Cloudera Data Engineering 1.5.2

Enabling and disabling Cloudera Data Engineering

Date published: 2020-07-30
Date modified: 2023-11-02

CLOUDERA

Legal Notice

Contents

Adding a Cloudera Data Engineering service

Before you can use the Cloudera Data Engineering (CDE) service, you must add the service to an environment that you want to use CDE on.

Before you begin

Make sure that you have a working environment for which you want to enable the CDE service. For more information about environments, see Environments.

Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.

2.
Click Administration on the left navigation menu, click ⊕ at the top to enable CDE service for an environment.

If the environment does not have any CDE service, the page displays a Enable a Service button that launches the same

Administration / Enable a Service

Name *

Service Name

Environment *

Select an Environment

Resource Pooling and Capacity

Resource Pool ⓘ *

Resource Pools in Quota management enable you to control how resources are allocated within your CDP Pr
It is recommended to create your own Resource Pool with restricted quota values instead of using default. Le

default.engineering

Capacity ⓘ

Set the Maximum number of resources your users can use in this CDE Service. Maximum is the limit for the C
includes resources for the CDE Service Infrastructure and all Virtual Clusters under it. Learn more 🔗

Note: CDE Infrastructure requires a minimum of 16 CPU cores and 48 GB in memory, including one Virtual C

CPU (cores) *

Maximum ⓘ

30

A Maximum capacity of 250 cores CPU can be set so that this cluster may scale under high demand situatior

Memory (GB) *

Maximum ⓘ

100

A Maximum capacity of 500 GB Memory can be set so that this cluster may scale under high demand situati

GPU (Technical Preview)

Maximum ⓘ

0

3. Enter a Name for the CDE service you are creating.

4. In the Environment text box, select or type the name of the environment that you want to enable CDE for. The displayed list dynamically updates to show environment names matching your input. When you see the correct environment, click on it to select it.

5. In Resource Pool (Technical Preview), select the name of the resource pool that you want to enable CDE service for.

   For information about configuring resource pool and capacity, see Managing cluster resources using Quota Management (Technical Preview).

6. In Capacity , use the slider to set the maximum number of CPU cores and the maximum memory in gigabytes that can be used by this CDE service.

7. GPU (Technical Preview), in Capacity , use the slider to set the maximum number of GPU cores in gigabytes that can be used by this CDE service. GPU resources are limited in the cluster and all data services like CML and CDE could share or dedicatedly set resource quotas for their experience. For information about configuring resource pool and capacity, see Managing cluster resources using Quota Management (Technical Preview).

8. Under Additional Configurations, in NFS Storage Class, specify the name of the custom NFS storage class. The storage provisioner must support `ReadWriteMany` access modes. By default, CDE uses `CephFS` provisioner in the OpenShift Container Platform and `Longhorn` provisioner in the Embedded Container Service. If it does not exist, the CDE service initialisation fails.

   You can specify the name of the Portworx storage class specified during the CDP Data Services installation to use the Portworx storage class. The storage provisioner must support ReadWriteMany access mode. You can obtain the name of the Portworx storage class from your cluster by running the `kubectl get sc` command. The CDE service and virtual clusters will now use the Portworx storage class instead of the default storage class of the platform.

   For more information, see *Installing in internet environment* and *Storage Classes*.

9. Default Virtual Cluster selection is enabled by default to create a default virtual cluster after enabling a CDE service. This helps you to create your jobs easily, without having to wait to create a CDE virtual cluster as mentioned in Creating virtual clusters, making the onboarding smoother. You can turn this toggle off if you do not wish to use a default virtual cluster.

10. Click Enable.

Results
The CDE Home page displays the status of the CDE service initialization. You can view logs for the service by clicking on the service vertical ellipsis (three dots) menu, and then clicking View Logs.
Related Information
Installing in internet environment
Storage classes

Managing a CDE Service

You can view configuration, metrics, and logs of existing CDE services. You can use the Edit option to make the configuration changes dynamically.

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Click Administration in the left navigation menu. The Administration page displays.
3. In the Services column, click the Service Details icon for the CDE service you want to manage.
4. On the Configuration tab, you can view details about the service, including the service name and CDP environment.
5. Optional: Click the Edit option to make the configuration changes dynamically which may take a few minutes to update.

You can switch between the following tabs to view additional information:

Configuration

> The Configuration tab lists details about the service name, CDP environment, and the CPU and memory capacity. You can modify the CPU and memory capacity dynamically.

Charts

> The Charts tab displays the charts related to CPU Requests, Memory Requests, Jobs, and Nodes.

Logs

> The Logs tab displays the latest log entries for the CDE service.

Diagnostics

> The Diagnostics tab provides option to generate and download diagnostics bundle.

Removing a Cloudera Data Engineering service

Disabling an existing Cloudera Data Engineering (CDE) service stops all jobs, and deletes all associated virtual clusters and virtual cluster metadata. Do not do this unless you are certain that you no longer need any of these. Disabling CDE does not delete your CDP data. If enabling the CDE service on an environment for the first time fails, you must disable the service before you can try again. In this scenario, there are no clusters or jobs, and you can safely perform this procedure.

Before you begin

Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Click Administration on the left navigation menu. The Administration page displays.
3. In the Services column, click ⋮ at the top right of the CDE service, and then click Disable Service.

   ⚠️ **Warning:** Disabling an existing Cloudera Data Engineering (CDE) service stops all jobs, deletes all associated virtual clusters and virtual cluster metadata. Do not do this unless you are certain that you no longer need any of these.

4. Confirm that you want to disable CDE by typing the environment name and then clicking Disable.

Results
The CDE Administration page displays the status of the environment that is being disabled.

What to do next
If you disabled CDE as a result of a failure to enable CDE for the first time on an environment, resolve any reported issues, and then try again.