Cloudera Data Warehouse Private Cloud 1.5.2

Cloudera Data Warehouse Private Cloud Environments

Date published: 2020-08-17
Date modified: 2023-11-02

CLOUDERA

Legal Notice

Contents

About Private Cloud environments

An environment is a logical entity that represents the association of your Private Cloud user account with compute resources. Using the compute resources of the environment, you can provision and manage Cloudera Data Warehouse (CDW), Cloudera Data Engineering (CDE), or Cloudera Machine Learning (CML) workloads.

For a CDP Private Cloud deployment, you can configure the environment on an OpenShift Container Platform or an Embedded Container Service (ECS) cluster. Deploying CDP Private Cloud Data Services (CDW, CDE, and CML) on OpenShift requires you to deploy and manage the Kubernetes infrastructure. ECS service creates and manages an embedded Kubernetes infrastructure for use with the CDP Private Cloud Experiences. To deploy CDP Experiences on ECS clusters, you only need to provide hosts on which you want to install the service and Cloudera Manager sets up an ECS cluster using an embedded Kubernetes service and provides a framework to manage and monitor the cluster.

Much of the installation and deployment configuration for private cloud is performed before you register environments using the Management Console. However, in addition to the steps described in the installation guide, you must perform additional configurations to activate an environment so you can use it with the CDW service.

The following diagram shows the components of CDP Private Cloud:
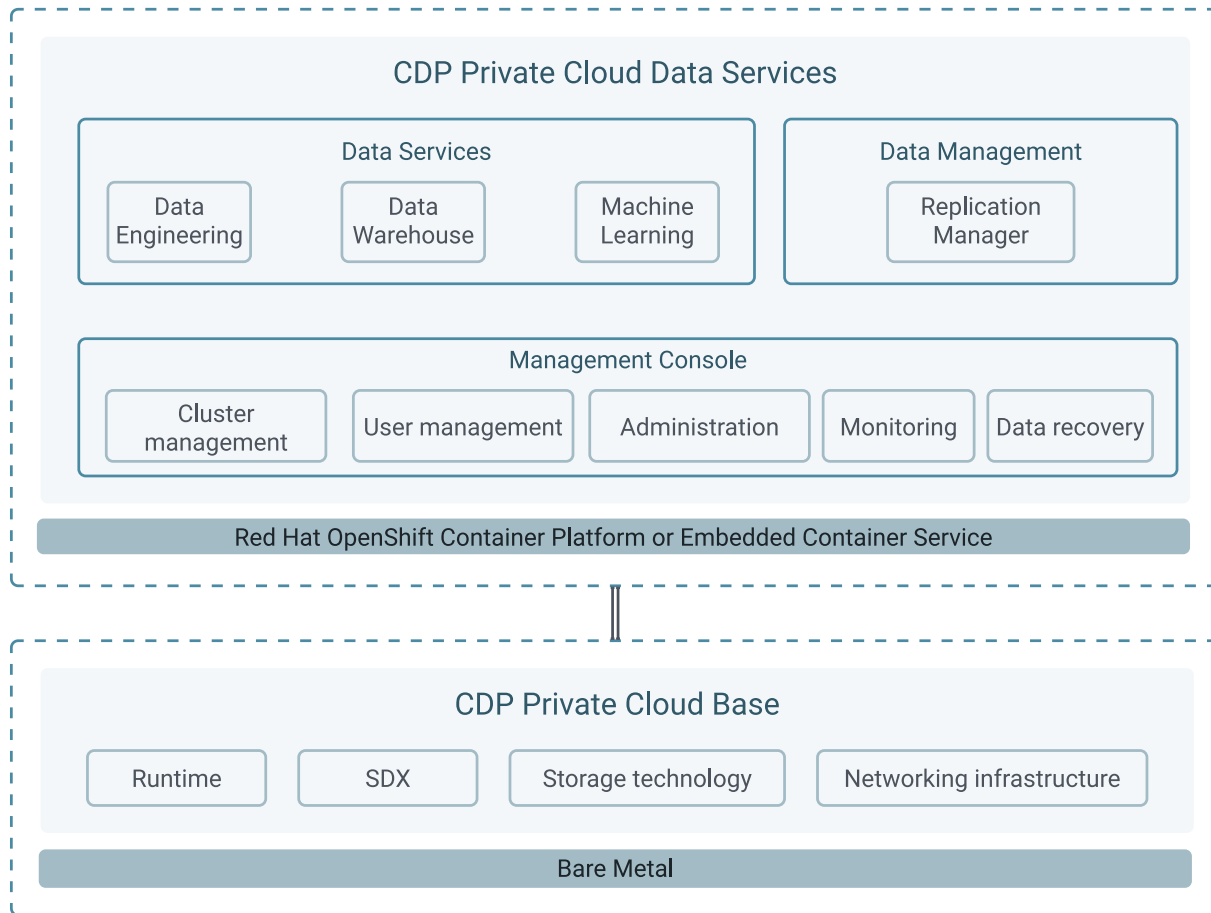
For more details about environment requirements and how to register an environment so CDP can access Kubernetes clusters on your OpenShift or ECS deployment, see CDP Private Cloud Environments.

