

# Enabling and disabling Cloudera Data Engineering

Date published: 2020-07-30

Date modified: 2025-11-20



# 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

<b>AWS Graviton instances in Cloudera Data Engineering.....</b>	<b>4</b>
<b>Enabling a Cloudera Data Engineering service.....</b>	<b>4</b>
Enabling a fully private network for a Cloudera Data Engineering service for Azure.....	8
Azure private DNS zones in a Cloudera Data Engineering service.....	9
Enabling a semi-private network for a Cloudera Data Engineering service with AWS.....	13
<b>Managing a Cloudera Data Engineering Service.....</b>	<b>13</b>
<b>Renewing the Cloudera Data Engineering certificate.....</b>	<b>14</b>
<b>Suspending and resuming Cloudera Data Engineering Services - Technical preview.....</b>	<b>15</b>
<b>Cloudera Data Engineering tier types.....</b>	<b>16</b>
<b>Removing a Cloudera Data Engineering service.....</b>	<b>18</b>
<b>Limiting Incoming Endpoint Traffic for Cloudera Data Engineering Services For AWS.....</b>	<b>19</b>

# AWS Graviton instances in Cloudera Data Engineering

AWS Graviton is a general purpose, ARM-based processor family. AWS Graviton delivers currently the best price performance for cloud workloads running in AWS Elastic Compute Cloud (EC2). With AWS Graviton, you can optimize costs and achieve better performance. From the AWS Graviton family, Cloudera Data Engineering supports AWS Graviton 3.

## Prerequisites for using AWS Graviton in Cloudera Data Engineering

For information on the prerequisites, such as the Data Lake and the Spark supported versions, see [Compatibility for Cloudera Data Engineering and Runtime components](#).

If your Data Lake version is not supported by Graviton, see [Upgrading a Data Lake](#).

## Considerations for using AWS Graviton in your cluster

By default, AWS Graviton is available to all customers for Cloudera Data Engineering service compute nodes if the component version requirements are met.

To use AWS Graviton, when enabling a Cloudera Data Engineering service, select from the available Graviton instances listed under the Workload type drop-down list.

With AWS Graviton, you can use SSD instances and spot instances.

If you select the AWS Graviton workload when creating the service:

- All Virtual Clusters will run AWS Graviton. Running a subset of Virtual Clusters or jobs on non-Graviton instances is not supported.
- To check the list of Spark versions supported by Graviton, see [Compatibility for Cloudera Data Engineering and Runtime components](#).

# Enabling a Cloudera Data Engineering service

Before you can use the Cloudera Data Engineering service, you must enable it on each environment that you want to use Cloudera Data Engineering on.

## Before you begin

Make sure that you have a working environment for which you want to enable the Cloudera Data Engineering service. For more information about environments, see [Environments](#).



**Note:** Cloudera Data Engineering on Microsoft Azure does not currently support SSD or Spot instances.

## Procedure

1. In the Cloudera console, click the Data Engineering tile. The Cloudera Data Engineering Home page displays.
2. Click Administration in the left navigation menu. The Administration page displays.
3. In the Services column, click the plus icon at the top to enable Cloudera Data Engineering for an environment.
4. Type the name of the service that you want to enable Cloudera Data Engineering for.

## 5. Select the Environment from the drop-down menu.

After you select the Environment, you can view the Cloudera Runtime version, the supported Apache Airflow runtime component version, and the Spark runtime component version on the Service creation and details pages.

The following screenshot illustrates enabling a Service, where you can view the supported runtime component versions.

Administration / Enable a Service

\* Name  
Service Name

\* Environment  
test Cloudera Runtime 7.3.1

\* Workload Type ☐ Use SSD instances ⓘ  
General - m5.2xlarge

EBS Size ⓘ  
100 GB

Select your Cloudera Data Engineering version  
Set a version of the service you would like to use. Please note selecting an older version will require you to upgrade in the future to support new features.  
Need more information on compatibility of runtime components? [Review our compatibility matrix](#)

☒ Current Version (1.24.1-b193) ☐ Previous Version (1.23.1-h2-b3)

Supported Spark versions  
3.5.1

Airflow Runtime Version  
Airflow 2.10.4 Python 3.11

Need more information on compatibility of runtime components? [Review our compatibility matrix](#)

## 6. Select the Workload Type.

The workload type corresponds to the instance size that will be deployed to run your submitted Spark jobs. When you select a type, the corresponding cloud provider instance size is displayed in the Summary section to the right.

- a) If you want to use SSD storage, check the box labeled Use SSD instances. In this configuration, SSD storage is used for the workload filesystem, such as the Spark local directory. If your workload requires more space than is available in the instance storage, select a larger instance type with sufficient local storage or select an instance type without SSD, and then configure the EBS volume size.

## 7. Set the Auto-Scale Range under the Capacity & Costs section:

The range you set here creates an [auto scaling group](#) with the specified minimum and maximum number of instances that can be used. The Cloudera Data Engineering service launches and shuts down instances as needed within this range. The instance size is determined by the Workload Type you selected.

