synced Team

Add Groups

Name	Role
	Viewer
	Operator
	Contributor
	Admin

No data

Procedure

1. In Site Administration Teams , select New Team.

2. Enter the name of the team.

3. Select Local or Synced Team.

Cloudera manages the member data of a Synced Team. The members and information about the members of a Local team is not managed by Cloudera.

4. If Synced Team is selected, under Add Groups, select a group name and the role for the group and click Add. You can add multiple groups and roles using the Add option.



Note: By default, each member will inherit the role of the groups they belong to. If a member belongs to multiple groups, their effective role in the team is the highest role assigned to the member (Viewer < Operator < Contributor < Admin).

5. Enter a Description, if needed.

6. Add or invite team members. Team members can have one of the following privilege levels:

- Viewer - The Viewer has read-only access to team projects. The Viewers cannot create new projects within the team but can be added to existing ones.
- Operator - The Operator has read-only access to team projects. Additionally, Operators can start and stop existing jobs in the projects that they have access to.
- Contributor - The Contributor has write-level access to all team projects to all team projects with Team or Public visibility. The Contributor can create new projects within the team. They can also be added to existing team projects.
- Admin - The Administrator has complete access to all team projects, can add new team members, and modify team account information. The creator of the team is assigned the Administrator privilege, and can also assign other team members the Administrator privilege. Each team must have at least one Administrator user.

7. Select Create Team.

8. Select Sync Teams to update the teams with information in the Management Console.

Managing a Team Account

Team administrators can modify account information, add or invite new team members, and view/edit privileges of existing members.

Procedure

1. From the upper right drop-down menu, switch context to the team account.
2. Click Settings to open up the Account Settings dashboard.
3. Modify any of the following settings:

Profile

Modify the team description on this page.

Members

You can add new team members on this page, and modify privilege levels for existing members.

SSH Keys

The team's public SSH key resides here. Team SSH keys provide a useful way to give an entire team access to external resources such as databases. For instructions, see *SSH Keys*. Generally, team SSH keys should not be used to authenticate against Git repositories. Use your personal key instead.

Related Information

[SSH Keys](#)

Managing a Synced Team

Team administrators and Site administrators can view members of a group, delete a group within a team, update roles for a group within a team, and update a custom role for a member within a group.

Viewing members of a group

You can view the members of a group along with their roles for a particular group within a team.

1. In the Cloudera console, click the Cloudera AI tile.
The **Home** page displays.
2. Click **Site Administration** in the left navigation menu.
3. Click Teams tab.
4. In the Teams page, click on the group name to view the members' information.
5. In the Groups tab, all the groups and their role is displayed.
6. Click the Members tab to view the list of members, their highest role, and all the groups that they belong to.

The Inherit label is displayed if the role is inherited as part of the group. The Custom Role Set label is displayed if a custom role is assigned to the user.

Adding a custom role for a member

By default, members inherit the role of the group. You can set a custom role for a specific member within a group.

1. In the Cloudera console, click the Cloudera AI tile.
The **Home** page displays.
2. Click **Site Administration** in the left navigation menu.
3. Click Teams tab.
4. In the Teams page, click on the group name.

5. Click the Members tab.

List of the members, their highest role, and all the groups that they belong to is displayed.

6. Under Role, select the role you want to assign to the member from the drop-down list.

The Custom Role Set label is displayed after a custom role is assigned to the user.



Note: When you change the role of the group, the custom role of the member does not change. You must change the custom role to the Inherit role from the Role drop-down list for the member to inherit the group's role.

Updating the role of a group within a team

You can update the role for a particular group within a team.

1. In the Cloudera console, click the Cloudera AI tile.

The **Home** page displays.

2. Click **Site Administration** in the left navigation menu.

3. Click Teams tab.

4. In the Teams page, click on the group name.

5. In the Groups tab, under Role, select the role you want to assign to the group from the drop-down list.

The new role will be implemented for all members who inherit their roles within the group. Members with custom roles will remain unaffected.

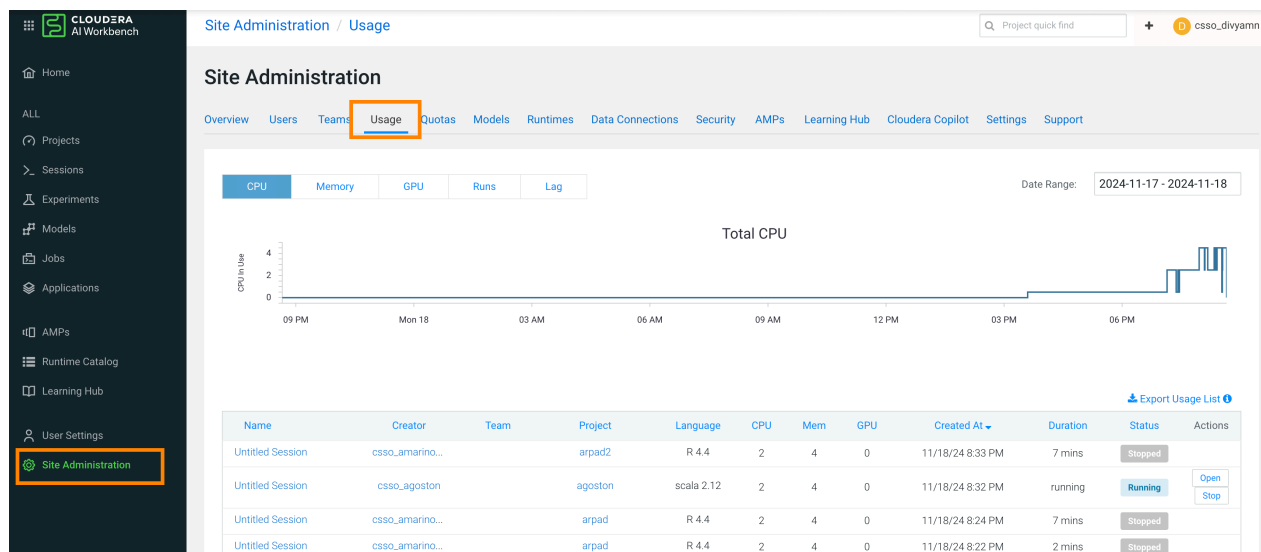
Monitoring Cloudera AI Activity

This topic describes how to monitor user activity on an Cloudera AI Workbench.

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 AI 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 AI 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 Cloudera AI 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 AI.

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.
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	-

Event	Description
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	<p>The logged event indicates when the sharing status for a job is changed from one of the following options to another:</p> <ul style="list-style-type: none"> • 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	-
update session sharing status	<p>The logged event indicates when the sharing status for a session is changed from one of the following options to another:</p> <ul style="list-style-type: none"> • All anonymous users with the link • All authenticated users with the link • Specific users and teams

Table 10: Events Related to Admin Settings

Event	Description
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	-

Event	Description
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 PostgreSQL database that is embedded within the Cloudera AI deployment to monitor or audit user events.

About this task

Querying the PostgreSQL database that is embedded within the Cloudera AI deployment requires root access to the Cloudera AI Master host.

