

## GCP Environments

Date published: 2019-08-22

Date modified:



# 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

|  |           |
|--|-----------|
| <b>Working with GCP environments.....</b>                    | <b>4</b>  |
| <b>Introduction to Google Cloud environments.....</b>        | <b>4</b>  |
| <b>Register environment (UI).....</b>                        | <b>5</b>  |
| <b>Register environment (CLI).....</b>                       | <b>8</b>  |
| <b>Enabling admin and user access.....</b>                   | <b>12</b> |
| <b>Understanding environment UI options.....</b>             | <b>12</b> |
| <b>Monitoring an environment.....</b>                        | <b>13</b> |
| Environment status options.....                              | 14        |
| <b>Stop and restart an environment.....</b>                  | <b>16</b> |
| <b>Delete an environment.....</b>                            | <b>17</b> |
| <b>Change environment's credential.....</b>                  | <b>18</b> |
| <b>Defining custom tags.....</b>                             | <b>18</b> |
| <b>Adding a customer managed encryption key for GCP.....</b> | <b>20</b> |

# Working with GCP environments

Refer to the following documentation to learn about creating and managing GCP environments in CDP:

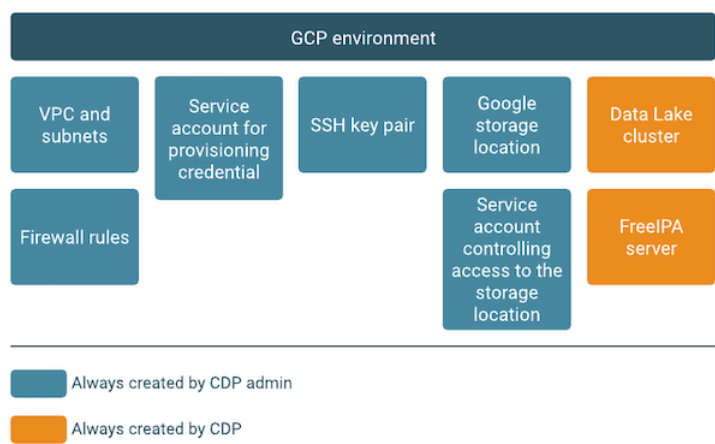
- Related Information
- [Managing provisioning credentials for GCP](#)
- [Managing Data Lakes](#)
- [Managing FreeIPA](#)

## Introduction to Google Cloud environments

In CDP, an environment is a logical subset of your cloud provider account including a specific virtual private network. You can register as many environments as you require.


Registering your GCP environment in CDP provides CDP with limited access to your GCP account and identifies a set of resources in your GCP account that CDP services can access. Once you’ve registered your GCP environment in CDP, you can start provisioning CDP resources such as clusters, which run on the physical infrastructure in an GCP data center.

The following diagram enumerates the components of a GCP environment:



The diagram illustrates all major user-created and CDP-created components of an environment:

- The items in dark blue boxes must be pre-created by your CDP administrator prior to environment registration and then specified when registering an environment.
- The items in orange boxes are automatically provisioned on GCP by CDP as part of environment provisioning.

 **Note:** The items that are user-created don’t get terminated during environment deletion.

As shown in the diagram, an environment consists of the following resources:

| Environment component        | Description  |
|------------------------------|--|
| Virtual network with subnets | An environment corresponds to one specific VPC network and one or more subnets in which CDP resources are provisioned. |

| Environment component  | Description   |
|--|---|
| Firewall rules   | <p>Firewall rules (similar to security groups on other cloud providers) act as a virtual firewall for your instances to control inbound and outbound traffic.</p> <p>All VM instances provisioned within an environment use your specified security access settings allowing inbound access to your instances from your organization's computers.</p>             |
| Service account for provisioning credential                            | <p>CDP uses a provisioning credential for authorization to provision resources (such as compute instances) within your cloud provider account.</p> <p>In GCP, credential creation involves creating a service account, assigning a set of minimum IAM permissions to it, and providing CDP with the access key generated for the service account.</p>             |
| SSH public key   | <p>When registering an environment on a public cloud, a CDP administrator provides an SSH public key. This way, the administrator has root-level access to the Data Lake instance and Data Hub cluster instances.</p>   |
| Google storage location and a service account controlling access to it | <p>When registering an environment, you must provide a Google storage location for storing:</p> <ul style="list-style-type: none"> <li>• All workload cluster data</li> <li>• Cluster service logs and Ranger audits</li> </ul> <p>Furthermore, you must create and assign a service account on the scope of this storage location so that CDP can access it.</p> |
| Data Lake cluster  | <p>A data lake is automatically provisioned when an environment is created. It provides a mechanism for storing, accessing, organizing, securing, and managing data.</p>  |
| FreeIPA server   | <p>A FreeIPA server is automatically provisioned when an environment is created. It is responsible for synchronizing your users and making them available to CDP services, Kerberos service principal management, and more.</p>   |

Once your environment is running, you can provision Data Hub clusters in it.

## Register a GCP environment from CDP UI

Once you've met the Google Cloud cloud provider requirements, register your GCP environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in [GCP requirements](#).