You can use Reserved Instances (RIs) in your account to save cost. Reserved Instances are not physical instances, but a billing discount that is applied to the use of On-Demand Instances in your account. To benefit from the billing discount, the On-Demand Instances in your account must match certain attributes. For more information,

see the AWS documentation on Reserved Instances. If you want to use RIs, you must also specify an instance family for them.

For Azure: Set the On-demand Instances range. This option displays for Core (Tier 1) cluster types during service creation. After the service and cluster is created, you can edit the service and set the All purpose On-demand Instances range under the Capacity & Costs section.

For AWS: After the service and clusters are created, you can set the All purpose On-demand Instances range and the All purpose Spot Instances range under the Capacity & Costs section. This option displays for the All-purpose (Tier 2) cluster types.



**Important:** For both Microsoft Azure and Amazon Web Services, to avoid being charged for All purpose compute nodes, even if you do not create an All purpose Virtual Cluster:

Navigate to the Service Details page > Configurations tab > Capacity & Costs > Autoscale Range and set the minimum value of All purpose On-demand Instances to 0.

For more information, see [Cloudera Data Engineering cluster types](#).

8. If you want to use spot instances, check the box labeled Use Spot instances and select a range of spot instances to request. This creates another auto scaling group of spot instances. Spot instances are requested with similar CPU and memory profiles as the instances selected for the Workload Type. For more information, see [Cloudera Data Engineering Spot Instances](#).



**Note:** Duplicate auto-scaling groups of the same size are created for supporting tiered pricing. These can be edited at a later time if needed.

9. Optional: Enable a private network (preview feature). This feature ensures that the Cloudera Data Engineering service is deployed with a secure network setup based on the cloud provider such as Microsoft Azure or Amazon Web Services (AWS). For more information see the links below on how to configure this feature as there are prerequisites needed before you see the Enable Private Network option in the Cloudera Data Engineering user interface under Network & Storage.
10. If you create the service in an AWS environment using a non-transparent proxy, you find a Proxy CIDR Source Ranges field. You are only required to enter the proxy CIDR ranges for the proxy instances if you registered your proxies using host names, as the Cloudera Control Plane has no way to resolve those to IPs. If your proxy instances were registered using IP addresses, you can leave this field blank.  
For example, if you have a load balanced proxy with static IPs running on 10.80.199.105 and 10.80.200.45, add "10.80.199.105/32" and "10.80.200.45/32". If your proxy instances are dynamic (behind a load balancer or virtual IP) then you would enter a wider range, e.g.: "10.80.200.0/24".



**Note:** Currently non-transparent support is only available on Amazon Web Services.

11. If you want to create a load balancing endpoint in a public subnet, check the box labeled Enable Public Loadbalancer. If you leave this unchecked, the load balancing endpoint will be created in a private subnet, and you will need to configure access manually in your cloud account.



**Note:** If you enable the [Public Endpoint Access Gateway](#) with public subnets, the load balancer is created in the public subnet, even if you do not check the Enable Public Loadbalancer option.



**Important:** When you enable the public loadbalancer, you need at least one public subnet configured in the environment.

12. Specify a comma-separated list of CIDRs in API server Authorized IP Ranges that can access the Kubernetes master API server.

You may specify a comma-separated list of CIDRs that can access the Kubernetes master API server.



**Attention:** Leaving this field empty renders the Kubernetes API server open to all traffic.

Make sure that the provided IP addresses do not overlap with the following ranges:

- 0.0.0.0 - 0.255.255.255
- 10.0.0.0 - 10.255.255.255
- 100.64.0.0 - 100.127.255.255
- 127.0.0.0 - 127.255.255.255
- 169.254.0.0 - 169.254.255.255s
- 172.16.0.0 - 172.31.255.255
- 192.0.0.0 - 192.0.0.255
- 192.0.2.0 - 192.0.2.255
- 192.88.99.0 - 192.88.99.255
- 192.168.0.0 - 192.168.255.255
- 198.18.0.0 - 198.19.255.255
- 198.51.100.0 - 198.51.100.255
- 203.0.113.0 - 203.0.113.255
- 224.0.0.0 - 239.255.255.255
- 240.0.0.0 - 255.255.255.254
- 255.255.255.255

13. Specify a comma-separated list of client IP ranges in Load Balancer Source Ranges that should be allowed to access the load balancer.
14. Specify which subnets to use for the Kubernetes worker nodes. Select from available Subnets in the drop-down list.
15. Specify which subnets to use for the Load balancer. Select from available Subnets in the drop-down list.
16. Optional: Check the box labeled Enable Observability Analytics if you want diagnostic information about jobs and query execution sent to Cloudera Observability. This helps optimize troubleshooting.
17. Check the box labeled Enable Workload Analytics to automatically send diagnostic information from job execution to [Cloudera Workload Manager](#).
18. Optionally add Tags as needed. Tags are applied to the cloud provider resources associated with the Cloudera Data Engineering service (including virtual clusters created in that service). For more information about tags, see the cloud provider documentation:

#### Amazon AWS

[Tagging AWS resources](#)

#### Microsoft Azure

[Use tags to organize your Azure resources and management hierarchy](#)



**Note:** The following tags are added automatically by Cloudera Data Engineering along with the custom Tags: "cde-cluster-id", "cde-provisioner-id", "cde-owner-email", "Cloudera-Resource-Name", "owner". After Cloudera Data Engineering 1.20.3, the "owner" tag is no longer added. If you ever rely on the "owner" tag, please use "cde-owner-email" instead.

