

Using quota management (Tech Preview)

Date published: 2020-07-16

Date modified: 2023-10-31



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Quota Management overview.....	4
Creating a resource pool for CML.....	4
Enabling Quota Management in CML.....	7
Provisioning a workspace using a configured resource pool.....	8
Quota for CML workloads.....	8
Limitations for CML Quota Management.....	9

Quota Management overview

Quota management enables you to control how resources are allocated within your Cloudera Machine Learning (CML) workspace.

In order to prevent a single session, job, or other workload from consuming all of the available cluster resources, you can limit the number of CPUs and memory allocated by user, business units, or Data Service by defining resource pools that define resource limits.

Pools are organized in a hierarchical manner by defining nodes in the hierarchy with resource limits.

A regular workflow of managing quotas constitutes of the following activities:

1. Creating a resource pool for CML
2. Enabling Quota management in CML
3. Provisioning a workplace using a configured resource pool



Note:

It is recommended for CML administrators to create at least one resource pool under root.default for CML during the Technical Preview (TP) release. Do not use the root.default resource pool for creating a workspace.

Quota Management cannot be enabled for an existing workspace. It is recommended for CML administrators to provision a new workspace to enable and test this feature.



Note: This feature is in Technical Preview and not recommended for production deployments. Cloudera recommends that you try this feature in test or development environments.

Related Information

[Creating a resource pool for CML](#)

[Enabling Quota Management in CML](#)

[Provisioning a workspace using a configured resource pool](#)

[Limitations for CML Quota Management](#)

Creating a resource pool for CML

In technical preview (TP), do not use the 'root.default' resource pool. You must create at least one resource pool for quota management.

Procedure

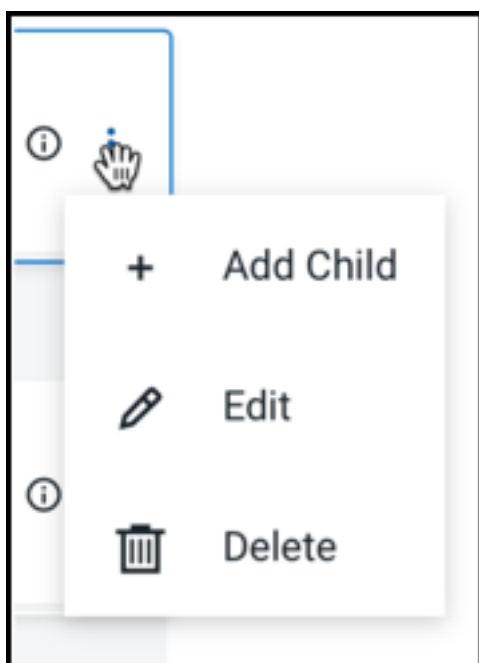
1. Go to Cloudera Manager Resource Utilization tab .
2. Under the root.default resource pool (or any already created custom resource pool), create a child node for Cloudera Machine Learning (CML).

Pools are organized in a hierarchical manner.

CML reserves 30 GB Memory and 20 CPU for the CML workspace installation. The resource pool created for CML should have at least 38 GB of Memory and 22 CPU.

3. Click the actions menu in a pool and select Add Child.

Figure 1: Add Child




The Add Quota dialog box opens.

Figure 2: Add Quota


Add Quota ✕

*** Name** ?


✓ Total Quota

 Memory

▼

 CPU

▼

 GPU

▼

> Guaranteed Quota

Validity ?

Distribution Policy ?

☒ Inelastic ☐ Elastic

Tags ?

Key	Value	✕

⊕ Add Tag

Cancel

OK

4. Enter the following information:

- Name – Enter a name for this resource pool.
- Memory – Use the sliders to select the memory allocation for the pool. Use the drop-down list to select the units.



Note: It is recommended to configure memory in units GB

- CPU – Use the sliders to select the CPU allocation for the pool. You can choose how CPUs are counted by selecting Cores or Millicores from the drop-down list.



Note: It is recommended to configure CPU units in cores.