Required role: EnvironmentCreator

Steps

1. Navigate to the Management Console > Environments > Register environment.
2. On the On the Register Environment page, provide the following information:page, provide the following information:

| Parameter                   | Description   |
|-----------------------------|---|
| General Information         |   |
| Environment Name (Required) | <p>Enter a name for your environment. The name:</p> <ul style="list-style-type: none"> <li>• Must be between 5 and 28 characters long.</li> <li>• Can only include lowercase letters, numbers, and hyphens.</li> <li>• Must start with a lowercase letter.</li> </ul> |

| Parameter                                   | Description  |
|---|--|
| Description                                 | Enter a description for your environment.  |
| Select Cloud Provider (Required)            | Select Google Cloud.   |
| Google Cloud Platform Credential (Required) |  |
| Select Credential                           | Select an existing credential or select Create new credential.<br>For instructions on how to create a credential for Google Cloud, refer to <a href="#">Create a provisioning credential for GCP</a> . |

3. Click Next.

4. On the Data Access and Data Lake Scaling page, provide the following information:

| Parameter                               | Description   |
|---|---|
| Data Lake Settings                      |   |
| Data Lake Name (Required)               | Enter a name for the Data Lake cluster that will be created for this environment. The name: <ul style="list-style-type: none"> <li>Must be between 5 and 100 characters long</li> <li>Must contain lowercase letters</li> <li>Cannot contain uppercase letters</li> <li>Must start with a letter</li> <li>Can only include the following accepted characters are: a-z, 0-9, -, .</li> </ul> |
| Data Lake Version (Required)            | Select Cloudera Runtime version that should be deployed for your Data Lake. The latest stable version is used by default.<br><br>All Data Hub clusters provisioned within this Data Lake will be using the same Runtime version.<br><br>Note: Google Cloud environments can only be provisioned in CDP with Runtime version 7.2.8 or newer.   |
| Data Access and Audit                   |   |
| Assumer Service Account (Required)      | Select the IDBroker service account created in <a href="#">Minimum setup for cloud storage</a> .  |
| Storage Location Base (Required)        | Select the Google Storage location created for data in <a href="#">Minimum setup for cloud storage</a> .  |
| Data Access Service Account (Required)  | Select the Data Lake Admin service account created in <a href="#">Minimum setup for cloud storage</a> .   |
| Ranger Audit Service Account (Required) | Select the Ranger Audit service account created in <a href="#">Minimum setup for cloud storage</a> .  |
| IDBroker Mappings                       | We recommend that you leave this out and set it up after registering your environment as part of <a href="#">Onboarding CDP users and groups for cloud storage</a> .  |
| Scale (Required)                        | Select Data Lake scale. By default, "Light Duty" is used. For more information on Data Lake scale, refer to <a href="#">Data Lake scale</a> .   |

5. Click on Advanced Options to make additional configurations for your Data Lake. The following options are available:

| Parameter            | Description   |
|----------------------|---|
| Hardware and Storage | For each host group you can specify an instance type. For more information on instance types, see <a href="#">Machine type families</a> .                         |
| Cluster Extensions   |   |
| Recipes              | You can optionally select and attach previously registered recipes to run on a specific Data Lake host group. For more information, see <a href="#">Recipes</a> . |

6. Click Next.

7. On the Region, Networking and Security page, provide the following information:

| Parameter                              | Description  |
|--|--|
| <b>Region</b>                          |  |
| Select Region (Required)               | Select the region where your VPC network is located.   |
| Select Zone (Required)                 | Select a zone within the selected region.  |
| <b>Network</b>                         |  |
| Use shared VPC                         | <p>This option is disabled by default. Enable this if you would like to use your existing shared VPC. Next enter:</p> <ul style="list-style-type: none"> <li>Host project ID</li> <li>Network name</li> <li>Subnet name(s). If providing multiple, provide a comma separated list.</li> </ul>  |
| Select Network (Required)              | Select the existing VPC network that you created as a prerequisite in the <a href="#">VPC network and subnets</a> step. All CDP resources will be provisioned into this network.   |
| Select Subnets (Required)              | Select at least one existing subnet.   |
| Create Public IPs                      | This option is disabled by default when CCM is enabled and enabled by default when CCM is disabled.  |
| Proxies                                | Select a proxy configuration if previously registered. For more information refer to <a href="#">Setting up a proxy server</a> .   |
| <b>Security Access Settings</b>        |  |
| Select Security Access Type (Required) | <p>You have two options:</p> <ul style="list-style-type: none"> <li>Do not create firewall rule: If you are using a shared VPC you can set the firewall rules directly on the VPC. If you did so, you can select this option.</li> <li>Provide existing firewall rules: If not all of your firewall rules are set directly on the VPC, provide the previously created firewall rules for SSH an UI access. You should select two existing firewall rules, one for Knox gateway-installed nodes and another for all other nodes. You may select the same firewall rule in both places if needed.</li> </ul> <p>For information on required ports, see <a href="#">Firewall rules</a>.</p> |
| <b>SSH Settings</b>                    |  |
| New SSH public key (Required)          | <p>Upload a public key directly from your computer.</p> <p>Note: CDP does not use this SSH key. The matching private key can be used by your CDP administrator for root-level access to the instances provisioned for the Data Lake and Data Hub.</p>  |
| Add tags                               | You can optionally add tags to be created for your resources on GCP. Refer to <a href="#">Defining custom tags</a> .   |