19. Default Virtual Cluster selection is enabled by default to create a default virtual cluster after enabling a Cloudera Data Engineering service. This will help you get a jump start to create your jobs easily, without having to wait to create a Cloudera Data Engineering 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.
20. Click Enable.

**Results**

The Cloudera Data Engineering Administration page displays the status of the environment initialization. You can view logs for the environment by clicking on the environment vertical ellipsis menu, and then clicking View Logs.

**Related Information**

[Cluster Connectivity Manager](#)

[Enabling Cluster Connectivity Manager in the Management Console](#)

## Enabling a fully private network for a Cloudera Data Engineering service for Azure

Learn how to enable a fully private network setup for a Cloudera Data Engineering service for Azure services in Cloudera. Additionally, you can learn how to add User Defined Routing (UDR) in the UI or CLI. The UDR helps from exposing public IP addresses in your service.

This feature ensures that all Azure services used by Cloudera Data Engineering are provisioned as private (private Azure Kubernetes Service (AKS), MySQL, and Storage Accounts). The Azure cluster is deployed as a fully private network cluster when you enable a Cloudera Data Engineering service and enables VNet access through private endpoints and private links. Lastly, Cloudera Data Engineering on Microsoft Azure does not currently support SSD or Spot instances.

**For Cloudera Data Engineering UI**

Before you begin

- Ensure that you have created and enabled a Cloudera Data Engineering service. Additionally, the Cloudera must communicate with the Cloudera Data Engineering service on a private network in order to manage the Cloudera Data Engineering service lifecycle. This communication occurs using the Cluster Connectivity Manager v2; therefore, to enable this feature, the Cloudera environment must be enabled with the CCMv2. Once the CCMv2 is enabled at the Cloudera environment level, the Enable Private Network option displays in the Cloudera Data Engineering user interface when you enable a service. For more information on how to enable a Cloudera Data Engineering service and set up CCMv2, refer to the links below.



**Note:** To enable UDR, you must enable a private network flag and you must provide a subnet.

UI steps for enabling a private network and enabling UDR

1. While enabling a Cloudera Data Engineering service for an Azure environment, select Enable Private Network. Optionally, once you've enabled a private network on Microsoft Azure, you can select the User Defined Routing checkbox. Use this to avoid exposing public IP addresses in your service by using a user defined routing (UDR) table. After, you'll need to specify a Subnet.
2. Click Enable.

**For CDP CLI**

You can enable the user defined routing (UDR) with the CDP CLI using the `--network-outbound-type` CLI switch with a value of "UDR". See the example command:

```
./clients/cdpcli/cdp.sh de enable-service --name "test-service-cdpcli" --env "dex-priv-env" --instance-type "Standard_D8s_v4" --minimum-instance-s 0 --maximum-instances 10 --enable-private-network --subnets dex-dev.internal.19.westus2 --network-outbound-type UDR
```

**Related Information**

[Enabling a Cloudera Data Engineering service](#)

[Cluster Connectivity Manager](#)

[Enabling Cluster Connectivity Manager in the Management Console](#)



## Azure private DNS zones in a Cloudera Data Engineering service

With the Azure private DNS zones feature, you can optimize costs by leveraging your existing Azure private DNS zone resources. When creating the Cloudera Data Engineering service, you can use Azure private DNS zones for the Azure Kubernetes Service (AKS), the Storage Account File Share, and for the database.

You can create private DNS zones in the same subscription where there is the Cloudera subscription, or you can create private DNS zones in a different subscription. For this reason, you must specify the full Azure resource ID of the Azure private DNS zones for the database and for the Storage Account when adding them to your Cloudera Data Engineering service. You can obtain the full resource ID of a private DNS zone using the JSON view option in the Azure Portal in Private DNS Zones.



**Important:** Before configuring the Azure private DNS zones, you must create them in the Azure Portal (one for the AKS, one for the database, and one for the Storage Account File Share) and link the private DNS zones to the Azure virtual networks.

### Configuring an existing private DNS Zone for an AKS

At the Cloudera-level, when you provision the environment, you can configure an existing Azure Private DNS zone for an Azure Kubernetes Service (AKS). Alternatively, if you do not have an existing private DNS zone configured, the system can create a new private DNS zone for the AKS.

#### About this task



**Important:** To configure an existing Azure Private DNS zone for an AKS, you must create the environment using the CDP CLI.