Procedure

1. SSH to the Cloudera AI 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 AI 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 Workbench

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

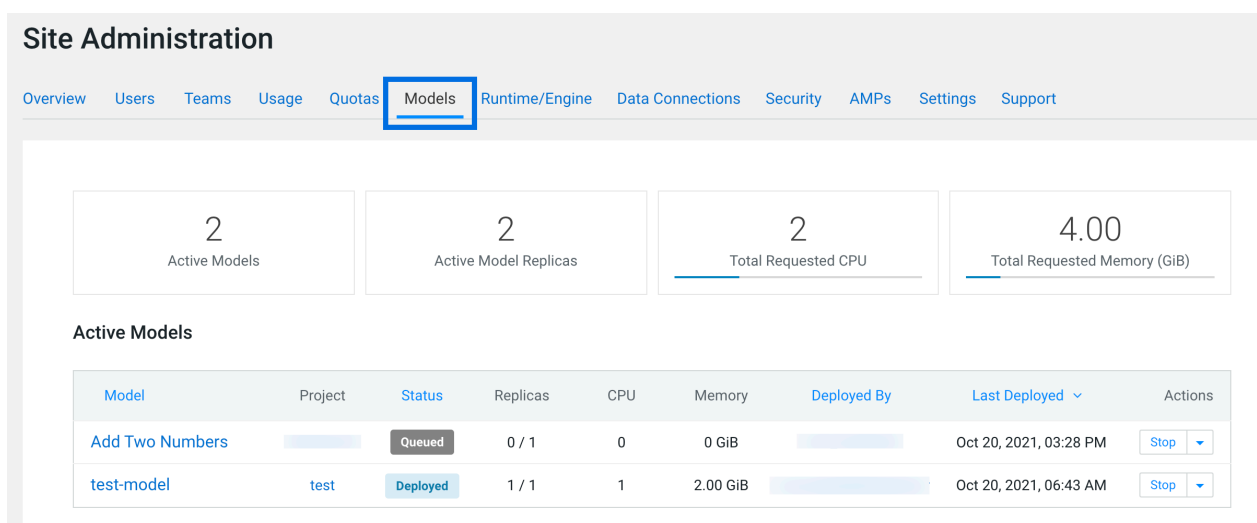
What is an active Model?

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

Monitoring all active Models across the Workbench

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.



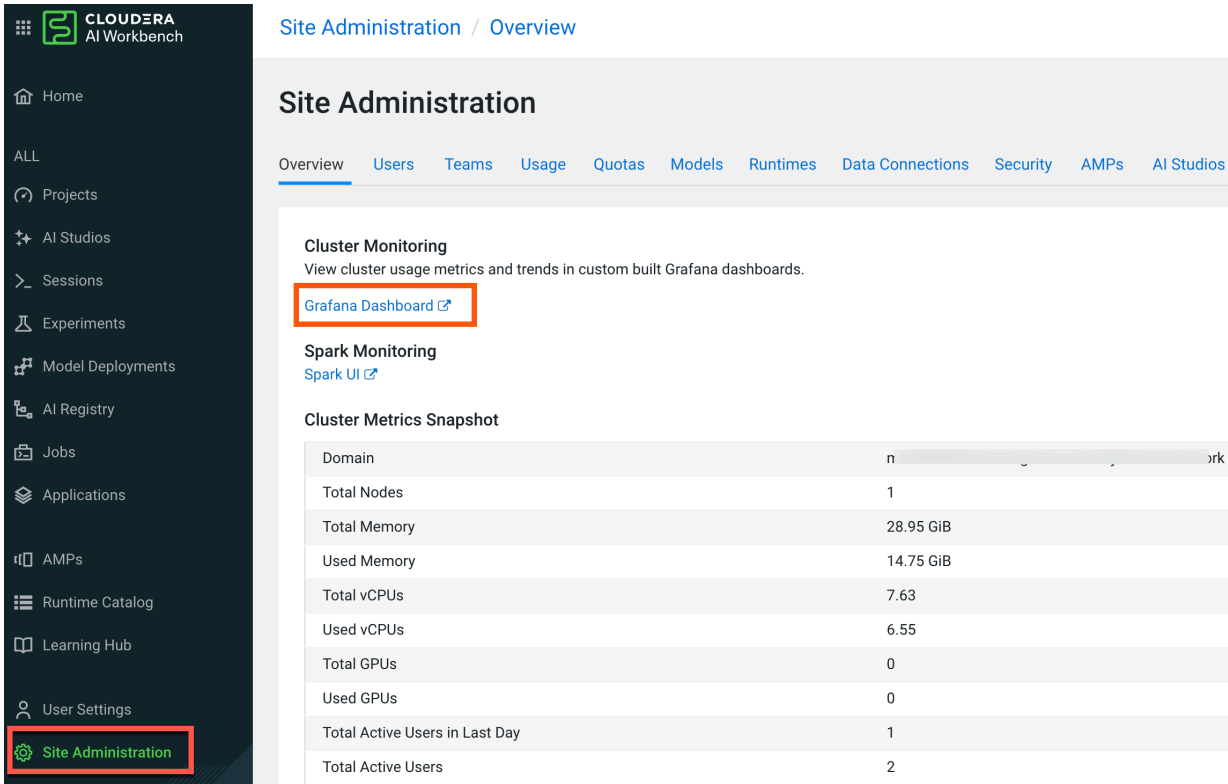
Monitoring and alerts

Cloudera AI leverages Cloudera 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 Cloudera AI Workbenches.

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, Cloudera AI 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.

1. In the Cloudera console, click the Cloudera AI tile.
The **Cloudera AI Workbenches** page displays.
2. Click on the name of the workbench.
The workbench **Home** page displays.
3. Click Site Administration in the left navigation pane.

4. In the Overview tab, click Grafana Dashboard.



Information about CPU, memory, storage, and other resources are being consumed by your workbench is displayed.

Related Information
[Grafana documentation](#)

Application polling endpoint

The Cloudera AI 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_ENDPOINT 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 on premises, the corresponding role is EnvironmentAdmin.

Make sure you are assigned the MLAdmin role in Cloudera. Only users with the MLAdmin role will be logged into Cloudera AI Workbenches with Site Administrator privileges.

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

Legacy engines contain the machinery necessary to run sessions using all four interpreter options that Cloudera AI 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 web interface.
2. Click Cloudera AI Workbenches , then open the workbench 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 workbench.



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 Runtime (for example, when starting a Session or Job)
 - The remaining information on the page applies to site-level settings specific for legacy Engines.

Controlling User access to features

Cloudera AI 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 Cloudera AI Workbench UI.

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

Table 13: Security Tab

Property	Description
Allow remote editing	Disable this property to prevent users from connecting to the Cloudera AI 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 workbench 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 workbench. When disabled, Cloudera AI 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 AI 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.

Setting custom Spark configurations at workbench-level

Administrators can configure custom Spark settings at the Cloudera AI Workbench level. These configurations will then be applied to all projects and newly launched Spark sessions within that workbench. Non-administrator users can view the applied configurations, but cannot modify them at this level.

About this task

Understanding Spark configuration layers and precedence

Spark configuration layers applied to workbenches have the following hierarchy and precedence:

1. Project-level configurations that are set in the `spark-defaults.conf` file: These are configurations set within specific project files and have the highest precedence, overriding any workbench-level defaults. For details on setting project-level defaults, see [Spark configuration files](#).
2. Custom workbench level: These are the custom settings you can configure in this topic, applied by administrators at the workbench level. The configurations in the workbench defaults are applied unless overridden in the custom workbench level.
3. Workbench defaults level: These are the default Spark configurations applied by the Cloudera AI Workbench system to all Spark sessions. Users can view these defaults, which are displayed in an uneditable textbox.