8. Click on Advanced Options to make additional configurations for FreeIPA. The following options are available:

| Parameter                 | Description   |
|---------------------------|---|
| Hardware and Storage      | For each host group you can specify an instance type. For more information on instance types, see <a href="#">Machine type families</a> .       |
| <b>Cluster Extensions</b> |   |
| Recipes                   | You can optionally select and attach previously registered recipes to run on FreeIPA nodes. For more information, see <a href="#">Recipes</a> . |

9. Click Next.

10. On the Storage page, provide the following information:

| Parameter                                 | Description   |
|---|---|
| Logs                                      |   |
| Logger Service Account (Required)         | Select the Logger service account created in <a href="#">Minimum setup for cloud storage</a> .  |
| Logs Location Base (Required)             | Select the Google Storage location created for logs in <a href="#">Minimum setup for cloud storage</a> .  |
| Backup Location Base                      | Select the Google Storage location created for FreeIPA backups in <a href="#">Minimum setup for cloud storage</a> . If not provided, the default Backup Location Base uses the Logs Location Base.                      |
| Telemetry                                 |   |
| Enable Workload Analytics                 | Enables Cloudera Observability support for workload clusters created within this environment. When this setting is enabled, diagnostic information about job and query execution is sent to the Cloudera Observability. |
| Enable Deployment Cluster Logs Collection | When this option is enabled, the logs generated during deployments will be automatically sent to Cloudera.  |

11. Click Register Environment to trigger environment registration.

12. The environment creation takes about 60 minutes. The creation of the FreeIPA server and Data Lake cluster is triggered. You can monitor the progress from the web UI. Once the environment creation has been completed, its status will change to “Running”.

After you finish

After your environment is running, perform the following steps:

- You must assign roles to specific users and groups for the environment so that selected users or user groups can access the environment. Next, you need to perform user sync. For steps, refer to [Enabling admin and user access to environments](#).
- You must onboard your users and/or groups for cloud storage. For steps, refer to [Onboarding CDP users and groups for cloud storage](#).
- You must create Ranger policies for your users. For instructions on how to access your Data Lake, refer to [Accessing Data Lake services](#). Once you've accessed Ranger, [create Ranger policies](#) to determine which users have access to which databases and tables.

## Register a GCP environment from CDP CLI

Once you’ve met the Google Cloud cloud provider requirements, register your GCP environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in [GCP requirements](#).

Required role: EnvironmentCreator

Steps

Unlike in the CDP web interface, in CDP CLI environment creation is a two-step process with environment creation and data lake creation being two separate steps. The following commands can be used to create an environment in CDP.

1. Once you’ve met the prerequisites, register your GCP environment in CDP using the `cdp environments create-gcp-environment` command and providing the CLI input parameters. For example:

```
cdp environments create-gcp-environment --cli-input-json '{
  "environmentName": "test-env",
  "description": "Test GCP environment",

```



```

    "credentialName": "test-gcp-crd",
    "region": "us-west2",
    "publicKey": "ssh-rsa AAAAB3NzaZlYc2EAAAADAQABAAQDwCI/wmQzbNn9YcA8v
dU+Ot41IIUWJfOfiDrUuNcUL0QL6ke5qcEKuboXzbLxV0YmQcPFvswbM5S4FlHjy2VrJ5spy
GhQajFEm9+PgrsybgzHkkssziX0zRq7U4BVD68kSn6CuAHj9L4wx8WBwefMzkw7u0lCkfifI
p8UE6ZcKKKwe2fLR6ErDaN9jQxIWhTPEiFjIhItPHrnOcfGKY/p6OlpDDUOuMRiFZh7qMzfg
vWI+UdN/qjnTlc/M53JftK6GJqK6osN+j7fCwKENPwWC/gmy8El7ZMHlIENxDut6X0qj9Okc/
JMMG0ebkSZAebhgNOBNLZYdP0oeQGCXjqdv",
    "enableTunnel": true,
    "usePublicIp": true,
    "existingNetworkParams": {
      "networkName": "eng-private",
      "subnetNames": [
        "private-us-west2"
      ],
      "sharedProjectId": "dev-project"
    },
    "logStorage": {
      "storageLocationBase": "gs://logs",
      "serviceAccountEmail": "logger@dev-project.iam.gserviceaccount.com"
    }
  },
  "
}

```

| Parameter             | Description  |
|-----------------------|--|
| environmentName       | Provide a name for your environment.   |
| credentialName        | Provide the name of the credential created earlier.  |
| region                | Specify the region where your existing VPC network is located. For example "us-west2" is a valid region.   |
| publicKey             | Paste your SSH public key.   |
| existingNetworkParams | <p>Provide a JSON specifying the following:</p> <pre> {   "networkName": "string",   "subnetNames": ["string", ...],   "sharedProjectId": "string" } </pre> <p>Replace the values with the actual VPC network name, one or more subnet names and shared project ID.</p> <p>The sharedProjectId value needs to be set in the following way:</p> <ul style="list-style-type: none"> <li>For a shared VPC, set it to the GCP host project ID</li> <li>For a non-shared VPC, set it to the GCP project ID of the project where CDP is being deployed.</li> </ul> |
| enableTunnel          | <p>By default CCM is enabled (set to "true"). If you would like to disable it, set it to "false". If you disable it, then you must also add the following to your JSON definition to specify two security groups as follows:</p> <pre> "securityAccess": {   "securityGroupIdForKnox": "string",   "defaultSecurityGroupId": "string" } </pre>   |

| Parameter   | Description  |
|-------------|--|
| usePublicIp | Set this to “true” or “false”, depending on whether or not you want to create public IPs.  |
| logStorage  | <p>Provide a JSON specifying your configuration for cluster and audit logs:</p> <pre>{   "storageLocationBase": "string",   "serviceAccountEmail": "string" }</pre> <p>The storageLocationBase should be in the following format: gs://my-bucket-name.</p> |



**Note:** CDP CLI includes the `cdp environments create-gcp-environment --generate-cli-skeleton` command option, which allows you to generate a CLI JSON template. You can also use CLI help to get some information about specific CLI command options.