If all the following statements are true, the system uses the private DNS zone that you specify in the `aksPrivateDnsZoneId` parameter while provisioning the environment for the AKS:

- The private network is enabled (the `privateNetwork.enabled` parameter is set to true)
- While provisioning the environment using CLI, the `aksPrivateDnsZoneId` parameter is configured
- The `LIFTIE_AKS_DISABLE_PRIVATE_DNS_ZONE` parameter is set to false (default)

If all the following statements are true, the system creates and enables a new private DNS zone for the AKS:

- The private network is enabled (the `privateNetwork.enabled` parameter is set to true)
- The `aksPrivateDnsZoneId` parameter is not set
- The `LIFTIE_AKS_DISABLE_PRIVATE_DNS_ZONE` parameter is set to false (default)

### Before you begin

To configure an existing Azure Private DNS zone for an AKS, you need the full resource ID of the private DNS zone. To obtain the full resource ID, use the JSON view option in Azure Private DNS Zones. Example for the full resource ID of a private DNS zone for an AKS:

```
/subscriptions/[***SUBSCRIPTIONID**]/resourceGroups/[***RESOURCEGROUP**]/providers/Microsoft.Network/privateDnsZones/privatelink.[***REGION**].azmk8s.io
```

For information on how to create the environment for Cloudera Data Engineering, see [Register an Azure environment from CDP CLI](#) and [Obtain CLI commands for registering an environment](#).



**Note:** Perform this procedure in the CDP CLI while creating the environment.

### Procedure

1. Open the CDP CLI.

2. To configure an existing private DNS zone for AKS, you must create an Azure environment using the CDP CLI. While creating the environment, specify the resource ID in `aksPrivateDnsZoneId` in the following CLI option:

```
--existing-network-params (object)
    Request object for creating an Azure environment using existing
    VNet
    and subnets.

    networkId -> (string)
        The id of the Azure VNet.

    resourceGroupName -> (string)
        The name of the resource group associated with the VNet.
    subnetIds -> (array)
        One or more subnet ids within the VNet.

    aksPrivateDnsZoneId -> (string)
        The full Azure resource ID of an existing Private DNS zone
used
    for the AKS.

    Shorthand Syntax:

        networkId=string,resourceGroupName=string,subnetIds=string,aksPrivateDnsZoneId=string

    JSON Syntax:

        {
            "networkId": "string",
            "resourceGroupName": "string",
            "subnetIds": ["string", ...],
            "aksPrivateDnsZoneId": "string"
        }
```



**Note:** You can specify the `aksPrivateDnsZoneId` parameter only through the CDP CLI.

## Results

You can use the AKS configured in the provided private DNS zone in a Cloudera Data Engineering service.

## Configuring an existing private DNS zone for a database

You can configure an existing Azure private DNS zone for your MySQL Flexi Server-based database to use in your Cloudera Data Engineering service. Alternatively, if you do not have an existing private DNS zone configured, the system can create a new private DNS zone for the database.

## About this task

If all the following statements are true, the system uses the private DNS zone that you specify in the `azure.database.privateDNSZoneId` parameter for the database:

- The `privateNetwork.enabled` parameter is set to true or false
- The `azure.database.privateDNSZoneId` parameter is set

If all the following statements are true, the system creates and enables a new private DNS zone for the database:

- The `privateNetwork.enabled` parameter is set to true or false
- The `azure.database.privateDNSZoneId` parameter is not set

For the database, you can configure a private DNS zone regardless of whether the `privateNetwork.enabled` parameter is set to true or false. This is because the MySQL Flexi server implementation requires a private DNS zone.

## Before you begin

For information on how to configure an existing private DNS zone for a database using the API, see [Configuring an existing private DNS zone for a database and a Storage Account File Share using the API](#).

To configure an existing private DNS zone for a database, you need the full resource ID of the private DNS zone. To obtain the full resource ID, use the JSON view option in [Azure Private DNS Zones](#). Example for the full resource ID of a private DNS zone for a database:

```
/subscriptions/[***SUBSCRIPTIONID**]/resourceGroups/[***RESOURCEGROUP**]/providers/Microsoft.Network/privateDnsZones/privatelink.mysql.database.azure.com
```

## Procedure

To configure an existing private DNS zone for a database, specify the full resource ID of the private DNS zone.

To configure the Azure private DNS zone ID for the database, use the enable-cluster CDP command with the `--azure-database-private-dns-zone-id` flag.

```
cdp de enable-service
--name "[***SERVICE NAME**]"
--env "[***ENVIRONMENT NAME**]"
--instance-type Standard_D8s_v4
--minimum-instances 0
--maximum-instances 50
--enable-private-network
--azure-database-private-dns-zone-id "[***AZURE DATABASE PRIVATE DNS ZONE RESOURCE ID**]"
--azure-fileshare-private-dns-zone-id "[***AZURE DATABASE FILE SHARE DNS ZONE RESOURCE ID**]"
```

## Configuring an existing private DNS zone for a Storage Account File Share

You can configure an existing Azure private DNS zone for your Storage Account File Share to use in your Cloudera Data Engineering service. Alternatively, if you do not have an existing private DNS zone configured, the system can create a new private DNS zone for the Storage Account File Share.