Procedure

1. In the Cloudera console, click the Cloudera AI tile.
The **Cloudera AI Workbenches** page displays.
2. Click on the name of the workbench.
The workbench **Home** page displays.
3. Click Site Administration in the left navigation pane.
4. Select Runtimes tab.
5. On the Runtimes page, scroll down to find the Spark Configuration section.

You see two main text areas. The Cloudera AI Workbench Defaults area displays the default Spark configurations applied by Cloudera AI. This section is not editable.

The Custom workbench level for spark on Kubernetes workloads area is the editable textbox in which you can enter your custom Spark properties.

6. Enter the desired custom Spark properties in the Custom workbench level for spark on Kubernetes workloads textbox. Each property must be on a new line, typically in the `key=value` format, similar to a `spark-defaults.conf` file.
7. Click Save Spark Configuration located at the bottom right of the section.

Configuring Job Retry settings

The Job Retry feature enables automatic retries of jobs based on their terminal execution states, namely failed, timed out, or skipped. It also supports concurrent execution of job retry runs, ensuring that scheduled job runs remain unaffected and are not blocked by retry processes. Users have the flexibility to configure various options to define the retry behavior. These retries are fully automated, eliminating the need for manual intervention.

About this task

The Administrator can define default values for the Job Retry parameters and only the Administrator can configure a hard limit on the maximum number of job retry runs that can be executed alongside normal job runs. This setting must only be enabled if you want to manage and limit resource usage for job retry runs.

Procedure

1. In the Cloudera console, click the Cloudera AI tile.

The Cloudera AI Workbenches page displays.

2. Click on the name of the workbench.

The workbench Home page displays.

3. Select Site Administration in the left Navigation pane.

4. Select the Settings tab.

5. Select Job Retry Configuration Limit Concurrent Retries .

6. Enable Limit Concurrent Retries by selecting the checkbox.

Enabling this option sets a limit to how many job retry runs (at maximum) can be active at the same time.

7. Define the limit value for Maximum Concurrent Retry Limit.

The Maximum Concurrent Retry Limit specifies the maximum number of job retry runs that can execute concurrently across the entire workbench, regardless of the total number of jobs running.

If the maximum limit value defined as Maximum Concurrent Retry Limit is reached, any additional job retry runs are rescheduled until the number of active retry runs falls below the limit.

Enable this hard limit only if job retry runs are consuming excessive resources, otherwise, avoid setting a hard limit.

Administrators can set this value if the Limit Concurrent Retires option is enabled.

8. Under Default Settings for all jobs, select Enable Retry to enable a retry run for the job.

If the administrator configures these settings, the specified values automatically populate the fields in the new job creation form when a user creates a job. In this case, the Job Retry settings act as default values that the administrator can recommend to users. If the administrator does not configure these settings, the fields remain blank in the new job creation form. In both cases, users have the flexibility to customize these values during job creation or update them later through the job settings page.

Define the following parameters for Job Retry:

- **Maximum Retry** – The maximum number of retry attempts which can be triggered for a single job run in case of continuous failure of retry job runs.

The minimum value is 1.

- **Retry Delay (minutes)** – The delay between two consecutive retry job runs for a failed instance of the run.

The minimum value is 1 minute.

- **Retry Conditions** – Different options can be configured to control the terminal states of a job run that trigger a retry. The Retry process completes as soon as at least one (or more) option is selected.

Select at least one of the following criteria if Retry is enabled, but you can select any combination of the following Retry Conditions options:

- **Script Failure** – Runs the Retry process for user script failures if the user script exits with a non-zero exit code after the execution of the script.
- **System Failure** – Runs the Retry process for any kind of system- or engine-related failures not including user script failures.
- **Timed-out Runs** – Runs the Retry process for timed-out job runs.
- **Skipped Runs** – Runs the Retry process for skipped job runs.

9. Click on Update to save the settings.

Cloudera AI email notifications

Cloudera AI 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 Cloudera AI Workbench is running, you can continue to use that server. Go to the `Admin Settings` 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).

Mail relay hosts often limit the authenticated sender reply address. Make sure to select a No reply email which you are allowed to use, otherwise email sending may fail.

Downloading diagnostic bundles for a workbench

This topic describes how to download diagnostic bundles for an Cloudera AI Workbench.

Before you begin

Required Role: MLAdmin

Make sure you are assigned the MLAdmin role in Cloudera. Only users with the MLAdmin role will be logged into Cloudera AI Workbenches with Site Administrator privileges.

Procedure

1. Log in to the web interface.
2. In Cloudera AI Workbenches, select a workbench.
3. Select `Site Administration Support Generate Log Archive`.
4. Select the time period from the dropdown.
5. Ensure `Include Engines` is selected if engine logs are needed (included by default).
6. Select `Send to Cloudera` to send the diagnostic logs to Cloudera Support.
7. Select `Create` to generate the logs.
8. When Status is Complete, select `Download` to download the diagnostics bundles to your machine.

What to do next

The data in the contained bundles may be incomplete. If it does not contain logs for time period you are looking for, there are a number of possible reasons:

- There is a delay between the time the logs are initially generated by a workload and the time they are visible in cloud storage. This may be approximately 1 minute due to buffering during streaming, but can be significantly longer due to eventual consistency in the cloud storage.
- Another user or process may have deleted data from your bucket; this is beyond the control of Cloudera AI.
- There may be a misconfiguration or an invalid parameter in your request. Retrieving logs requires a valid cloud storage location to be configured for logging, as well as authentication for Cloudera AI to be set up properly for it. Requests must pertain to a valid engine in a valid project.

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 orphaned 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`.

Ephemeral storage

Ephemeral storage space is scratch space that a Cloudera AI session, job, application or model can use. This feature helps in better scheduling of Cloudera AI pods, and provides a safety valve to ensure runaway computations do not consume all available scratch space on the node.

By default, each user pod in Cloudera AI is allocated 0 GB of scratch space, and it is allowed to use up to 10 GB. These settings can be applied to an entire site, or on a per-project basis.

How Spark uses ephemeral storage in Cloudera AI

Spark drivers and executors write shuffle files, spilled RDD/DataFrame blocks, broadcast variables, and task logs under directories referenced by `SPARK_LOCAL_DIRS`.

On Kubernetes these paths are mounted as one `emptyDir` volume per pod; `emptyDir` is wiped as soon as the pod terminates, so the data is **ephemeral**.

If this volume fills up, the kubelet evicts the pod and Spark surfaces errors such as:

- `java.io.IOException: No space left on device`
- `org.apache.spark.shuffle.MetadataFetchFailedException`

This is followed by a Kubernetes event similar to Evicted: The node was low on resource:#ephemeral#storage

How does the CML UI map to Kubernetes resources

Table 15: Mapping to Kubernetes Resources

CML field	Pod spec element	What it does
Ephemeral Storage (GB) – Request	resources.requests.ephemeral-storage	Scheduler bin#packing & cluster#autoscaler logic
Ephemeral Storage (GB) – Max	resources.limits.ephemeral-storage	Hard ceiling; usage ##limit # pod eviction

Both the driver and every executor inherit the values you set here (or an override in Project#Settings###Advanced

Sizing guidelines for common Spark workloads

Table 16: Sizing Guidelines

Work#load pattern	Rule of thumb (across all executors)	Rationale
SQL/ETL with light aggregations	##1#×#largest input size	Minimal shuffle spill
Joins, `groupByKey`, heavy shuffle	2#–#3#×#largest input size	Shuffle writes often exceed input volume
ML pipelines with .cache() / .persist()	Cached dataset size × #replicas	Cached blocks are duplicated