- To verify that your environment is running, use:

```
cdp environments list-environments
```

You can also log in to the CDP web interface to check the deployment status.

- Once your environment and Data Lake are running, you should set IDBroker Mappings. To create the mappings, run the `cdp environments set-id-broker-mappings` command. For example:

```
cdp environments set-id-broker-mappings \
--environment-name test-env \
--data-access-role dl-admin@dev-project.iam.gserviceaccount.com \
--ranger-audit-role ranger-audit@dev-project.iam.gserviceaccount.com \
--mappings '[{"accessorCrn": "crn:altus:iam:us-west-1:45ca3068-42a6-4227-8394-13a4493e2ac0:user:430c534d-8a19-4d9e-963d-8af377d16963", "role": "data-science@dev-project.iam.gserviceaccount.com"}, {"accessorCrn": "crn:altus:iam:us-west-1:45ca3068-42a6-4227-8394-13a4493e2ac0:machineUser:mfox-gcp-idbmms-test-mu/2cbca867-647b-44b9-8e41-47a01dea6c19", "role": "data-eng@dev-project.iam.gserviceaccount.com"}]'
```

| Parameter         | Description   |
|-------------------|---|
| environment-name  | Specify a name of the environment created earlier.  |
| data-access-role  | Specify an email address of the Data Lake admin service account created earlier.  |
| ranger-audit-role | Specify an email address of the Ranger audit service account created earlier.   |
| mappings          | <p>Map CDP users or groups to GCP service accounts created earlier. Use the following syntax:</p> <pre>[ {   "accessorCrn": "string",   "role": "string" } ... ]</pre> <p>You can obtain user or group CRN from the Management Console &gt; User Management by navigating to details of a specific user or group.</p> <p>The role should be specified as service account email.</p> |

4. Next, sync IDBroker mappings:

```
cdp environments sync-id-broker-mappings --environment-name demo3
```

5. Finally, check the sync status:

```
cdp environments get-id-broker-mappings-sync-status --environment-name demo3
```

6. Once your environment is running, you can create a Data Lake using the `cdp datalake create-gcp-datalake` command and providing the CLI input parameters:

```
cdp datalake create-gcp-datalake --cli-input-json '{
  "datalakeName": "my-dl",
  "environmentName": "test-env",
  "scale": "LIGHT_DUTY",
  "cloudProviderConfiguration": {
    "serviceAccountEmail": "idbroker@dev-project.iam.gserviceaccount.com",
    "storageLocation": "gs://data-storage"
  }
}'
```

| Parameter                               | Description  |
|---|--|
| <code>datalakeName</code>               | Provide a name for your Data Lake.   |
| <code>environmentName</code>            | Provide a name of the environment created earlier.   |
| <code>scale</code>                      | Provide Data Lake scale. It must be one of: <ul style="list-style-type: none"> <li><code>LIGHT_DUTY</code> or</li> <li><code>MEDIUM_DUTY_HA</code>.</li> </ul> |
| <code>cloudProviderConfiguration</code> | Provide the name of the data storage bucket and the email of the IDBroker service account.   |



**Note:** CDP CLI includes the `cdp datalake create-gcp-datalake --generate-cli-skeleton` command option, which allows you to generate a CLI JSON template. You can also use CLI help to get some information about specific CLI command options.

7. To verify that your Data lake is running, use:

```
cdp datalake list-datalakes
```

You can also log in to the CDP web interface to check the deployment status.

After you finish

After your environment is running, perform the following steps:

- You must assign roles to specific users and groups for the environment so that selected users or user groups can access the environment. Next, you need to perform user sync. For steps, refer to [Enabling admin and user access to environments](#).
- You must onboard your users and/or groups for cloud storage. For steps, refer to [Onboarding CDP users and groups for cloud storage](#).
- You must create Ranger policies for your users. For instructions on how to access your Data Lake, refer to [Accessing Data Lake services](#). Once you've accessed Ranger, [create Ranger policies](#) to determine which users have access to which databases and tables.

## Enabling admin and user access to environments

In order to grant admin and user access to an environment that you registered in CDP, you should assign the required roles.

You need to be an EnvironmentCreator in order to register an environment. Once an environment is running, the following roles can be assigned:

- **EnvironmentAdmin** - Grants all rights to the environment and Data Hub clusters running in it, except the ability to delete the environment. The user who registers the environment automatically becomes its EnvironmentAdmin.
- **EnvironmentUser** - Grants permission to view Data Hub clusters and set the workload password for the environment. This role should be used in conjunction with service-specific roles such as DataHubAdmin, DWAdmin, DWUser, MLAdmin, MLUser, and so on. When assigning one of these service-specific roles to users, make sure to also assign the EnvironmentUser role.
- **DataSteward** - Grants permission to perform user/group management functions in Ranger and Atlas Admin, manage ID Broker mappings, and start user sync for the environment.
- **Owner** - Grants the ability to manage the environment in CDP, including deleting the environment. The user who registers the environment automatically becomes its Owner. The Owner role does not grant access to the environment's clusters (Data Lakes, Data Hubs).