## About this task

If all the following statements are true, the system uses the private DNS zone that you specify in the `azure.fileshare.privateDNSZoneId` parameter for the private Storage Account:

- The private network is enabled (the `privateNetwork.enabled` parameter is set to true)
- The `DE_AZURE_PRIVATE_STORAGE` parameter is set to true on Cloudera-level
- The `azure.fileshare.privateDNSZoneId` parameter is set
- The custom Storage Account (also known as Bring Your Own Storage Account) is not used

If all the following statements are true, the system creates and enables a new private DNS zone for the private Storage Account:

- The private network is enabled (the `privateNetwork.enabled` parameter is set to true)
- The `DE_AZURE_PRIVATE_STORAGE` parameter is set to true on Cloudera-level
- The `azure.fileshare.privateDNSZoneId` parameter is not set
- The custom Storage Account (also known as Bring Your Own Storage Account) is not used



**Note:** You cannot use the private DNS zone for File Share together with a custom Storage Account that is also known as Bring Your Own (BYO) Storage Account. This means that you cannot use the Override Azure File Storage Server UI option with the private DNS zone for Storage Account File Share. The reason for this is that the BYO private custom Storage Account integrates with a private DNS zone from Azure, but Cloudera Data Engineering cannot accept an additional private DNS zone to manage the Storage Account File Share.

### Before you begin

For information on how to configure an existing private DNS zone for a Storage Account File Share using the API, see [Configuring an existing private DNS zone for a database and a Storage Account File Share using the API](#).

To configure an existing private DNS zone for a Storage Account File Share, make sure that the following prerequisites are fulfilled:

- The `DE_AZURE_PRIVATE_STORAGE` parameter is set to true on the Cloudera-side.
- The `privateNetwork.enabled` parameter is set to true.

To configure an existing private DNS zone for a Storage Account File Share, you need the full resource ID of the private DNS zone. To obtain the full resource ID, use the JSON view option in [Azure Private DNS Zones](#). Example for the full resource ID of a private DNS zone for Storage Account File Share:

```
/subscriptions/[***SUBSCRIPTIONID**]/resourceGroups/[***RESOURCEGROUP**]/providers/Microsoft.Network/privateDnsZones/privatelink.file.core.windows.net
```

### Procedure

To configure an existing private DNS zone for a Storage Account File Share, specify the full resource ID of the private DNS zone.

To configure the Azure private DNS zone ID for the Storage Account File Share, use the `enable-cluster CDP` command with the `--azure-fileshare-private-dns-zone-id` flag.

```
cdp de enable-service
--name "[***SERVICE NAME**]"
--env "[***ENVIRONMENT NAME**]"
--instance-type Standard_D8s_v4
--minimum-instances 0
--maximum-instances 50
--enable-private-network
--azure-database-private-dns-zone-id "[***AZURE DATABASE PRIVATE DNS ZONE RESOURCE ID**]"
--azure-fileshare-private-dns-zone-id "[***AZURE DATABASE FILE SHARE DNS ZONE RESOURCE ID**]"
```

### Enabling cross-subscription communication for private DNS zones

If the private DNS zones for the Storage Account File Share and the database are deployed in an Azure subscription that is outside of the Cloudera subscription, to enable the cross-subscription communication, you must configure the role assignment for Cloudera.

### About this task

In the subscription of the private DNS zone for the Storage Account Files Share and for the database, you must assign a contributor role to the Azure application that is used for Cloudera.

### Procedure

1. In Azure, navigate to the private DNS zone subscription of the database and the Storage Account File Share.

2. Using the Add role assignment option, assign a contributor role to the Azure application that is used for Cloudera. For more information, see the Microsoft Azure documentation on assigning Azure roles using the Azure portal.

## Enabling a semi-private network for a Cloudera Data Engineering service with AWS

Learn how to enable a semi-private network setup for a Cloudera Data Engineering service with Amazon Web Services (AWS) services in Cloudera. When you enable a Cloudera Data Engineering service with this feature, the Amazon Kubernetes Service (EKS) cluster is deployed as a private cluster but some services used by Cloudera Data Engineering such as MySQL and S3 are not provisioned as private.



**Note:** You need to contact Cloudera to have this feature enabled.

### For Cloudera Data Engineering UI

Before you begin

- Ensure that you have created and enabled a Cloudera Data Engineering service. Additionally, Cloudera must communicate with the Cloudera Data Engineering service on a private network in order to manage the Cloudera Data Engineering service lifecycle. This communication occurs using the Cluster Connectivity Manager v2; therefore, to enable this feature, the Cloudera environment must be enabled with the CCMv2. Once the CCMv2 is enabled at the Cloudera environment level, the Enable Private Network option displays in the Cloudera Data Engineering user interface when you enable a service. For more information on how to enable a Cloudera Data Engineering service and set up CCMv2, refer to the links below.

UI steps for enabling a private network

1. While enabling a Cloudera Data Engineering service for an AWS environment, under Network & Storage, select Enable Private Network.
2. Click Enable.

### For CDP CLI

You can enable a private network with the CDP CLI with the following commands:

```
cdp de enable-service --name dsp-private-eks-ntp-try1 --env dsp-aws-ntp-priv --instance-type m5.2xlarge --minimum-instances 0 --maximum-instances 4 --initial-instances 0 --root-volume-size 50 --no-skip-validation --enable-private-network
```

### Related Information

[Enabling a Cloudera Data Engineering service](#)

[Cluster Connectivity Manager](#)

[Enabling Cluster Connectivity Manager in the Management Console](#)

## Managing a Cloudera Data Engineering Service

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



**Note:** The active node count has been removed from the following UI pages:

- Administrator Overview
- Administrator Overview Service Details

For real-time node count information, check your Cloud Service Provider's website. For more information, see [Checking the node count on your Cloud Service Provider's website](#).

