Management Console

Cloudera Management Console Top Tasks

Date published: 2022-08-30
Date modified:

https://docs.cloudera.com/

Legal Notice

Contents

Verifying your credential prerequisites

Cross-account access IAM role

To allow Cloudera to create resources in your AWS account, you create a cross-account access IAM role in your AWS account and grant Cloudera access to the role as a trusted principal by specifying a specific AWS account and an external ID.

The policy for the cross-account access IAM role must have the permissions enumerated in the documentation linked below. In addition, the IAM role must reference the specific AWS account ID and external ID provided in the Cloudera Management Console.

Prerequisites for the provisioning credential

To allow Cloudera to create resources on your Azure account, you must create the app-based credential. The credential allows Cloudera to access and provision a set of resources in your Azure account.

Cloudera uses an app-based credential to authenticate your Azure account and obtain authorization to create resources on your behalf. The app-based credential requires that you to manually configure the service principal created within your Azure Active Directory. The app-based method requires Owner role to be able to create a service principal, which must be given Contributor role or its equivalent.

To meet Azure prerequisites for Cloudera:

- Review the provided policies
- Perform the step 1 and step 2 described in the documentation for creating an app-based credential:

Service account for the provisioning credential

The provisioning credential for Google Cloud relies on a service account that can be assumed by Cloudera.

The following flow describes how the Google Cloud provisioning credential works:

1. Your GCP account administrator creates a service account and assigns the minimum permissions allowing Cloudera to create and manage resources in your Google Cloud account. Next, the administrator generates a service account access key pair for the service account.
2. The service account is registered as a credential in Cloudera and its access key is uploaded to Cloudera.
3. The credential is then used for registering your Google Cloud environment in Cloudera.
4. Once this is done, Cloudera uses the credential for provisioning environment-related resources, workload clusters, and resources for other Cloudera services that you run in Cloudera.

Review the following to learn about the permissions required for the credential and how to create the service account:

Registering your first environment

Registering an AWS environment from the Cloudera UI

Once you've met the AWS cloud provider requirements, register your AWS environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in AWS requirements.

Required role: EnvironmentCreator

Steps

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

Parameter	Description
General Information	
Environment Name (Required)	Enter a name for your environment. The name: • Must be between 5 and 28 characters long. • Can only include lowercase letters, numbers, and hyphens. • Must start with a lowercase letter.
Description	Enter a description for your environment.
Select Cloud Provider (Required)	Select Amazon.
Credential (Required)	
Select Credential	Select an existing credential or select Create new credential. For instructions on how to create a credential, refer to Creating a role-based credential. **Note:** Activate the Enable Permission Verification button if you want Cloudera to check permissions for your credential. Cloudera will verify that you have the required permissions for your environment.

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: • Must be between 5 and 100 characters long • Must contain lowercase letters • Cannot contain uppercase letters • Must start with a letter • Can only include the following accepted characters are: a-z, 0-9, -.
Data Lake Version (Required)	Select Cloudera Runtime version that should be deployed for your Data Lake. The latest stable version is used by default. All Cloudera Data Hub clusters provisioned within this Data Lake will be using the same Cloudera Runtime version.
Fine-grained access control on S3	
Enable Ranger authorization for AWS S3 Identity	Enable this if you would like to use Fine-grained access control. Next, from the Select AWS IAM role for Ranger authorizer dropdown, select the DATALAKE_ADMIN_ROLE IAM role created in Minimal setup for cloud storage.
Data Access and Audit	
Assumer Instance Profile (Required)	Select the IDBROKER_ROLE instance profile created in Minimal setup for cloud storage.
Storage Location Base (Required)	Provide the S3 location created for data storage in Minimal setup for cloud storage.
Data Access Role (Required)	Select the DATALAKE_ADMIN_ROLE IAM role created in Minimal setup for cloud storage.
Ranger Audit Role (Required)	Select the RANGER_AUDIT_ROLE IAM role created in Minimal setup for cloud storage.

Parameter	Description
IDBroker Mappings	We recommend that you leave this out and set it up after registering your environment as part of Onboarding Cloudera users and groups for cloud storage.  **Note:** If you are using Fine-grained access control, this option is disabled, because you should onboard your users and groups via Ranger instead of using IDBroker mappings.
Scale (Required)	Select Data Lake scale. By default, “Light Duty” is used. For more information on Data Lake scale, refer to Data Lake scale.
Enable **Compute Cluster**	Enable Compute Clusters if you would like to deploy a containerized platform on Kubernetes for data services and shared services.