The roles are described in detail in Resource roles. The steps for assigning the roles are described in Assigning resource roles to users and Assigning resource roles to groups.

### Related Information

[Resource roles](#)

[Assigning resource roles to users](#)

[Assigning resource roles to groups](#)

## Understanding environment UI options

To access information related to your environment, navigate to the Management Console service > Environments and click on your environment.

The screenshot shows the Cloudera Management Console interface for an environment named 'cpxmon-az'. The left sidebar contains navigation links for Dashboard, Environments, Data Lakes, User Management, Data Hub Clusters, Data Warehouses, ML Workspaces, Classic Clusters, Cost Management, and Global Settings. The main content area displays the environment's details, including its CRN, resource group, and a summary of its components. The 'Summary' tab is active, showing a table with columns for Data Lake Name, Nodes, Data Lake Scale, Data Lake Status, and Reason. The Data Lake Name is 'cpxmon-az', Nodes is '2', Data Lake Scale is 'Light Duty', Data Lake Status is 'Running', and Reason is 'Datalake is running'. Below this, the 'General' tab shows the CRN and Resource Group. The 'Event History' tab on the right shows a list of events, including 'Environment sync is finished and new status is found...', 'Environment failed to start', 'Synchronize users started for Environment', 'Datahub started for Environment', and 'Datalake started for Environment'.

You need to have the EnvironmentUser role or higher for the environment in order to access details of that environment.

From environments details, you can access the following:

- From the Data Hub tab, you can create, manage, and access Data Hub clusters within the environment.
- From the Data Lake tab, you can monitor, manage, and access the Data Lake cluster.
- From the Cluster Definitions tab, you can access all cluster definitions that can be used with the environment.
- From the Summary tab, you can manage and monitor your environment.

The Summary includes the following information:

| Option                 | Description   |
|------------------------|---|
| General                | This includes your environment's CRN. CRN is an ID that CDP uses to identify a resource.  |
| Credential             | This links the provisioning credential associated with the environment and includes the option to change the credential.  |
| Region                 | This lists the region in which your environment is deployed.  |
| Network                | This lists the networking resources used by your environment, provided by you or created by CDP during environment registration. You can add additional subnets for Data Hub clusters deployed in the future. |
| Security Access        | This lists the firewall rules used by your environment, provided by you or created by CDP during environment registration. You can provide new firewall rules for Data Hub clusters deployed in the future.   |
| FreeIPA                | This includes details of a FreeIPA server running in the environment and includes an Actions menu with FreeIPA management options.  |
| Log Storage and Audits | This lists the cloud storage location used for logs and audits that you provided during environment registration. There is no way to update this location once your environment is running.                   |
| Telemetry              | This includes your environment's telemetry settings. You can change them for any Data Hub clusters created in the future.   |
| Advanced               | This lists the name of your root SSH key.   |

### Related Information

[Understanding Data Hub details](#)

[Understanding Data Lake details](#)

## Monitoring an environment

Once an environment exists, you can access it from the Management Console.

Required role: EnvironmentUser, EnvironmentAdmin, or Owner

Steps

### For CDP UI

1. To access an existing environment, navigate to Management Console > Environments and click on your environment.
2. Click on the Summary tab to access environment details.
3. You can monitor the status of your environment from this page.

### For CDP CLI

You can also list available environments from CDP CLI using the `cdp environments list-environments` command. For example:

```
cdp environments list-environments
```

```
{
  "environments": [
    {
      "environmentName": "cdp-demo",
      "crn": "crn:altus:environments:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:environment:d3361b40-39ab-4d87-bd5b-abc15f16b90c",
      "status": "DELETE_FAILED",
      "region": "us-east-2",
      "cloudPlatform": "AWS",
      "credentialName": "cdp-demo",
      "description": "Cdp demo"
    },
    {
      "environmentName": "cdp-new",
      "crn": "crn:altus:environments:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:environment:1d2bacde-5c96-47c1-a597-9f276b824028",
      "status": "AVAILABLE",
      "region": "us-east-2",
      "cloudPlatform": "AWS",
      "credentialName": "cdp-demo",
      "description": ""
    }
  ]
}
```

To get more information about a specific environment, you can use the following commands:

```
cdp environments describe-environment --environment-name <value>
```

```
cdp environments get-id-broker-mappings --environment-name <value>
```

## Environment status options

This topic lists all possible environment status options for the UI and CLI and explains what they mean.