1. In the Cloudera console, click the Data Engineering tile. The Cloudera Data Engineering 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 Cloudera Data Engineering service you want to manage.
4. On the Configuration tab, you can view details about the service, including the service name and Cloudera 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, Cloudera environment, and the CPU and memory capacity. You can modify the CPU and memory capacity dynamically.



**Important:** For both Microsoft Azure and Amazon Web Services, to avoid being charged for All purpose compute nodes, even if you do not create an All purpose Virtual Cluster:

At Capacity & Costs > Autoscale Range, set the minimum value of All purpose On-demand Instances to 0.

For more information, see [Cloudera Data Engineering cluster types](#).

### 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 Cloudera Data Engineering service.

### Access

The Access tab displays the option to add Amazon Web Services (AWS) user's Amazon Resource Names (ARNs) for the Cloudera Data Engineering Service. The Access tab also displays the users you have added to the Cloudera Data Engineering Service which will provide them with the following:

- Access to all pods
- Access to all secrets including TGTs
- Bypass Istio security
- Access to tgtgen host keytab with ability to create, delete, and modify users in FreeIPA

### Diagnostics

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

## Renewing the Cloudera Data Engineering certificate

Learn about how to renew the Cloudera Data Engineering certificate when the certificate is about to expire.

### Procedure

1. Open the Cloudera Data Engineering UI and edit the URI to open the Swagger UI.  
Example: Edit the Cloudera Data Engineering landing page console URL `https://console.us-west-1.cdp.cloudera.com/dex/home` to the following URL to open the Swagger UI: `https://console.us-west-1.cdp.cloudera.com/dex/swagger/index.html#/Cluster/patch-cluster`
2. In the Swagger UI, provide the Cluster-ID for the affected Cloudera Data Engineering Service, and for the Data section, pass only the following tag and remove additional fields.  

```
{ "status_update": "renewTLSCerts" }
```

After running the command, the Cloudera Data Engineering Service Log shows that the TLS CertRenewal is triggered and completed with `NewExpiryDate`, and `NextRenewalAttempt` is set to 15 days before the expiry.
3. In the `Dex-Base-xxxxxx` namespace, restart the `dex-base-NGinx-Controller-xxxxx` pod.

## Suspending and resuming Cloudera Data Engineering Services - Technical preview

The suspend and resume feature allows you to temporarily pause your entire Cloudera Data Engineering Service when it is not needed.

### Benefits of suspending your services

Suspending your Services when they are not needed offers significant cost savings, primarily on compute resources, while ensuring your Services remain available when required. During suspension, all Service and Virtual Cluster configurations, as well as jobs, are preserved, similar to hibernating a laptop during idle periods.

For instance, to reduce operational costs, you can suspend your Services during weekends, or overnight on weekdays.

#### In AWS

Suspending the Service scales down all compute nodes used by the Cloudera Data Engineering Service and its Virtual Clusters (VCs) to zero.

#### In Azure

The feature stops the AKS cluster during suspension and restarts it when the Service is resumed.

### Prerequisites

To suspend and resume your Cloudera Data Engineering Services, you need to:

- Request the `DE_ENABLE_SUSPEND_RESUME` entitlement
- Have DE Admin user privileges at the environment level

### Limitations

The suspend and resume feature has these specific limitations:

- This feature is not available on the Cloudera Data Engineering UI. It is available only through the API and CLI.
- Only scales down compute nodes, application nodes, and base nodes. It does not stop or shut down the Relational Database Service (RDS) or managed database instances.
- Does not allow selective scaling down or suspending specific VCs.
- When you suspend the service, the following operations are not supported:
  - Generating diagnostic bundles; however, you can still download old diagnostic bundles.
  - Generating Grafana charts.



**Important:** Grafana is empty during the suspension. When you resume a Cloudera Data Engineering Service after being suspended, you lose the old Grafana charts history that was generated before the suspension. For example, if you suspend a Service at 12 a.m. midnight on 10 March 2025, and resume it at 6 a.m. on 14 March 2025, you lose all the Grafana history generated before 12 a.m. midnight on 10th March.

### Troubleshooting

If the suspend or resume operation fails, try once again before raising a support ticket.

For more information, see [Troubleshooting failures in suspend and resume](#).

### Related Information

[Suspending and resuming Cloudera Data Engineering Services using the API](#)

[Suspending and resuming Cloudera Data Engineering Services using the CLI](#)

[Suspending and resuming a Cloudera Data Engineering Service using a CLI script](#)

[Troubleshooting failures in suspend and resume](#)

## Cloudera Data Engineering tier types

Learn more about the available tier types that you can use in your Cloudera Data Engineering Service.

### Overview

On the Cloudera Data Engineering UI, DE and Platform administrators can create new Virtual Clusters and select the required cluster type at Administration Create a Virtual Cluster .

Cloudera Data Engineering offers these cluster types:

- Core
- All Purpose

In the same Cloudera Data Engineering Service, you can run some workloads on Core and some workloads on All purpose clusters.

### Tier functionalities

On the Core Tier, you can develop jobs for operational deployment and monitor batch based transformations.