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

Parameter	Description
Network and Availability	
Enable Multiple Availability Zones for Data Lake	Click the Enable Multiple Availability Zones for Data Lake toggle button to enable multi-AZ for the Data Lake. This option is disabled by default and is only available when a Medium Duty Data Lake is selected. Refer to Deploying Cloudera in multiple AWS availability zones.
Hardware and Storage	For each host group you can specify an instance type. For more information on instance types, see Amazon EC2 instance types.
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 Recipes.

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

Parameter	Description
Region	
Select Region (Required)	Select the region that you would like to use for Cloudera. If you would like to use a specific existing virtual network, the virtual network must be located in the selected region.
Customer-managed Keys	
Enable Customer-Managed Keys	Enable this if you would like to provide a Customer-Managed Key (CMK) to encrypt environment's disks and databases. Next, under Select Encryption Key, select an existing CMK. For more information, refer to Customer managed encryption keys.
Select Encryption Key	Select an existing CMK.
Network	
Select Network (Required)	Select the existing virtual network where you would like to provision all Cloudera resources. Refer to VPC and subnet.
Select Subnets (Required)	This option is only available if you choose to use an existing network. Multiple subnets must be selected and Cloudera distributes resources evenly within the subnets.
Enable Public Endpoint Access Gateway	When Cluster Connectivity Manager is enabled, you can optionally enable Public Endpoint Access Gateway to provide secure connectivity to UIs and APIs in Data Lake and Cloudera Data Hub clusters deployed using private networking. Under Select Endpoint Access Gateway Subnets, select the public subnets for which you would like to use the gateway. The number of subnets must be the same as under Select Subnets and the availability zones must match. For more information, refer to Public Endpoint Access Gateway documentation.
Proxies	
Select Proxy Configuration	Select a proxy configuration if previously registered. For more information refer to Setting up a proxy server.

Parameter	Description
Security Access Settings	
Select Security Access Type (Required)	This determines inbound security group settings that allow connections to the Data Lake and Cloudera Data Hub clusters from your organization's computers. You have two options: • Create new security groups - Allows you to provide custom CIDR IP range for all new security groups that will be created for the Data Lake and Cloudera Data Hub clusters so that users from your organization can access cluster UIs and SSH to the nodes. This must be a valid CIDR IP in IPv4 range. For example: 192.168.27.0/24 allows access from 192.168.27.0 through 192.168.27.255. You can specify multiple CIDR IP ranges separated with a comma. For example: 192.168.27.0/24,192.168.28.0/24 If you use this setting, several security groups will get created: one for each Data Lake host group the Data Lake and one for each host group), one for each FreeIPA host group, and one for RDS; Furthermore, the security group settings specified will be automatically used for Cloudera Data Hub, Cloudera Data Warehouse, and Cloudera AI clusters created as part of the environment. • Provide existing security groups (Only available for an existing VPC) - Allows you to select two existing security groups, one for Knox-installed nodes and another for all other nodes. If you select this option, refer to Security groups to ensure that you open all ports required for your users to access environment resources.
Kubernetes	
Select **Private Kubernetes Cluster** or provide **Authorized IP Ranges**	If you have enabled Compute Clusters, you have the following options to configure the necessary networking information for the Kubernetes cluster: • Enable Private Kubernetes Cluster to create a private cluster that blocks all access to the API Server endpoint. • Provide the CIDRs to the Kubernetes API Server Authorized IP Ranges field to specify a set of IP ranges that will be allowed to access the Kubernetes API server. You need to provide the advanced configurations only once when creating your environment. The configurations will be applied to all compute clusters in the environment.
Worker Node Subnets	Uses the same set of subnets provided in **Network** section. You have the option to not use all of the previously provided subnets.
SSH Settings	
New or existing SSH public key (Required)	You have two options for providing a public SSH key: • Select a key that already exists on your AWS account within the specific region that you would like to use. • Upload a public key directly from your computer.  **Note:** Cloudera does not use this SSH key. The matching private key can be used by your Cloudera administrator for root-level access to the instances provisioned for the Data Lake and Cloudera Data Hub.
Add tags	You can optionally add tags to be created for your resources on AWS. Refer to Defining custom tags.

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

Parameter	Description
Network and Availability	
Enable Multiple Availability Zones for Data Lake	Click the Enable Multiple Availability Zones for Data Lake toggle button to enable multi-AZ for the FreeIPA cluster. Refer to Deploying Cloudera in multiple AWS availability zones.
Hardware and Storage	For each host group you can specify an instance type. For more information on instance types, see Amazon EC2 instance types.
Cluster Extensions	
Recipes	You can optionally select and attach previously registered recipes to run on a specific FreeIPA host group. For more information, see Recipes.