| Environment status                     | Description  |
|--|--|
| Environment creation                   |  |
| CREATION_INITIATED                     | Environment creation request was registered in the database and CDP is starting the environment creation flow.               |
| ENVIRONMENT_INITIALIZATION_IN_PROGRESS | Setting up the region and network metadata (public/private and cidr).  |
| ENVIRONMENT_VALIDATION_IN_PROGRESS     | Setting up the region and network metadata (public/private and cidr).  |
| NETWORK_CREATION_IN_PROGRESS           | If the user chose the create new network option, then CDP creates the network on cloud provider side.                        |
| PUBLICKEY_CREATE_IN_PROGRESS           | If the user choose the create new SSH key option, then CDP creates the SSH key on cloud provider side.                       |
| FREEIPA_CREATION_IN_PROGRESS           | Creating the FreeIPA resources for an environment.   |
| Environment update                     |  |
| UPDATE_INITIATED                       | Environment update was requested and CDP is starting the update flow (network update, load balancer update, SSH key update). |
| Environment deletion                   |  |
| DELETE_INITIATED                       | Environment deletion request was registered and CDP is starting the deletion flow.   |
| NETWORK_DELETE_IN_PROGRESS             | If the user chose the create new network option, then CDP deletes the network on cloud provider side.                        |

| Environment status                   | Description  |
|--------------------------------------|--|
| PUBLICKEY_DELETE_IN_PROGRESS         | If the user choosing the create new SSH key option, then CDP deletes the SSH key on cloud provider side. |
| FREEIPA_DELETE_IN_PROGRESS           | Deleting the FreeIPA resources for an environment.   |
| EXPERIENCE_DELETE_IN_PROGRESS        | Deleting all the attached clusters (CDW, CML, and so on).  |
| RDBMS_DELETE_IN_PROGRESS             | Deleting all the provisioned RDS instances that are related to an environment.                           |
| CLUSTER_DEFINITION_DELETE_PROGRESS   | Deleting all the cluster definitions that are created for an environment.                                |
| UMS_RESOURCE_DELETE_IN_PROGRESS      | Deleting all the related UMS resources for an environment.   |
| IDBROKER_MAPPINGS_DELETE_IN_PROGRESS | Deleting all the IBroker mapping for an environment.   |
| S3GUARD_TABLE_DELETE_IN_PROGRESS     | Deleting all the Dynamo DB tables for an environment.  |
| DATAHUB_CLUSTERS_DELETE_IN_PROGRESS  | Deleting all the attached Data Hub clusters.   |
| DATALAKE_CLUSTERS_DELETE_IN_PROGRESS | Deleting the attached Data Lake cluster.   |
| ARCHIVED                             | Environment has been deleted (not shown on the UI).  |
| Environment is running               |  |
| AVAILABLE                            | Environment is available (ready to use).   |
| Environment process failed           |  |
| CREATE_FAILED                        | Environment creation failed (Detailed message in the statusReason).                                      |
| DELETE_FAILED                        | Environment deletion failed (Detailed message in the statusReason).                                      |
| UPDATE_FAILED                        | Environment update failed (Detailed message in the statusReason).  |
| Environment stop                     |  |
| STOP_DATAHUB_STARTED                 | Stopping all the Data Hub clusters in an environment.  |
| STOP_DATAHUB_FAILED                  | Stopping all the Data Hub clusters in an environment failed (Detailed message in the statusReason).      |
| STOP_DATALAKE_STARTED                | Stopping the Data Lake cluster in an environment.  |
| STOP_DATALAKE_FAILED                 | Stopping the Data Lake cluster in an environment failed (Detailed message in the statusReason).          |
| STOP_FREEIPA_STARTED                 | Stopping the FreeIPA instances in an environment.  |
| STOP_FREEIPA_FAILED                  | Stopping the FreeIPA instances in an environment failed (Detailed message in the statusReason).          |
| ENV_STOPPED                          | Environment was successfully stopped.  |
| Environment start                    |  |
| START_DATAHUB_STARTED                | Starting all the Data Hub clusters in an environment.  |
| START_DATAHUB_FAILED                 | Starting all the Data Hub clusters in an environment failed (Detailed message in the statusReason).      |
| START_DATALAKE_STARTED               | Starting the Data Lake cluster in an environment.  |
| START_DATALAKE_FAILED                | Starting the Data Lake cluster in an environment failed (Detailed message in the statusReason).          |
| START_FREEIPA_STARTED                | Starting all the FreeIPA instances in an environment.  |
| START_FREEIPA_FAILED                 | Starting all the FreeIPA instances failed in an environment (Detailed message in the statusReason).      |
| START_SYNCHRONIZE_USERS_STARTED      | Starting user sync for all the clusters in an environment.   |
| START_SYNCHRONIZE_USERS_FAILED       | Starting user sync for all the clusters in an environment failed (Detailed message in the statusReason). |

| Environment status                 | Description  |
|------------------------------------|--|
| FreeIPA instance deletion          |  |
| FREEIPA_DELETED_ON_PROVIDER_SIDE   | The FreeIPA instance has been deleted on cloud provider side.  |
| Load balancer                      |  |
| LOAD_BALANCER_ENV_UPDATE_STARTED   | Start updating the LoadBalancer on Data Lake in an environment.  |
| LOAD_BALANCER_ENV_UPDATE_FAILED    | Failed to update the LoadBalancer on Data Lake in an environment (Detailed message in the statusReason). |
| LOAD_BALANCER_STACK_UPDATE_STARTED | Start updating the LoadBalancer on Data Hubs in an environment.  |
| LOAD_BALANCER_STACK_UPDATE_FAILED  | Failed to update the LoadBalancer on Data Hubs in an environment (Detailed message in the statusReason). |

## Stop and restart an environment

You can stop an environment if you need to suspend but not terminate the resources within the environment. When you stop an environment, all of the resources within the environment are also stopped, including Data Lakes and Data Hubs. You can also restart the environment.



### Warning:

The Machine Learning service does not support environment stop and restart. This means that if ML workspaces are running or expected to be provisioned within an environment, then the environment should not be stopped. If done, this will disrupt running CML workspaces and prevent successful provisioning of ML workspaces in the environment.