- GPU – Use the sliders to select the GPU allocation for the pool.



Note: It is recommended to configure GPU units in cores.

- Validity – Enter the length of time that the pool can remain active. To keep the pool active indefinitely, enter NEVER_EXPIRES.
- Tags – Tags provide a way to add user-defined name/value pairs as metadata for the pools.



Note: Use the tag key: experience, value: cml, which helps CML to identify that the resource pool is configured for CML.

Note that the maximum value of the sliders is limited by the available resources, i.e. total resources of the parent pool and the amount of resources not used by other sibling pools.

Enabling Quota Management in CML

To enable Quota Management in Cloudera Machine Learning (CML) it needs to be configured. Follow the recommended configuration guidelines.

Before you begin

Set up the Kubernetes and kubectl as defined in [Prerequisites](#).

Procedure

1. Configure the kubectl using the kubeconfig file.
2. Edit the CML Control Plane deployment:

```
kubectl get deploy dp-mlx-control-plane-app -n cdp-namespace -o yaml
> file-name
```

This will save the CML control plane deployment specification.

3. Take a backup of the above file.
4. Search for the environment variable 'ENABLE_UMBRA_INTEGRATION' in the file, and change the value to true.
5. Save the deployment file and run:

```
kubectl apply -f file-name.yaml
```

Wait for the new pod to come up.

6. Verify:

```
kubectl get pods -n cdp-namespace
```

Provisioning a workspace using a configured resource pool

Configuring the quota for a specified workspace requires additional configuration settings besides provisioning the workspace for Quota Management.

Cloudera Machine Learning (CML) reserves 30 GB Memory and 20 CPU for the CML workspace installation.







The minimum requirement for a CML workspace to be provisioned is 38 GB of Memory and 22 CPU.

Umbra Integration Options

Resource Pool

root.default.cml

root.default.cml

	Memory		45	GB
	CPU		45	Cores
	GPU		0	Cores

In this example out of 45 Cores and 45 GB allocated for the resource pool root.default.cml, 20 Cores and 30 GB are reserved for the CML workspace. Out of these resources, 20 cores and 30 GBs will be used for CML, while 25 cores and 15 GBs are available to run workloads.

The rest of the resources is available in root.default.cml resource pool and can be allocated to other ML Workspaces or other Data Services.

Quota for CML workloads

A Machine Learning (ML) Workspace is allocated a set amount of resources based on configured parameters at provisioning time. Within a workspace, resources available for workloads can be further subdivided into quotas at user level.

Quota Management is implemented at user level. By default 8 GB memory and 2 vCPU cores are configured for each user. The configured resources are sufficient for running sessions but neither the spark workloads nor the executors find additional resources. CML administrators can configure custom quota for the user on the Site Administration page.

Site Administration

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

☒ ON ?
Default Quota 7 Users on this cluster

Quotas	CPU (vCPU)	Memory (GiB)	GPU
Default (per user)	2	8	0

Custom Quota Add User

Users	CPU (vCPU)	Memory (GiB)	GPU
No custom quotas have been created			

Add Custom Quota

If the quota for a user is exhausted, the workload remains in the pending state until the required resources are available.

If the quota for users is modified, it will be reflected when the next workload is submitted.



Note: Do not toggle off this default quota once you have enabled the quota management in CML.



Note: Do not edit the GPU resources on the user quota tab under Site Administration when the workspace is provisioned without any GPU resources.

GPU resources can be edited if the workspace is provisioned with GPU resources.

Related Information

[Creating a team](#)

Limitations for CML Quota Management

This section provides an overview of the limitations for Cloudera Machine Learning (CML) Quota Management.



Note: Do not use the root.default resource pool in the Technical Preview (TP) release to provision the CML workspace. Create at least one resource pool for CML.

If the resource pool gets deleted, the underlying CML workspace and workloads will be deleted. However, the stale entries will still be available in the CML UI, reflecting that the workspace still exists but the underlying applications on the cluster will be deleted.

If the quota for users is modified, it will be reflected when the next workload (session or job) is submitted.