Quick workflow: Start with a generous limit, run once, open Spark UI##Executors#Shuffle Spill#(Disk) and set the per#pod limit to peak#spill ÷##executors

Tips to reduce Spark's scratch#disk footprint

Table 17:

Goal	Knob	Notes
Fewer shuffle bytes	spark.sql.shuffle.partitions (closer to number of executors) and spark.sql.adaptive.enabled=true	Adaptive Query Execution coalesces partitions on the fly
Eliminate shuffle joins	Broadcast the small side: '/*+ BROADCAST(t) */'	Keeps data in RAM when feasible
Compress spill data	Ensure spark.shuffle.compress=true (default)	Small CPU cost, large disk savings
Use RAM#backed volumes (SSD#less nodes)	spark.kubernetes.local.dirs.tmpfs=true and raise spark.kubernetes.{driver,executor}.memoryOverheadFactor	Mounts emptyDir as tmpfs
Persist scratch across pod restarts	Mount a PVC at /spark-local with spark.kubernetes.executor.volumes.persistentVolumeClaim.<name>.mount.path	Gives Spark a dedicated disk

Change site-wide ephemeral storage configuration

In Site Administration Settings Advanced , you can see the fields to change the ephemeral storage request (minimum) and maximum limit.

Ephemeral Storage Settings**Ephemeral Storage Request (in GB)**

The amount of scratch space requested by the session pod.

Ephemeral Storage Limit (in GB)

The maximum amount of scratch space the session pod is permitted to use. Kubernetes terminates the pod if it exceeds this limit.

Update

Override Site-wide ephemeral storage configuration

If you want to customize the ephemeral storage settings, you can do so on a per-project basis. Open your project, then click on **Project Settings Advanced** and adjust the ephemeral storage parameters.

Ephemeral Storage Settings

The amount of scratch space requested by the session pod. The value set here is for the specific project.

Ephemeral Storage Request

GB

The maximum amount of scratch space the session pod is permitted to use. Kubernetes terminates the pod if it exceeds this limit. The value set here is for the specific project.

Ephemeral Storage Limit

GB

Apply

Click on the below button to reset the project-level ephemeral storage values to match the values set on site level.

Reset Ephemeral Storage

AWS Known Issues

There is a known issue with the cluster autoscaler that affects autoscaling from 0->1 if a non-zero value for Ephemeral Storage Request is set. This affects both CPU and GPU node groups of the Cloudera AI Workbench. The autoscaler throws the following error when this happens:

```
pod didn't trigger scale-up: 1 Insufficient ephemeral-storage
```

This is occurring even though the nodes in the Cloudera AI autoscaling groups have sufficient ephemeral storage space in their group templates. See this [github issue](#) for details. Even though the issue is closed, the problem still persists.

The issue only affects node groups that have [0, x] autoscaling range.

Set the Ephemeral Storage Request value to 0 in both the site-wide and project settings if you run into this issue.

Ports used by Cloudera AI

Certain ports must be accessible through the firewall for proper operation of Cloudera AI.

Firewall restrictions must be disabled across Cloudera AI and Cloudera cluster hosts. Internally, the Cloudera AI master and worker hosts require full connectivity with no firewalls. Externally, end users connect to Cloudera AI exclusively through a web server running on the master host, and therefore do not need direct access to any other internal Cloudera AI or Cloudera services.

Communication with the Cloudera cluster	Cloudera AI -> Cloudera As a gateway service, Cloudera AI must have access to all the ports used by Cloudera and Cloudera Manager .
Communication with the Web Browser	The Cloudera AI web application is available at port 80. HTTPS access is available over port 443.

Export Usage List

You can export a list of sessions, jobs, workers, and experiments. You can either download a complete list of workloads or you can filter the workloads by date to download a more concise list. Timestamps in the list are given in Coordinated Universal Time (UTC).

Procedure

1. Select Site Administration in the navigation pane.
2. Select the Usage tab.
3. If you want a list of workloads specific to a date range, you can filter the list of workloads by setting the Date Range.
4. Select Export Usage List to download the list of workloads.

Results

The list downloads to your computer as a .csv file.

Private cluster support

Each type of network architecture supported by Cloudera has a unique set of tradeoffs among ease of setup, security, workloads (Experiences) supported, and so on.

on premises Clusters provide a simple way to create a secure cluster, where the API server and the workloads themselves only rely on private IP addresses that are not accessible from the internet. Connectivity to the cluster from the Cloudera control plane is provided by the Cluster Connectivity Manager v2 (CCM v2). CCMv2 uses an agent running in the cluster, and an inverting proxy running on Cloudera, which creates a HTTPS tunnel between the workload and the control plane.



Note: on premises clusters are Generally Available (GA) on AWS, but are in Preview on Azure, and require an entitlement.

Enable a private cluster

To enable a private cluster, select the option when provisioning the workbench.

Procedure

1. In Cloudera AI Workbenches, select Provision Workbench.
2. Enter a Workbench Name, and select Environment.
3. Select the Advanced Options toggle.
4. In Network Settings, select Enable Fully Private Cluster.

5. Make any other settings needed, and select Provision Workbench.

Network Settings

Subnets ⓘ

Select Subnets

Load Balancer Source Ranges ⓘ

0.0.0.0/0

☐ Enable Fully Private Cluster

☐ Enable Public IP Address for Load Balancer

Results

The workbench is provisioned using a fully private cluster.

What to do next

User Defined Routing (UDR)

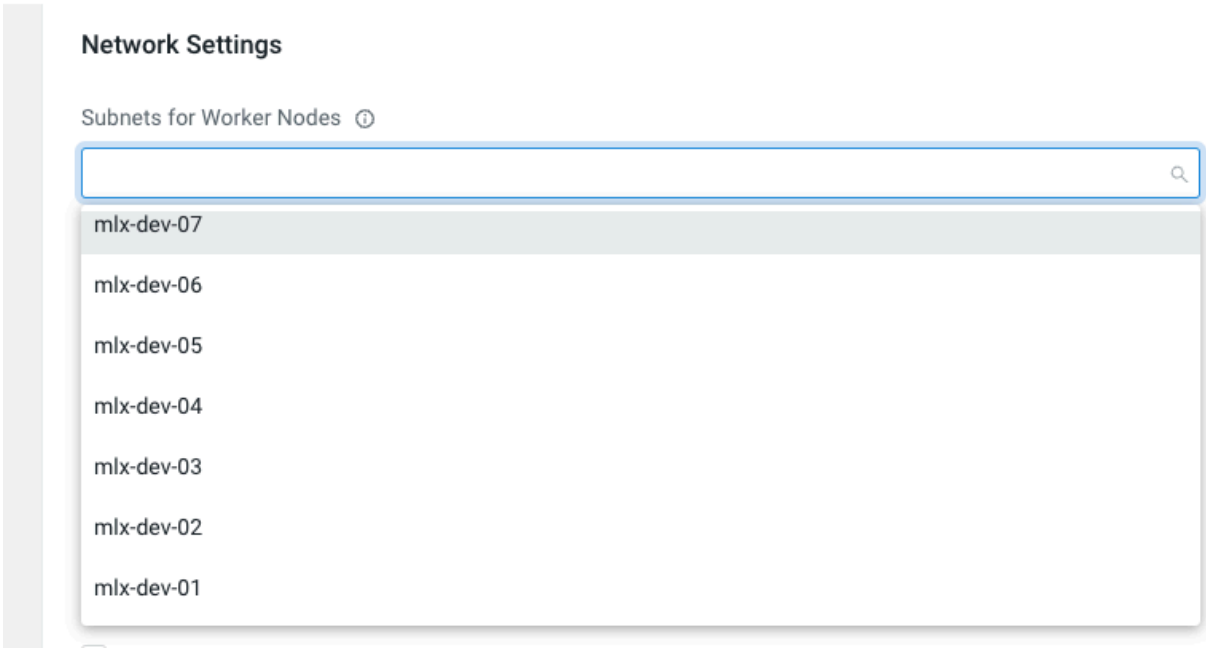
With the Fully Private Cluster configuration, Azure still creates some public IP resources to support load balancer egress. If necessary, you can avoid creating public IP addresses in the Cloudera AI cluster by using a User Defined Routing (UDR) table. A UDR table can be configured in the cluster subnet to route packets to a customer-configured firewall, for example to limit internet access or analyze traffic. For more information on setting up UDR, see the Microsoft articles [Virtual appliance scenario](#) or [Virtual network traffic routing](#).