9. Click Next.

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

Parameter	Description
Logs	
Logger Instance Profile (Required)	Select the LOG_ROLE instance profile created in Minimal setup for cloud storage.
Logs Location Base (Required)	Provide the S3 location created for log storage in Minimal setup for cloud storage.
Backup Location Base	Provide the S3 location created for FreeIPA and Data Lake backups in Minimal setup for cloud storage. 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 Cloudera Observability. For more information, refer to Enabling workload analytics and logs collection.
Enable Deployment Cluster Logs Collection	When this option is enabled. the logs generated during deployments will be automatically sent to Cloudera. For more information, refer to Enabling workload analytics and logs collection.

11. Click on 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 Cloudera 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.

Registering an AWS environment from the CDP CLI

Once you’ve met the AWS cloud provider requirements, register your AWS environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in AWS requirements.

Required role: EnvironmentCreator

Steps

Unlike in the Cloudera web interface, in CDP CLI environment creation is a three-step process with environment creation, setting IDBroker mappings and Data Lake creation being three separate steps. The easiest way to obtain the correct commands is to provide all parameters in Cloudera web interface and then generate the CDP CLI commands on the last page of the wizard. For detailed steps, refer to Obtain CLI commands for registering an environment.

To learn more about how to create Compute Cluster enabled environments with CLI, see Enabling default Compute Cluster for new environments.

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 Cloudera 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.

Registering an Azure environment from the Cloudera UI

Once you've met the Azure cloud provider requirements, register your Azure environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in Azure requirements.

Required role: EnvironmentCreator

Steps

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

Parameter	Description
General Information	
Environment Name (Required)	Enter a name for your environment. The name: • Must be between 5 and 28 characters long. • Can only include lowercase letters, numbers, and hyphens. • Must start with a lowercase letter.
Description	Enter a description for your environment.
Select Cloud Provider (Required)	Select Azure.
Microsoft Azure Credential (Required)	
Select Credential	Select an existing credential or select Create new credential. For instructions on how to create a credential, refer to Create an app-based credential.

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: • Must be between 5 and 100 characters long • Must contain lowercase letters • Cannot contain uppercase letters • Must start with a letter • Can only include the following accepted characters are: a-z, 0-9, -.
Data Lake Version (Required)	Select Cloudera Runtime version that should be deployed for your Data Lake. The latest stable version is used by default. All Cloudera Data Hub clusters provisioned within this Data Lake will be using the same Cloudera Runtime version.
Fine-grained access control on ADLS Gen2	

Parameter	Description
Enable Ranger authorization for ADLS Gen2 Identity	If you would like to use Fine-grained access control, enable this option and then select the Ranger RAZ managed identity created in the Minimal setup for cloud storage.
Data Access and Audit	
Assumer Identity (Required)	Select the Assumer managed identity created in Minimal setup for cloud storage.
Storage Location Base (Required)	Provide the ADLS Gen2 location created for data storage in Minimal setup for cloud storage.
Data Access Identity (Required)	Select the Data Lake Admin managed identity created in Minimal setup for cloud storage.
Ranger Audit Identity (Required)	Select the Ranger Audit managed identity created in Minimal setup for cloud storage.
IDBroker Mappings	We recommend that you leave this out and set it up after registering your environment as part of Onboarding Cloudera users and groups for cloud storage.  **Note:** If you are using Fine-grained access control, this option is disabled, because you should onboard your users and groups via Ranger instead of using IDBroker mappings.
Scale (Required)	Select Data Lake scale. By default, “Light Duty” is used. For more information on data lake scale, refer to Data Lake scale.
Enable **Compute Cluster**	Enable Compute Clusters if you would like to deploy a containerized platform on Kubernetes for data services and shared services.

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 Sizes for virtual machines in Azure.
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 Recipes.