Related Information
Activating OpenShift environments
Activating Embedded Container Service environments

Viewing environment details

You can view information about an activated environment, such as when it was created and last updated, or how many Database Catalogs and Virtual Warehouses use the environment from the Cloudera Data Warehouse (CDW) web interface.

Procedure

1. Log in to the Data Warehouse service as DWAdmin.
2. Go to the Environments tab on the **Overview** page.
3. Locate the environment you want to view and then click the ⋮ Edit .

   Environment details for the selected environment are displayed.

Creating custom pod configurations for Impala Virtual Warehouses

You can configure the resources used by Impala Virtual Warehouses in Cloudera Data Warehouse (CDW) Private Cloud environments to optimize Impala performance or to control resource usage in the environment.

About this task

When you create a Virtual Warehouse, CDW allocates standard resources to the Warehouses that are suitable for most workloads. You can control the size of the Virtual Warehouse at the time of creation by choosing the number of nodes to be used. By using custom pod configurations, you can also change the resources used by the critical Impala components, such as the coordinators, executors, and catalog daemons to pack a particular number of pods into a Kubernetes node or to create extra-large daemons to handle specific workloads.

Before you begin

Ensure that the Customize pod sizes option is selected on the Advanced Configuration Advanced Settings page.

You must have already activated an environment in CDW.

Procedure

1. Log in to the Data Warehouse service as a DWAdmin.
2. Click on the Environments tab.
3. Locate your environment and click ⋮ Edit .

   The **Environment Details** page is displayed.

4. Click the POD CONFIGURATIONS tab.

The following two read-only options are displayed: Cdw Defaults and 1 x Node. You can create a copy of these options and customize them.

- The Cdw Defaults option is selected by default. CDW uses default values for the pods if a specific pod configuration is not used.
- The 1 x Node option calculates and shows the resources based on the attributes of a random node from a cluster, assuming all of them are the same.

A pod configuration is a named resource that is configured at the environment level.

5. Click Actions corresponding to the configuration you want to copy and customize and click Make A Copy.

6. Specify a name for the new configuration in the Cloned Config Name field. Optionally, specify a description and click Create New Config.

A new pod configuration is created, which you can now customize.

7. Specify the values for the following parameters under the Coordinator section:

- Memory
- CPUs
- Xmx (maximum memory allocation pool for a Java Virtual Machine)
- Xms (initial memory allocation pool for a Java Virtual Machine)
- AC Slots (admission_control_slots flag)
- Cache size (size of the data cache)
- Scratch size (limit of Impala scratch space)
- Overhead size (size for resources used by tools run by the containers)

» test

Make A Copy   Delete

Name: test

Description :

☐ Set As Default

⌄ Coordinator

Memory   117760

CPUS   28

Xmx   29G

Xms   2G

AC Slots   128

Cache size   200GiB

Scratch size   300GiB

Overhead size   58GiB

⚠ **Important:** By default, the total space allocated for scratch, cache, and overhead is 600 GiB, with scratch=280GiB, cache=280GiB, and overhead=40GiB. You can allocate space for these components in the following format: [***QUANTITY***][***UNIT***].

The quantity can be a decimal number, separated by a period (.). For example, 240.5GiB. The unit must be in GiB (gibibytes).

1 GiB (gibibyte) equals 1.074 GB (gigabytes).

8. Specify the values for the following parameters under the Executor section:

- Memory
- CPUS
- Xmx (maximum memory allocation pool for a Java Virtual Machine)
- Xms (initial memory allocation pool for a Java Virtual Machine)
- AC Slots (admission_control_slots flag)
- Cache size (size of the data cache)
- Scratch size (limit of Impala scratch space)
- Overhead size (size for resources used by tools run by the containers)

∨ Executor

Memory	116736
CPUS	14
Xmx	4G
Xms	2G
AC Slots	36
Cache size	200GiB
Scratch size	300GiB
Overhead size	58GiB

> **Note:** By default, the total space allocated for scratch, cache, and overhead is 600 GiB, with scratch=280GiB, cache=280GiB, and overhead=40GiB. You can allocate space for these components in the following format: [***QUANTITY***][***UNIT***].
>
> The quantity can be a decimal number, separated by a period (.). For example, 240.5GiB. The unit must be in GiB (gibibytes).
>
> 1 GiB (gibibyte) equals 1.074 GB (gigabytes).