About this task

To utilize a UDR and firewall in the Azure Cloudera AI on premises cluster, select the following when setting up the cluster.

Procedure

- 1. Select a subnet with a default route configuration to forward the traffic to the network appliance or firewall.



- 2. Create load balancers with private IP addresses. This is the default choice when creating clusters in Cloudera AI.

Network Settings

Subnets for Worker Nodes ⓘ

mlx-dev-01 ▾

Load Balancer Source Ranges ⓘ

0.0.0.0/0 - +

☒ Enable Fully Private Cluster

3. Select Enable User Defined Routing.

Network Settings

Subnets for Worker Nodes ⓘ

mlx-dev-01

Load Balancer Source Ranges ⓘ

0.0.0.0/0

☒ When enabled, the Azure CML Workspace will use the configured UDR.

☒ Enable User Defined Routing ⓘ

Embed a Cloudera AI application in an external website

You can embed a Cloudera AI application into an I-frame on a web page. You need to specify the frame-ancestor attribute, otherwise the browser security policy will prevent the application from rendering in the page.

The frame-ancestor attribute prevents "Clickjacking" attacks by specifying which domains are allowed to provide embedded content to your site. To enable a domain to embed a Cloudera AI application, set the environmental variable CDSW_FRAME_ANCESTORS to contain one or more websites as follows:

- The name of the embedding website, specified in host-source form.
- You can specify multiple websites as a comma-separated list.

You can set the environmental value in **Project Settings Advanced**.

Resource Profile

1 vCPU / 2 GiB Memory

Environment Variables

CDSW_APP_POLLING_ENDPOINT

/

CDSW_FRAME_ANCESTORS

http://*.dev.cldr.work, https://*.dev.cldr.w

Environmental variables will override the [project environment](#).

For more information on host-source form, see: *CSP: frame-ancestors*.

Related Information

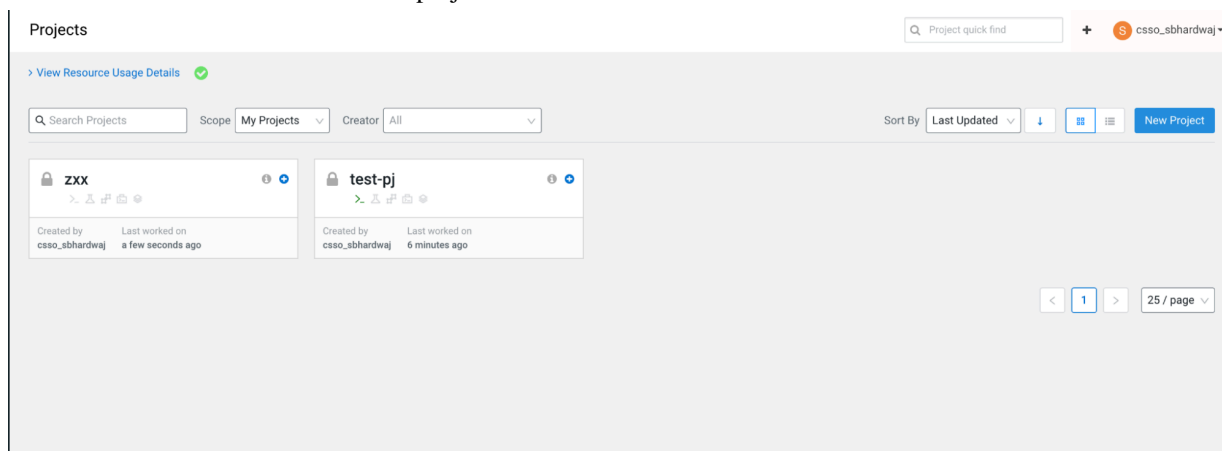
[CSP: frame-ancestors](#)

Setting up Cloudera AI Workbenches for high volume Workloads

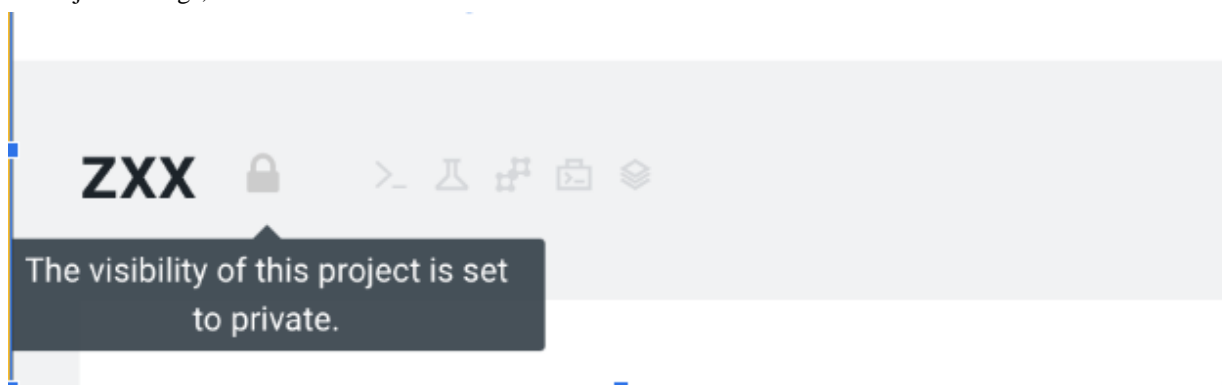
Autoscaling in Cloudera AI enables seamless scaling up of clusters, accommodating sessions, experiments, model metrics, jobs, and applications with increased user demand. In order to ensure seamless functioning of Cloudera AI Workbenches with high volume workloads, core Cloudera AI flows such as workbench suspend, resume, backup, upgrade, and editing are validated with high volume workloads. Additionally, clusters can be efficiently downsized by adjusting the autoscale range in workloads without disrupting control plane capabilities.

To prepare a Cloudera AI Workbench for high volume workloads, you need to perform the following steps to modify the pod quota limit for the Cloudera AI tenant, and then do the same for the Cloudera AI Workbench.

1. Go inside the workbench and create a project.



2. In Project settings, click the lock button.



3. In Advanced settings, add the key `OVERRIDE_PODQUOTA` and enter the value for the new pod limit to set.

csso_sbhardwaj / zxx / Project Settings / Advanced

Project Settings

General Runtime **Advanced** SSH Tunnels Data Connections Delete Project

Environment Variables

Set project environment variables that can be accessed from your scripts.