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

Parameter	Description
Region	
Select Region (Required)	Select the region that you would like to use for accessing and provisioning resources from Cloudera. If you would like to use a specific existing virtual network, the virtual network must be located in the selected region.
Resource Group	
Select Resource Group (Required)	You have two options: • Select one existing resource group. If you select this, all Cloudera resources will be provisioned into that resource group. • Select Create new resource groups to have Cloudera create multiple resource groups.
Customer Managed Encryption Keys	

Parameter	Description
Enable Customer-Managed Keys	Enable this if you would like to provide a Customer-Managed Key (CMK) to encrypt environment's disks and databases. For more information, refer to Customer managed encryption keys.
Select Encryption Key Resource Group	Select the resource group where the CMK is located.
Encryption key URL	Provide the URL of the key value where the CMK resides. This is the same as the key identifier that you can copy directly from Azure Portal.
Managed identity for encryption	If using Azure Database for PostgreSQL Flexible Server, you can optionally select a managed identity created for encrypting it. For more information, refer to Managed identity for encrypting Azure Database for PostgreSQL Flexible Server.
Network	
Select Network (Required)	You have two options: • Select the existing virtual network where you would like to provision all Cloudera resources. Refer to VNet and subnets. • Select Create new network to have a new network with three subnets created.
Select Subnets (Required)	This option is only available if you choose to use an existing network. Multiple subnets must be selected and Cloudera distributes resources evenly within the subnets.
Network CIDR (Required)	This option is only available if you select to create a new network. If you selected to create a new network, provide Network CIDR that determines the range of private IPs that VMs will use. This must be a valid private IP CIDR IP in IPv4 range. For example 10.10.0.0/16 are valid IPs. /16 is required to allow for enough IP addresses.
Create Private Subnets	This option is only available if you select to have a new network and subnets created. Is is turned on by default so that private subnets are created in addition to public subnets. If you disable it, only public subnets will be created. **Important:** For production deployments, Cloudera recommends that you use private subnets. Work with your internal IT teams to ensure that users can access the browser interfaces for cluster services.
Enable Public Endpoint Access Gateway	When Cluster Connectivity Manager is enabled, you can optionally enable Public Endpoint Access Gateway to provide secure connectivity to UIs and APIs in Data Lake and Cloudera Data Hub clusters deployed using private networking. If you are using your existing VPC, under Select Endpoint Access Gateway Subnets, select the public subnets for which you would like to use the gateway. The number of subnets must be the same as under Select Subnets and the availability zones must match. For more information, refer to Public Endpoint Access Gateway documentation.
Create Private Endpoints	By default, the PostgreSQL Azure database provisioned for your Data Lake is reachable via a service endpoint (public IP address). To increase security, you can optionally select to have it reachable via a private endpoint instead of a service endpoint. **Note:** This option is only available if an existing resource group is selected. **Note:** Only the subnets that have Azure private endpoint network policies turned off are eligible for private endpoint creation. At least one such subnet is required. If you select to create a private endpoint and you are using your own VNet, you have two options: • Select “Create new private DNS zone” and Cloudera creates and manages a private DNS zone for you in the provided existing resource group. • Select your existing private DNS zone. If you select to create a private endpoint and you would like for Cloudera to create a new VNet, Cloudera creates a private DNS zone for you. For more information, refer to Private endpoint for Azure Postgres.
Create Public IPs	This option is disabled by default when Cluster Connectivity Manager is enabled and enabled by default when Cluster Connectivity Manager is disabled.