Required role: EnvironmentAdmin or Owner

### Steps

#### For CDP UI

1. Navigate to the environment in Management Console > Environments.
2. Click **Actions Stop Environment** and confirm the action.
3. To restart the environment, click **Actions Start Environment**.



**Warning:** We have not tested or certified restarting the environment while Cloudera Data Engineering (CDE) is running.

#### For CDP CLI

Use the following command to stop an environment:

```
cdp environments stop-environment --environment-name <ENVIRONMENT_NAME>
```

Use the following commands to start an environment:

```
cdp environments start-environment --environment-name <ENVIRONMENT_NAME>
[--with-datahub-start]
```

Use the following commands to start an environment and all Data Hubs running in it:

```
cdp environments start-environment --environment-name <ENVIRONMENT_NAME>
--with-datahub-start
```



## Delete an environment

Deleting an environment terminates all resources within the environment including the Data Lake.


Before you begin

To delete an environment, you should first terminate all clusters running in that environment.

Required role: Owner or PowerUser

Steps

### For CDP UI

1. In Management Console, navigate to Environments.
  2. Click on your environment.
  3. Click **Actions Delete** and confirm the deletion.
    - Check the box next to "I would like to delete all connected resources" if you have Data Lake and Data Hub clusters running within the environment. This will delete the Data Lake and Data Hub clusters together with the rest of the environment.
-  **Note:** The "I would like to delete all connected resources" option does not delete any Data Warehouse or Machine Learning clusters running within the environment, so these always need to be terminated prior to environment termination.
- Check the box next to "I understand this action cannot be undone". This is required.
4. Click Delete.

### For CDP CLI

When terminating an environment from the CDP CLI, you need to first terminate the Data Lake:

1. Terminate the Data Lake using the following command:

```
cdp datalake delete-datalake --datalake-name <value>
```

2. Wait until the Data Lake terminates before proceeding. Use the following commands to check on the status of Data Lake:

```
cdp datalake get-cluster-host-status --cluster-name <value>
```

```
cdp datalake list-datalakes
```

3. Delete the environment using the following command:

```
cdp environments delete-environment --environment-name <value> --cascading
```

The `--cascading` option deletes all Data Hubs running in the environment.

If environment deletion fails, you can:

- Repeat the environment deletion steps, but also check "I would like to force delete the selected environment". Force deletion removes CDP resources from CDP, but leaves cloud provider resources running.
- Clean up cloud resources that were left on your cloud provider account.

Only the resources that were provisioned as part of the environment are deleted. For example, if a new network was created by CDP for the environment, the network will be deleted; But if you provided your existing network, it will not be deleted as part of environment deletion.

## Change environment's credential



You can change the credential attached to an environment as long as the new credential provides the required level of access to the same GCP account as the old credential.

Required roles:

- EnvironmentAdmin or Owner of the environment
- SharedResourceUser or Owner of the credential

Steps

### For CDP UI

1. Log in to the CDP web interface.
2. Navigate to the Management Console.
3. Select Environments from the navigation pane.
4. Click on a specific environment.
5. Navigate to the Summary tab.
6. Scroll down to the Credential section.
7. Click  to access credential edit options.
8. Select the new credential that you would like to use.
9. Click  to save the changes.

### For CDP CLI

If you would like to delete a credential from the CDP CLI, use:

```
cdp environments update-environment-credential --environment-name <value>
--credential-name <value>
```

## Defining custom tags

In the Management Console user interface, you can define tenant-level or environment-level custom tags across all instances and resources provisioned in your organization's cloud provider account.

### Resource tagging

When you create an environment or other resources shared across your cloud provider account, CDP automatically adds default tags to the Cloudera-created resources in your cloud provider account. You can also define additional custom tags that CDP applies to the cluster-related resources in your account.

You can use tags to protect the cloud resources of your CDP environment. Using the tags, you can exclude the resources that should not be removed during housekeeping or security deletion activities that can result in data corruption and data loss.

Default tags

By default, CDP applies certain tags to cloud provider resources whenever you create the resource, for example an environment.

CDP applies the following tags by default:

- **Cloudera-Resource-Name:** the workload-appropriate Cloudera resource name. For example, an IPA CRN for an IPA, a data lake CRN for a data lake, or a Data Hub CRN for a Data Hub cluster. This CRN serves as a unique identifier for the resource over time.
- **Cloudera-Creator-Resource-Name:** Cloudera resource name of the CDP user that created the resource.
- **Cloudera-Environment-Resource-Name:** name of the environment with which the resource is associated.

### Custom tags

There are two types of custom tags that you can define in the Management Console: tenant-level tags that apply to Cloudera-created resources across your organization's entire cloud provider account, and environment-level tags.

In the Management Console user interface, you can define tenant-level tags across all instances and resources provisioned in your organization's cloud provider account. These resources include not only provisioned instances, but disks, networks, and other resources as well. In your cloud provider account you can search or filter on either the tag key or value. Tenant-level tags cannot be overridden during environment creation.

In addition to tenant-level tags, you can also define environment-level tags. Environment-level tags are inherited by the resources specific to an environment. For example, environment-level tags are inherited by the following resources:

- FreeIPA
- Data lakes
- Data Hubs
- Operational Databases

As with tenant-level tags, you can search or filter on the key tag or key value in your cloud provider account.