Environment variable **values** are only visible to **collaborators** with **write** or higher access. They are a great way to securely store confidential information such as your AWS or database credentials. Names are available to all users with access to the project.

CDSW_APP_POLLING_ENDPOINT	/	- +
PROJECT_OWNER	csso_sbhardwaj	- +
OVERRIDE_PODQUOTA	1	- +

Submit

Verified flows and configurations

1. Enhanced Pod Count Limit

Scaling up to 250 Pods within a single user namespace is verified, assuming that the workbench possesses sufficient computational resources and storage capacity to accommodate these pods.

2. Auto-Scaling Certification

The Control Plane actions have been verified to scale up to 100 Nodes.

3. Suspension and resumption of 100 Nodes has been successfully tested.

4. Workbench Management

- Backup of workbenches with up to 100 Nodes.
- Upscale and downscale of cluster nodes between 1-100 is now verified by changing the worker node range within the edit section of workbench.
- Successfully tested upgrading a 100 node Workbench from version at 'V-1'.

5. Basic sanity testing of the following workload flows was conducted for workbenchess with 100 Nodes:

- Creation of Sessions, Applications, Jobs, Model Metrics and Experiments for a user after workbench creation and upgrading.
- Deletion of Sessions, Applications, Jobs, Model Metrics and Experiments for high volume workloads.
- I/O, memory usage validation during upscale and downscale operations.

Known issues

There are a few known issues that have been identified during the certification process. These issues are currently being addressed.

1. During concurrent creation of sessions via Cloudera AI Workbench users, some pods may fail to come up, with a ~3% failure rate.
2. For high volume workbenches, Cloudera AI Workbench applications may fail to restart after modify/resume and upgrade operations. In this case, manually restart the affected application.

Host name required by Learning Hub

Learning Hub requires internet access to link to the displayed content. Learning Hub cannot be supported on a fully airgapped cluster.

The following domain must be added to the allow list so that links from the content will work:

- *.raw.githubusercontent.com

Configuring external authentication with LDAP and SAML



Important: Cloudera recommends you leverage Single Sign-On for users via the Cloudera Management Console. For instructions on how to configure this, see [Configuring User Access to Cloudera AI](#). If you cannot do this, we recommend contacting Cloudera Support before attempting to use the LDAP or SAML instructions provided in this section.

Cloudera AI supports user authentication against its internal local database, and against external services such as Active Directory, OpenLDAP-compatible directory services, and SAML 2.0 Identity Providers. By default, Cloudera AI performs user authentication against its internal local database. This topic describes the signup process for the first user, how to configure authentication using LDAP, Active Directory or SAML 2.0, and an optional workaround that allows site administrators to bypass external authentication by logging in using the local database in case of misconfiguration.

Configuring LDAP/Active Directory authentication

This topic describes how to set up LDAP authentication for a workbench.



Important: This is not the recommended method to set up LDAP authentication. Cloudera recommends you use the Cloudera Management Console to set this up: [Configuring User Access to Cloudera AI](#).

Cloudera AI supports both search bind and direct bind operations to authenticate against an LDAP or Active Directory directory service. The search bind authentication mechanism performs an ldapsearch against the directory service, and binds using the found [Distinguished Name \(DN\)](#) and password provided. The direct bind authentication mechanism binds to the LDAP server using a username and password provided at login.

You can configure Cloudera AI to use external authentication methods by clicking the Admin link on the left sidebar and selecting the Security tab. Select LDAP from the list to start configuring LDAP properties.

LDAP general settings

Lists the general settings required to configure LDAP authentication.

- **LDAP Server URI:** Required. The URI of the LDAP/Active Directory server against which Cloudera AI shall authenticate. For example, ldaps://ldap.COMPANY.com:636.
- **Use Direct Bind:** If checked, the username and password provided at login are used with the LDAP Username Pattern for binding to the LDAP server. If unchecked, Cloudera AI uses the search bind mechanism and two configurations, LDAP Bind DN and LDAP Bind Password, are required to perform the ldapsearch against the LDAP server.
- **LDAP Bind DN:** Required when using search bind. The DN to bind to for performing ldapsearch. For example, cn=admin,dc=company,dc=com.
- **LDAP Bind Password:** Required when using search bind. This is the password for the LDAP Bind DN.
- **LDAP Search Base:** Required. The base DN from which to search for the provided LDAP credentials. For example, ou=Engineering,dc=company,dc=com.
- **LDAP User Filter:** Required. The [LDAP filter](#) for searching for users. For example, (&(sAMAccountName={0}))(objectclass=person)). The {0} placeholder will be replaced with the username provided at login.

- **LDAP User Username Attribute:** Required. The case-sensitive username attribute of the LDAP directory service. This is used by Cloudera AI to perform the bind operation and extract the username from the response. Common values are uid, sAMAccountName, or userPrincipalName.

General Settings

LDAP Server URI *

ldaps://ldap.company.com

☐ **Use Direct Bind**

By default, Cloudera Data Science Workbench searches for the users by binding to the provided **LDAP Bind DN** and **LDAP Bind Password**. When checked, Cloudera Data Science Workbench will attempt to search for the user by binding to the user-provided username and password. In such case, please make sure the users have permissions to search for themselves in the LDAP server.

LDAP Bind DN *

CN=Test1 Person,OU=People,DC=company,DC=com

[Update LDAP Bind Password](#)

LDAP User Search Base *

OU=People,DC=company,DC=com

LDAP User Search Filter

objectClass=person

LDAP User Username Attribute *

sAMAccountName

When you select Use Direct Bind, Cloudera AI performs a direct bind to the LDAP server using the LDAP Username Pattern with the credentials provided on login (not LDAP Bind DN and LDAP Bind Password).

By default, Cloudera AI performs an LDAP search using the bind DN and credentials specified for the LDAP Bind DN and LDAP Bind Password configurations. It searches the subtree, starting from the base DN specified for the LDAP Search Base field, for an entry whose attribute specified in LDAP User Username Attribute, has the same value as the username provided on login. Cloudera AI then validates the user-provided password against the DN found as a result of the search.

LDAP group settings

In addition to the general LDAP settings, you can use group settings to restrict the access to Cloudera AI to certain groups in LDAP.

- **LDAP Group Search Base:** The base distinguished name (DN) where Cloudera AI will search for groups.
- **LDAP Group Search Filter:** The LDAP filter that Cloudera AI will use to determine whether a user is affiliated to a group.

A group object in LDAP or Active Directory typically has one or more member attributes that stores the DNs of users in the group. If LDAP Group Search Filter is set to member={0}, Cloudera AI will automatically substitute the {0} placeholder for the DN of the authenticated user.

- **LDAP User Groups:** A list of LDAP groups whose users have access to Cloudera AI. When this property is set, only users that successfully authenticate themselves AND are affiliated to at least one of the groups listed here, will be able to access Cloudera AI.

If this property is left empty, all users that can successfully authenticate themselves to LDAP will be able to access Cloudera AI.

- **LDAP Full Administrator Groups:** A list of LDAP groups whose users are automatically granted the site administrator role on Cloudera AI.

The groups listed under LDAP Full Administrator Groups do not need to be listed again under the LDAP User Groups property.

Figure 3: Example

If you want to restrict access to Cloudera AI to members of a group whose DN is:

```
CN=MLUsers,OU=Groups,DC=company,DC=com
```

And automatically grant site administrator privileges to members of a group whose DN is:

```
CN=MLAdmins,OU=Groups,DC=company,DC=com
```

Add the CNs of both groups to the following settings in Cloudera AI:

- LDAP User Groups: MLUsers
- LDAP Full Administrator Groups: MLAdmins

Test LDAP Configuration

Use the **Test LDAP Configuration** form to test your settings.