Parameter	Description
Flexible Server	During environment registration in Cloudera, the Flexible Server in public service mode is used by default, but you can specify to use the Flexible Server in private service mode (“Flexible Server with Private Link” or "Flexible Server with Delegated Subnet (deprecated)"). For more information, refer to Using Azure Database for PostgreSQL Flexible Server.
Proxies	
Select Proxy Configuration	Select a proxy configuration if previously registered. For more information refer to Setting up a proxy server.
Security Access Settings	
Select Security Access Type (Required)	This determines inbound security group settings that allow connections to the Data Lake and Cloudera Data Hub clusters from your organization’s computers. You have two options: • Create new security groups - Allows you to provide custom CIDR IP range for all new security groups that will be created for the Data Lake and Cloudera Data Hub clusters so that users from your organization can access cluster UIs and SSH to the nodes. This must be a valid CIDR IP in IPv4 range. For example: 192.168.27.0/24 allows access from 192.168.27.0 through 192.168.27.255. You can specify multiple CIDR IP ranges separated with a comma. For example: 192.168.27.0/24,192.168.28.0/24. If you use this setting, several security groups will get created: one for each Data Lake host group the Data Lake and one for each host group), one for each FreeIPA host group, and one for RDS; Furthermore, the security group settings specified will be automatically used for Cloudera Data Hub, Data Warehouse, and Cloudera AI clusters created as part of the environment. • Provide existing security groups (Only available for an existing VPC) - Allows you to select two existing security groups, one for Knox-installed nodes and another for all other nodes. If you select this option, refer to Security groups to ensure that you open all ports required for your users to access environment resources.
Kubernetes	
Enable **Private Kubernetes Cluster** or provide **Authorized IP Ranges**	If you have enabled Compute Clusters, you have the following options to configure the necessary networking information for the Kubernetes cluster: • Enable Private Kubernetes Cluster to create a private cluster that blocks all access to the API Server endpoint. • Provide the CIDRs to the Kubernetes API Server Authorized IP Ranges field to specify a set of IP ranges that will be allowed to access the Kubernetes API server. You need to provide the advanced configurations only once when creating your environment. The configurations will be applied to all compute clusters in the environment.
Enable **User Defined Routing**	Enable User Defined Routing (UDR) in case public IPs are blocked for egress.In case you enable UDR, you must select the specific worker node subnet where the UDR is configured.
AKS Private DNS Zone ID	When selecting **Private Kubernetes Cluster**, you also need to select an existing private DNS zone or select creating a new private DNS zone by Cloudera on your Azure account for the database.
Worker Node Subnets	Uses the same set of subnets provided in **Network** section. You have the option to not use all of the previously provided subnets.
SSH Settings	
New SSH public key (Required)	Upload a public key directly from your computer.  **Note:** Cloudera does not use this SSH key. The matching private key can be used by your Cloudera administrator for root-level access to the instances provisioned for the Data Lake and Cloudera Data Hub.
Add tags	You can optionally add tags to be created for your resources on Azure. Refer to Defining custom tags.

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 Sizes for virtual machines in Azure.
Cluster Extensions	
Recipes	You can optionally select and attach previously registered recipes to run on FreeIPA nodes. For more information, see Recipes.

9. Click Next.
10. On the Storage page, provide the following information:

Parameter	Description
Logs	
Logger Identity (Required)	Select the Logger managed identity created in Minimal setup for cloud storage.
Logs Location Base (Required)	Provide the ADLS Gen2 location created for log storage in Minimal setup for cloud storage.
Backup Location Base	Provide the ADLS Gen2 location created for FreeIPA and Data Lake backups in Minimal setup for cloud storage. 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 Cloudera Observability. For more information, refer to Enabling workload analytics and logs collection.
Enable Deployment Cluster Logs Collection	When this option is enabled. the logs generated during deployments will be automatically sent to Cloudera. For more information, refer to Enabling workload analytics and logs collection.

11. Click on 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 Cloudera 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.

Registering an Azure environment from the CDP CLI

Once you’ve met the Azure cloud provider requirements, register your Azure environment.

Before you begin

This assumes that you have already fulfilled the environment prerequisites described in Azure requirements.

Required role: EnvironmentCreator

Steps

Unlike in the Cloudera web interface, in CDP CLI environment creation is a three-step process with environment creation, setting IDBroker mappings and Data Lake creation being three separate steps. The easiest way to obtain the correct commands is to provide all parameters in the Cloudera web interface and then generate the CDP CLI commands on the last page of the wizard. For detailed steps, refer to Obtain CLI commands for registering an environment.

To learn more about how to create Compute Cluster enabled environments with CLI, see Enabling default Compute Cluster for new environments.

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 Cloudera 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.

Registering a GCP environment from the Cloudera 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 Cloudera Management Console > Environments > Register environment.
2. On the Register Environment page, provide the following information:

Parameter	Description
General Information	
Environment Name (Required)	Enter a name for your environment. The name: • Must be between 5 and 28 characters long. • Can only include lowercase letters, numbers, and hyphens. • Must start with a lowercase letter.
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. For instructions on how to create a credential for Google Cloud, refer to Create a provisioning credential for GCP.

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

Parameter	Description
Data Lake Settings	

Parameter	Description
Data Lake Name (Required)	Enter a name for the Data Lake cluster that will be created for this environment. The name: • Must be between 5 and 100 characters long • Must contain lowercase letters • Cannot contain uppercase letters • Must start with a letter • Can only include the following accepted characters are: a-z, 0-9, -.
Data Lake Version (Required)	Select Cloudera Runtime version that should be deployed for your Data Lake. The latest stable version is used by default. All Data Hub clusters provisioned within this Data Lake will be using the same Cloudera Runtime version. Note: Google Cloud environments can only be provisioned in Cloudera with Cloudera Runtime version 7.2.8 or newer.
Data Access and Audit	
Assumer Service Account (Required)	Select the IDBroker service account created in Minimum setup for cloud storage.
Storage Location Base (Required)	Select the Google Storage location created for data in Minimum setup for cloud storage.
Data Access Service Account (Required)	Select the Data Lake Admin service account created in Minimum setup for cloud storage.
Ranger Audit Service Account (Required)	Select the Ranger Audit service account created in Minimum setup for cloud storage.
IDBroker Mappings	We recommend that you leave this out and set it up after registering your environment as part of Onboarding Cloudera users and groups for cloud storage.
Scale (Required)	Select Data Lake scale. By default, “Light Duty” is used. For more information on Data Lake scale, refer to Data Lake scale.