**Note:** CDP applies custom tags during creation of the resources. For example, you can only define environment-level tags during environment registration. If you want to add or update cloud provider resource tags, you must do so through the cloud provider API.

For more information about using tags on cloud provider resources, consult AWS, Azure, or Google Cloud documentation. It is your responsibility to ensure that your tags meet your cloud provider requirements.

## Supported services

While some CDP services such as Data Hub inherit environment-level tags, others require that you add tags when provisioning or enabling the data service. The following table shows how tags can be added for various CDP services:

| CDP service          | How to add tags   |
|----------------------|---|
| Data Lake            | Inherits tenant or environment level tags.  |
| FreeIPA              | Inherits tenant or environment level tags.  |
| Data Hub             | Inherits tenant or environment level tags and you can add more tags when creating a Data Hub.             |
| Operational Database | Inherits tenant or environment level tags and you can add more tags when creating a COD database via CLI. |

## Defining tenant-level tags

Required roles: PowerUser can define tags for the whole tenant.

- EnvironmentAdmin or Owner can set environment telemetry settings for a specific environment.

### Steps

1. In the Management Console, click **Global Settings Tags**.
2. Click **Add**.

3. Define both a key and a value for the tag. Both the key and the value must be between 4- 255 characters, with the following restrictions:

#### Key

Allowed characters are hyphens (-), underscores (\_), lowercase letters, and numbers. Keys must start with a lowercase letter and must not start or end with spaces.

#### Value

Allowed characters are hyphens (-), underscores (\_), lowercase letters, and numbers. Values must not start or end with spaces. You can configure variables in the `{{{variableName}}}` format. The following variables are supported for tenant-level tags:

- `{{{cloudPlatform}}}` = AWS, AZURE or GCP
- `{{{userName}}}` = CDP username
- `{{{userCrn}}}` = Customer Resource Number (CRN) of CDP user
- `{{{creatorCrn}}}` = CRN of CDP resource creator
- `{{{time}}}` = Actual time
- `{{{accountId}}}` = CDP account ID
- `{{{resourceCrn}}}` = Generated string of CDP resource CRN

4. Click Add, and if necessary repeat the process for additional tags.



**Note:** Tenant-level tags are applied only to resources created after you define the tag. Any changes to the tag value do not propagate to existing resources that have the tag.

### Defining environment-level tags

You define environment-level tags during environment registration.

Required roles: EnvironmentCreator can set tags for a specific environment during environment registration.

Steps

1. In the Management Console, click **Environments Register Environment**.
2. Proceed through the environment registration steps.
3. After you define data access, add any environment-level tags by clicking **Add** and defining the tag key and value.

### Related Information

[Label format](#)

## Adding a customer managed encryption key for GCP

By default, a Google-managed encryption key is used to encrypt disks and Cloud SQL instances in Data Lake, FreeIPA, and Data Hub clusters, but you can optionally configure CDP to use a customer-managed encryption key (CMEK) instead.

To set up a CMEK, you should:

1. Meet the CMEK prerequisites.
2. Register a GCP environment in CDP via CDP web UI or CLI. During environment registration, specify the encryption key that you would like to use.

### CMEK prerequisites

Refer to [GCP Prerequisites: Customer managed encryption keys](#).

### Create a CDP environment with a CMEK

You can pass the CMEK during GCP environment registration in CDP via CDP web interface or CDP CLI.

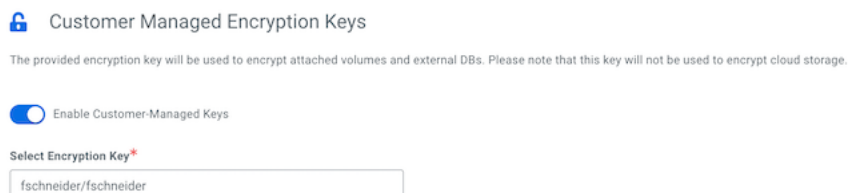
## Steps

**For CDP UI**

You can register your environment as described in [Register a GCP environment from CDP UI](#), just make sure that on the Data Access and Audit page you enable the following:

1. Under Customer-Managed Encryption Keys, click Enable Customer-Managed Keys.
2. In the same section, select the CMEK:

The following screenshot shows the UI options:

**For CDP CLI**

The steps below can only be performed via CDP CLI. Create an environment passing the `--encryption-key` parameter as shows in this example:

```
cdp environments create-gcp-environment \
  --no-use-public-ip \
  --environment-name <ENVIRONMENT_NAME> \
  --credential-name <EXISTING_CREDENTIAL-NAME> \
  --region <REGION> \
  --security-access securityGroupIdForKnox=<SG_NAME1>,defaultSecurityGroupId=<SG_NAME2> \
  --public-key <PUBLIC_SSH_KEY> \
  --log-storage storageLocationBase=<LOGS_STORAGE_LOCATION> \
  --existing-network-params networkName=<NETWORK>,subnetNames=<SUBNET>,
sharedProjectId=<PROJECT_ID> \
  --workload-analytics \
  --encryption-key <PATH_TO_THE_ENCRYPTION_KEY>
```



**Note:** If the `--encryption-key` parameter is not provided the GCP resources are not encrypted using CMEK, falling back to the default behavior of Google managed encryption.

Next, create a Data Lake and IDBroker mappings using the usual commands. Once the environment is running, Data Hubs can be created using the usual steps.