You can test your LDAP/Active Directory configuration by entering your username and password in the Test LDAP Configuration section. This form simulates the user login process and allows you to verify the validity of your LDAP/Active Directory configuration without opening a new window.

Before using this form, make sure you click Update to save the LDAP configuration you want to test.

Configuring SAML authentication

This topic describes how to set up SAML for Single Sign-on authentication for a workbench.



Important: This is not the recommended method to set up SSO. Cloudera recommends you use the Cloudera Management Console to set this up: [Configuring User Access to Cloudera AI](#).

Cloudera AI supports the [Security Assertion Markup Language \(SAML\)](#) for [Single Sign-on \(SSO\)](#) authentication; in particular, between an identity provider (IDP) and a service provider (SP). The SAML specification defines three roles: the principal (typically a user), the IDP, and the SP. In the use case addressed by SAML, the principal (user agent) requests a service from the service provider. The service provider requests and obtains an identity assertion from the IDP. On the basis of this assertion, the SP can make an access control decision—in other words it can decide whether to perform some service for the connected principal.



Note: The user sync feature only works with the SAML IDP provided by the control plane. If a custom SAML IDP is provided then customer has to make sure to turn usersync off. Otherwise, there is a risk that users will be deactivated and therefore causing cron jobs scheduled by users that are deactivated to fail.

The primary SAML use case is called web browser single sign-on (SSO). A user with a user agent (usually a web browser) requests a web resource protected by a SAML SP. The SP, wanting to know the identity of the requesting user, issues an authentication request to a SAML IDP through the user agent. In the context of this terminology, Cloudera AI operates as a SP.

Cloudera AI supports both SP- and IDP-initiated SAML 2.0-based SSO. Its [Assertion Consumer Service \(ACS\)](#) API endpoint is for consuming assertions received from the Identity Provider. If your Cloudera AI domain root were `cdsw.COMPANY.com`, then this endpoint would be available at `http://cdsw.COMPANY.com/api/v1/saml/acs`. SAML 2.0 metadata is available at `http://cdsw.COMPANY.com/api/v1/saml/metadata` for IDP-initiated SSO. Cloudera AI uses [HTTP Redirect Binding](#) for authentication requests and expects to receive responses from [HTTP POST Binding](#). Note: When visiting these pages, view the Page Source in the browser to see the full XML.

When Cloudera AI receives the SAML responses from the Identity Provider, it expects to see at least the following user attributes in the SAML responses:

- The unique identifier or username. Valid attributes are:
 - uid
 - urn:oid:0.9.2342.19200300.100.1.1
- The email address. Valid attributes are:
 - mail
 - email
 - urn:oid:0.9.2342.19200300.100.1.3
- The common name or full name of the user. Valid attributes are:
 - cn
 - urn:oid:2.5.4.3

In the absence of the cn attribute, Cloudera AI will attempt to use the following user attributes, if they exist, as the full name of the user:

- The first name of the user. Valid attributes are:
 - givenName
 - urn:oid:2.5.4.42
- The last name of the user. Valid attributes are:
 - sn
 - urn:oid:2.5.4.4

Configuration options

List of properties to configure SAML authentication and authorization in Cloudera AI.

Cloudera AI settings

- Entity ID: Required. A globally unique name for Cloudera AI as a Service Provider. This is typically the URI.
- NameID Format: Optional. The name identifier format for both Cloudera AI and Identity Provider to communicate with each other regarding a user. Default: urn:oasis:names:tc:SAML:1.1:nameid-format:emailAddress.
- Authentication Context: Optional. [SAML authentication context](#) classes are URIs that specify authentication methods used in SAML authentication requests and authentication statements. Default: urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport.

Signing SAML authentication requests

- CDSW Private Key for Signing Authentication Requests: Optional. If you upload a private key, you must upload a corresponding certificate as well so that the Identity Provider can use the certificate to verify the authentication requests sent by Cloudera AI. You can upload the private key used for both signing authentication requests sent to Identity Provider and decrypting assertions received from the Identity Provider.
- Cloudera AI Certificate for Signature Validation: Required if the Cloudera AI Private Key is set, otherwise optional. You can upload a certificate in the [PEM format](#) for the Identity Provider to [verify the authenticity](#) of the authentication requests generated by Cloudera AI. The uploaded certificate is made available at the `http://cdsw.COMPANY.com/api/v1/saml/metadata` endpoint.

SAML assertion decryption

Cloudera AI uses the following properties to support SAML assertion encryption & decryption.

- Cloudera AI Certificate for Encrypting SAML Assertions - Must be configured on the Identity Provider so that Identity Provider can use it for encrypting SAML assertions for Cloudera AI

- Cloudera AI Private Key for Decrypting SAML Assertions - Used to decrypt the encrypted SAML assertions.

Identity provider

- Identity Provider SSO URL: Required. The entry point of the Identity Provider in the form of URI.
- Identity Provider Signing Certificate: Optional. Administrators can upload the [X.509](#) certificate of the Identity Provider for Cloudera AI to validate the incoming SAML responses.

Cloudera AI extracts the Identity Provider SSO URL and Identity Provider Signing Certificate information from the uploaded Identity Provider Metadata file. Cloudera AI also expects all Identity Provider metadata to be defined in a <md:EntityDescriptor> XML element with the namespace "urn:oasis:names:tc:SAML:2.0:metadata", as defined in the [SAMLMeta-xsd schema](#).

For on-premises deployments, you must provide a certificate and private key, generated and signed with your trusted Certificate Authority, for Cloudera AI to establish secure communication with the Identity Provider.

Authorization

When you're using SAML 2.0 authentication, you can use the following properties to restrict the access to Cloudera AI to certain groups of users:

- SAML Attribute Identifier for User Role: The Object Identifier (OID) of the user attribute that will be provided by your identity provider for identifying a user's role/affiliation. You can use this field in combination with the following SAML User Groups property to restrict access to Cloudera AI to only members of certain groups.

For example, if your identity provider returns the OrganizationalUnitName user attribute, you would specify the OID of the OrganizationalUnitName, which is urn:oid:2.5.4.11, as the value for this property.

- SAML User Groups: A list of groups whose users have access to Cloudera AI. When this property is set, only users that are successfully authenticated AND are affiliated to at least one of the groups listed here, will be able to access Cloudera AI.

For example, if your identity provider returns the OrganizationalUnitName user attribute, add the value of this attribute to the SAML User Groups list to restrict access to Cloudera AI to that group.

If this property is left empty, all users that can successfully authenticate themselves will be able to access Cloudera AI.

- SAML Full Administrator Groups: A list of groups whose users are automatically granted the site administrator role on Cloudera AI.

The groups listed under SAML Full Administrator Groups do not need to be listed again under the SAML User Groups property.

Configuring HTTP Headers for Cloudera AI

This topic explains how to customize the HTTP headers that are accepted by Cloudera AI.

Required Role: Site Administrator

These properties are available under the site administrator panel at Admin Security .

Enable Cross-Origin Resource Sharing (CORS)

Most modern browsers implement the [Same-Origin Policy](#), which restricts how a document or a script loaded from one origin can interact with a resource from another origin. When the Enable cross-origin resource sharing property is enabled on Cloudera AI, web servers will include the Access-Control-Allow-Origin: * HTTP header in their HTTP responses. This gives web applications on different domains permission to access the Cloudera AI API through browsers.