9. Specify the values for the following parameters under the Catalog section:

- Memory
- CPUS
- Xmx (maximum memory allocation pool for a Java Virtual Machine)
- Xms (initial memory allocation pool for a Java Virtual Machine)

∨ Catalog

Memory	8192
CPUS	1
Xmx	6G
Xms	2G

10. Specify the values for the following parameters under the Default Settings section:

- MaxQueryMemLimit
- MinQueryMemLimit
- mt_dop

∨ Default Settings

| MaxQueryMemLimit	53687091200
MinQueryMemLimit	2147483648
mt_dop	12

11. Make the configuration default by selecting the **Set As Default** option.

This makes a pod configuration the default configuration at the environment level.

12. Click Apply Changes under the **POD CONFIGURATION** tab to save the custom settings.

The Configuration update initiated message is displayed.

The new pod configuration becomes available in the Pod Config drop-down menu as shown in the following image. You can select this Impala pod configuration while creating a new Impala Virtual Warehouse.
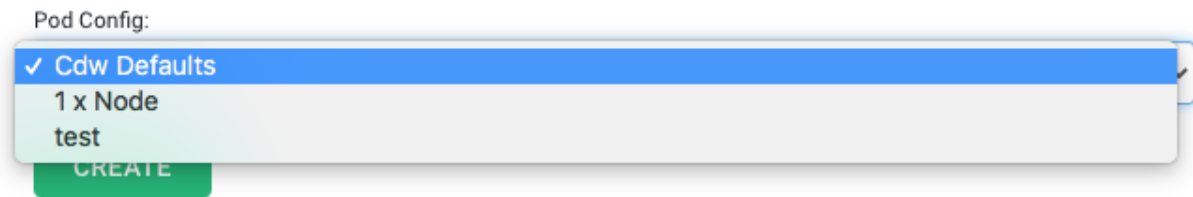
Results

While adding a new Impala Virtual Warehouse, you can select the Pod Configuration to be used for resource allocation. The default value is "Cdw Defaults", but you can select other configurations available in your environment that you created using these steps.

Scheduling executor pods on dedicated worker nodes in CDW Private Cloud

To optimize resource utilization on your cluster and improve query performance, Cloudera Data Warehouse (CDW) allows you to schedule Hive and Impala executor and coordinator pods on dedicated executor nodes that are tainted for CDW and equipped with more local storage, either using SSDs or NVMEs.

About this task

To block CDW pods, other than executors and coordinators to be scheduled on a dedicated worker node with higher local storage capacity, you must enable the Use dedicated nodes for executors option while activating an environment in CDW or by editing an existing environment. This allows nodes with available local storage to be reserved for CDW query executor pods over all other CDW or other data services pods marking them exclusively for CDW executor pods.

By default, the Use dedicated nodes for executors option is disabled. When this option is disabled, all CDW pods, including Hive MetaStore (HMS), Data Visualization, Hive and Impala executors and coordinators can be scheduled on the tainted worker nodes. When you enable this option, only Hive and Impala executor and coordinator pods can be scheduled on the tainted worker nodes.

To enable the Use dedicated nodes for executors option while activating the environment, see *Activating OpenShift environments* or *Activating Embedded Container Service environments*. This task describes how to enable the Use dedicated nodes for executors option on an existing environment.

Before you begin

On Embedded Container Service (ECS), you must dedicate the Embedded Container Service cluster nodes for specific workloads using Cloudera Manager, as described in the *Dedicating ECS nodes for specific workloads* section.

On OpenShift Container Platform (OCP), you must dedicate the OCP cluster nodes for specific workloads using the `kubectl taint` command, as described in the *Dedicate an SSD node for CDW workloads* section.

Procedure

1. Log in to the Data Warehouse service as DWAdmin.

2. Go to the **Environment(s)** tab.

3.
Locate the environment on which you want to enable this feature and click ⋮ Edit .

4. Go to the **CONFIGURATIONS** tab and enable the Use dedicated nodes for executors option.

5. Click Apply Changes.

6. Refresh the associated Database Catalogs and Virtual Warehouses.

   You can also upgrade or update the associated Database Catalogs and Virtual Warehouses depending on your needs.

Related Information

Dedicating ECS nodes for specific workloads

Dedicating OCP nodes for specific workloads

Activating OpenShift environments

Activating Embedded Container Service environments