Core Tier functionalities:

- Autoscaling cluster
- Spot instances
- Cloudera Shared Data Experience
- Open Lakehouse with Iceberg
- Job lifecycle management
- Centralized monitoring
- Workflow orchestration (Airflow)

On the All Purpose Tier, you can develop jobs using interactive sessions and deploy both batch and streaming (preview) workloads.

All Purpose Tier functionalities:



**Note:** All Purpose supports all functionalities of Core Tier.

- Interactive sessions
- External IDE connectivity



- Spark streaming
- JDBC connector (Coming soon)

The following screenshot from the Cloudera Data Engineering UI provides a detailed feature comparison for the Core and All Purpose Tiers.

## Cloudera Data Engineering Tier Types



Cloudera Data Engineering offers two different types of clusters that are each suited to your workloads. Explore these below:

Cluster Type	Core	All-Purpose
<b>Infrastructure</b>		
Autoscaling Cluster	✓	✓
Spot Instance	✓	✓
SDX/Lakehouse	✓	✓
<b>Spark</b>		
Job Lifecycle	✓	✓
Monitoring	✓	✓
Workflow Orchestration	✓	✓
Streaming (Preview)		✓
<b>Development Endpoints</b>		
Shell Session - CLI and Web		✓
JDBC/SparkSQL (Coming soon)		✓
IDE		✓

### Autoscaling range considerations for tier and instance types

In the Cloudera Data Engineering UI, in the Administration Enable a Service Capacity & Costs Autoscale Range section, you can set the autoscale range for the Core Tier and for the All Purpose Tier.

In the **Capacity & Costs** section, you can configure the autoscaling range for On-demand instances and if you select an AWS environment, for Spot instances as well. Currently, Spot instances in Azure are not supported in Cloudera Data Engineering.

The key difference between On-demand instances and Spot instances is that On-demand instances are consistently available, while the availability of Spot instances depends on the cloud provider.


You can use Spot instances and All Purpose Spot instances to save compute costs. For more information about Spot instances, see [Cloudera Data Engineering Spot Instances](#) and the AWS document, [Amazon EC2 Spot Features](#).



**Important:** Before configuring the minimum value for the autoscaling ranges, keep in mind that you are charged at least for the minimum number of instances that you set, regardless of the usage. This means that if you set the minimum value to 1, even if you do not use the instance, you are charged for that 1 instance.

## Capacity & Costs

Configure the capacity based on your estimated consumption and set the autoscaling ranges for the Core and All-Purpose Tiers, which differ in functionality and pricing. [Learn more](#)

 You will be charged at least for the minimum number of instances set, regardless of the usage.  
[Learn More](#)

### Autoscaling Range

Set the autoscaling range for the Core and All-Purpose Tiers. For more information about the tiers, review the [feature comparison](#).

#### Core Tier

On-demand instances

Minimum	<input type="text" value="0"/>	Maximum	<input type="text" value="50"/>
---------	--------------------------------	---------	---------------------------------

Spot instances

Minimum	<input type="text" value="0"/>	Maximum	<input type="text" value="25"/>
---------	--------------------------------	---------	---------------------------------

Enter a minimum and maximum value between 0 and 100

#### All Purpose Tier

On-demand instances

Minimum	<input type="text" value="0"/>	Maximum	<input type="text" value="50"/>
---------	--------------------------------	---------	---------------------------------

Spot instances

Minimum	<input type="text" value="0"/>	Maximum	<input type="text" value="25"/>
---------	--------------------------------	---------	---------------------------------

Enter a minimum and maximum value between 0 and 100

## Important factors to consider before using All purpose On-demand instances

With All Purpose On-demand Instances, consider the following:

- Since version 1.20.3, Cloudera Data Engineering offers price tiering to provide better cost-control for the consumed features.



**Important:** The All Purpose Tier On-demand Instances are charged at a higher rate than the Core Tier On-demand Instances. For more information, see the [pricing page](#).

- To avoid being charged for All Purpose compute nodes, set the minimum value of All Purpose On-demand Instances to 0. However, in such a case, consider that if you create a new session, the session can take several minutes to start.

## Removing a Cloudera Data Engineering service

Disabling an existing Cloudera Data Engineering 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 Cloudera Data Engineering does not delete your Cloudera data. If enabling the Cloudera Data Engineering 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



**Important:** The user interface for Cloudera Data Engineering 1.17 and above has been updated. The left-hand menu was updated to provide easy access to commonly used pages. The steps below will vary slightly, for example, the Overview page has been replaced with the Home page. You can remove a Cloudera Data Engineering service by clicking Administration on the left-hand menu, then proceed to step 2 listed below. The new home page still displays Virtual Clusters, but now includes quick-access links located at the top for the following categories: Jobs, Resources, and Download & Docs.

### Procedure

1. In the Cloudera console, click the Data Engineering tile and click Overview.
2. In the Cloudera Data Engineering Services column, click the menu icon for the environment for which you want to disable the Cloudera Data Engineering service, and then click Disable Cloudera Data Engineering.