This property is disabled by default .

If this property is disabled, web applications from different domains will not be able to programmatically communicate with the Cloudera AI API through browsers.

Enable HTTP security headers

When Enable HTTP security headers is enabled, the following HTTP headers will be included in HTTP responses from servers:

- X-XSS-Protection
- X-DNS-Prefetch-Control
- X-Frame-Options
- X-Download-Options
- X-Content-Type-Options

This property is enabled by default .

Disabling this property could leave your Cloudera AI deployment vulnerable to clickjacking, cross-site scripting (XSS), or any other injection attacks.

Enable HTTP Strict Transport Security (HSTS)



Note: Without TLS/SSL enabled, configuring this property will have no effect on your browser.

When both TLS/SSL and this property (Enable HTTP Strict Transport Security (HSTS)) are enabled, Cloudera AI will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera AI using HTTP will automatically be converted to HTTPS.

This property is disabled by default .

If you ever need to downgrade to back to HTTP, use the following sequence of steps: First, deactivate this checkbox to disable HSTS and restart Cloudera AI. Then, load the Cloudera AI web application in each browser to clear the respective browser's HSTS setting. Finally, disable TLS/SSL across the cluster. Following this sequence should help avoid a situation where users get locked out of their accounts due to browser caching.

Enable HTTP security headers

When Enable HTTP security headers is enabled, the following HTTP headers will be included in HTTP responses from servers:

- X-XSS-Protection
- X-DNS-Prefetch-Control
- X-Frame-Options
- X-Download-Options
- X-Content-Type-Options

This property is enabled by default .

Disabling this property could leave your Cloudera AI deployment vulnerable to clickjacking, cross-site scripting (XSS), or any other injection attacks.

Enable HTTP Strict Transport Security (HSTS)



Note: Without TLS/SSL enabled, configuring this property will have no effect on your browser.

When both TLS/SSL and this property (Enable HTTP Strict Transport Security (HSTS)) are enabled, Cloudera AI will inform your browser that it should never load the site using HTTP. Additionally, all attempts to access Cloudera AI using HTTP will automatically be converted to HTTPS.

This property is disabled by default .

If you ever need to downgrade to back to HTTP, use the following sequence of steps: First, deactivate this checkbox to disable HSTS and restart Cloudera AI. Then, load the Cloudera AI web application in each browser to clear the respective browser's HSTS setting. Finally, disable TLS/SSL across the cluster. Following this sequence shall help to avoid a situation where users get locked out of their accounts due to browser caching.

Enable Cross-Origin Resource Sharing (CORS)

Most modern browsers implement the [Same-Origin Policy](#), which restricts how a document or a script loaded from one origin can interact with a resource from another origin. When the Enable cross-origin resource sharing property is enabled on Cloudera AI, web servers will include the Access-Control-Allow-Origin: * HTTP header in their HTTP responses. This gives web applications on different domains permission to access the Cloudera AI API through browsers.

This property is disabled by default .

If this property is disabled, web applications from different domains will not be able to programmatically communicate with the Cloudera AI API through browsers.

SSH Keys

This topic describes the different types of SSH keys used by Cloudera AI, and how you can use those keys to authenticate to an external service such as GitHub.

Personal key

Cloudera AI automatically generates an SSH [key pair](#) for your user account. You can rotate the key pair and view your public key on your user settings page. It is not possible for anyone to view your private key.

Every console you run has your account's private key loaded into its [SSH-agent](#). Your consoles can use the private key to authenticate to external services, such as GitHub. For instructions, see [Adding an SSH Key to GitHub](#).

Related Information

[Adding an SSH Key to GitHub](#)

Team key

Team SSH keys provide a useful way to give an entire team access to external resources such as databases or GitHub repositories (as described in the next section).

Like Cloudera AI users, each Cloudera AI team has an associated SSH key. You can access the public key from the team's account settings. Click Account, then select the team from the drop-down menu at the upper right corner of the page.

When you launch a console in a project owned by a team, you can use that team's SSH key from within the console.

Adding an SSH key to GitHub

Cloudera AI creates a public SSH key for each account. You can add this SSH public key to your GitHub account if you want to use password-protected GitHub repositories to create new projects or collaborate on projects.

Procedure

1. Sign in to Cloudera AI.
2. Go to the upper right drop-down menu and switch context to the account whose key you want to add. This could be a individual user account or a team account.
3. On the left sidebar, click User Settings.
4. Go to the Outbound SSH tab and copy the User Public SSH Key.
5. Sign in to your GitHub account and add the Cloudera AI key copied in the previous step to your GitHub account. For instructions, refer the GitHub documentation on [Adding a new SSH key to your GitHub account](#).

Creating an SSH tunnel

You can use your SSH key to connect Cloudera AI to an external database or cluster by creating an SSH tunnel.

About this task

In some environments, external databases and data sources reside behind restrictive firewalls. A common pattern is to provide access to these services using a bastion host with only the SSH port open. Cloudera AI provides a convenient way to connect to such resources using an SSH tunnel.

If you create an [SSH tunnel](#) to an external server in one of your projects, then all engines that you run in that project are able to connect securely to a port on that server by connecting to a local port. The encrypted tunnel is completely transparent to the user and code.

Procedure

1. Open the Project Settings page.
2. Open the Tunnels tab.
3. Click New Tunnel.
4. Enter the server IP Address or DNS hostname.
5. Enter your username on the server.
6. Enter the local port that should be proxied, and to which remote port on the server.

What to do next

On the remote server, configure SSH to accept password-less logins using your individual or team SSH key. Often, you can do so by appending the SSH key to the file `/home/username/.ssh/authorized_keys`.

Hadoop authentication for Cloudera AI Workbenches

Cloudera AI does not assume that your Kerberos principal is always the same as your login information. Therefore, you will need to make sure Cloudera AI knows your Kerberos identity when you sign in.

About this task

This procedure is required if you want to run Spark workloads in an Cloudera AI Workbench. This is also required if connecting Cloudera Data Visualization running in Cloudera AI to an Impala instance using Kerberos for authentication.

Procedure

1. Navigate to your Cloudera AI Workbench.
2. Go to the top-right dropdown menu, click **Account settings Hadoop Authentication**.
3. To authenticate, either enter your password or click **Upload Keytab** to upload the keytab file directly.

Results

Once successfully authenticated, Cloudera AI uses your stored credentials to ensure you are secure when running workloads.

Cloudera AI and outbound network access

Cloudera AI expects access to certain external networks. See the related information *Outbound internet access and proxy* for further information.



Note: The outbound network access destinations listed in *Outbound internet access and proxy* are only the minimal set required for Cloudera installation and operation. For environments with limited outbound internet access due to using a firewall or proxy, access to Python or R package repositories such as Python Package Index or CRAN may need to be whitelisted if your use cases require installing packages from those repositories. Alternatively, you may consider creating mirrors of those repositories within your environment.

Related Information

[Outbound internet access and proxy](#)

Non-transparent proxy and egress trusted list

Cloudera AI, when used on AWS on cloud, supports non-transparent proxies. Non-transparent proxy enables Cloudera AI to proxy web requests without requiring any particular browser setup.

Egress Trusted List

In normal operation, Cloudera AI requires the ability to reach several external domains. See *Outbound internet access and proxy* for more information.

Related Information

[Outbound internet access and proxy](#)