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 Machine type families.
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 Recipes.

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

Parameter	Description
Region	
Select Region (Required)	Select the region where your VPC network is located.
Select Zone (Required)	Select a zone within the selected region.
Network	
Use shared VPC	This option is disabled by default. Enable this if you would like to use your existing shared VPC. Next enter: • Host project ID • Network name • Subnet name(s). If providing multiple, provide a comma separated list.

Parameter	Description
Select Network (Required)	Select the existing VPC network that you created as a prerequisite in the VPC network and subnets step. All Cloudera 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 Cluster Connectivity Manager is enabled and enabled by default when Cluster Connectivity Manager is disabled.
Proxies	Select a proxy configuration if previously registered. For more information refer to Setting up a proxy server.
Security Access Settings	
Select Security Access Type (Required)	You have two options: • 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. • 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. For information on required ports, see Firewall rules.
SSH Settings	
New SSH public key (Required)	Upload a public key directly from your computer. Note: Cloudera 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 Cloudera Data Hub.
Add tags	You can optionally add tags to be created for your resources on GCP. Refer to Defining custom tags.

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 Machine type families.
Cluster Extensions	
Recipes	You can optionally select and attach previously registered recipes to run on FreeIPA nodes. For more information, see Recipes.

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 Minimum setup for cloud storage.
Logs Location Base (Required)	Select the Google Storage location created for logs in Minimum setup for cloud storage.
Backup Location Base	Select the Google Storage location created for FreeIPA backups in Minimum setup for cloud storage. If not provided, the default Backup Location Base uses the Logs Location Base.
Telemetry	

Parameter	Description
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 Cloudera 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.

Registering a GCP environment from the 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 Cloudera web interface, in the 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 Cloudera.

1. Once you’ve met the prerequisites, register your GCP environment in Cloudera 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 AAAAB3NzaZ1yc2EAAAADAQABAAABAQDwCI/wmQzbNn9YcA8v
dU+Ot41IIUWJfOfiDrUuNcULOQL6ke5qcEKuboXzbLxV0YmQcPFvswbM5S4FlHjy2VrJ5spy
GhQajFEm9+PgrsybgzHkkssziX0zRq7U4BVD68kSn6CuAHj9L4wx8WBwefMzkw7uO1CkfifI
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	Provide a JSON specifying the following: `{ "networkName": "string", "subnetNames": ["string", ...], "sharedProjectId": "string" }` Replace the values with the actual VPC network name, one or more subnet names and shared project ID. The sharedProjectId value needs to be set in the following way: • For a shared VPC, set it to the GCP host project ID • For a non-shared VPC, set it to the GCP project ID of the project where Cloudera is being deployed.
enableTunnel	By default Cluster Connectivity Manager 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: `"securityAccess": { "securityGroupIdForKnox": "string", "defaultSecurityGroupId": "string" }`
usePublicIp	Set this to “true” or “false”, depending on whether or not you want to create public IPs.

Parameter	Description
logStorage	Provide a JSON specifying your configuration for cluster and audit logs: `{ "storageLocationBase": "string", "serviceAccountEmail": "string" }` The storageLocationBase should be in the following format: gs://my-bucket-name.

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.

2. To verify that your environment is running, use:

```
cdp environments list-environments
```

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

3. 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-1:45ca3068-42a6-422
7-8394-13a4493e2ac0: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	Map Cloudera users or groups to GCP service accounts created earlier. Use the following syntax: `[ { "accessorCrn": "string", "role": "string" } ... ]` You can obtain user or group CRN from the Management Console > User Management by navigating to details of a specific user or group. The role should be specified as service account email.

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 d
emo3
```

6. One 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
datalakeName	Provide a name for your Data Lake.
environmentName	Provide a name of the environment created earlier.
scale	Provide Data Lake scale. It must be one of: • LIGHT_DUTY or • MEDIUM_DUTY_HA.
cloudProviderConfiguration	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 Cloudera 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 Cloudera 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.