**Warning:** Disabling an existing Cloudera Data Engineering 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. Additionally, if you're prompted to perform a Force Disable, in the event that a Disable is not successful, you must perform a manual cleanup of cloud infrastructures such as Security Group, EBS Volume, and S3 Bucket. A manual cleanup is not required for a standard Disable.

3. Confirm that you want to disable Cloudera Data Engineering by typing the environment name and then clicking Disable.

### Results

The Cloudera Data Engineering Administration page displays the status of the environment that is being disabled.

### What to do next

If you disabled Cloudera Data Engineering as a result of a failure to enable Cloudera Data Engineering for the first time on an environment, resolve any reported issues, and then try again.

## Limiting Incoming Endpoint Traffic for Cloudera Data Engineering Services For AWS

You can limit incoming endpoint traffic for a Cloudera Data Engineering service.

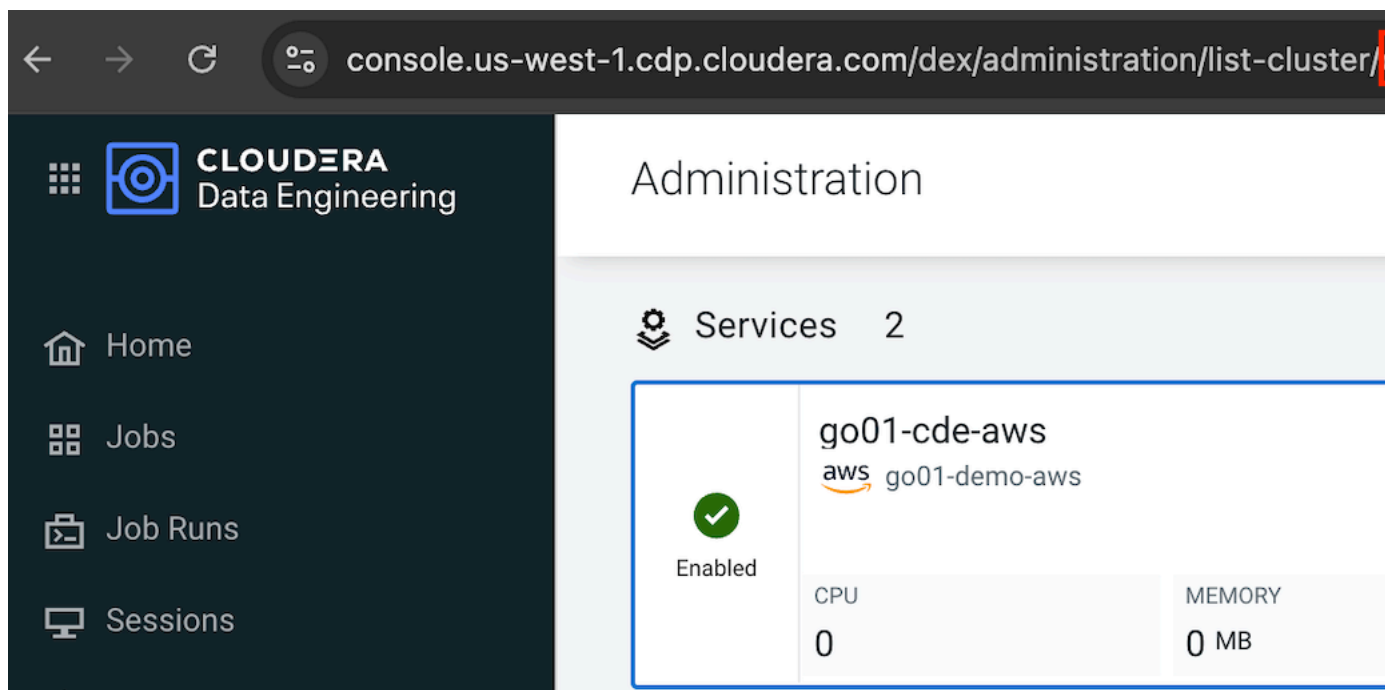
### Before you begin



**Important:** The user interface for Cloudera Data Engineering 1.17 and above has been updated. The left-hand menu was updated to provide easy access to commonly used pages. The steps below will vary slightly, for example, the Overview page has been replaced with the Home page. The new home page still displays Virtual Clusters, but now includes quick-access links located at the top for the following categories: Jobs, Resources, and Download & Docs.

**Procedure**

1. Note the Cloudera Data Engineering service ID, which you can obtain from the URL in the Cloudera Data Engineering Management Console when the service is highlighted:



2. Go to the AWS console for the account where the Cloudera Data Engineering service is enabled.
3. Navigate to EC2 -> Load Balancers in the AWS console and enter the following filter:

```
tag:cde-cluster-id : <id_from_step_1>
, e.g. tag:cde-cluster-id : cluster-cn92cs9g
```

4. Select the Load Balancer instance and then under the Description tab in the Detail window, click on the link to the Source Security Group.
5. In the subsequent view, select the correct security group ID.
6. In the subsequent window, click Edit inbound rules.
7. Modify the "0.0.0.0/0" CIDR ranges for the HTTPS rule to your desired CIDR ranges.

Add additional ranges and rules as required but note that HTTPS traffic must be enabled for each range. The HTTP (port 80) and ICMP rules can be removed.