Cloudera Public Cloud

GCP Requirements

Date published: 2019-08-22
Date modified: 2025-08-18

https://docs.cloudera.com/

Legal Notice

Contents

GCP requirements

Use the following guidelines to ensure that your Google Cloud account has all the necessary resources required by Cloudera and that Cloudera can access these resources:

Related Information

GCP permissions

GCP resources and services

Overview of GCP resources used by Cloudera

GCP outbound network access destinations

Access to workload UIs

Supported browsers

GCP permissions

As an administrator, you must be able to create and manage the resources in the Google Cloud project where Cloudera environments and clusters run. You must be able to perform all administrative tasks and have administrative rights to all resources.

Cloudera recommends that the administrator has the role of Owner in the project used for Cloudera in your GCP account.

GCP resources and services

Cloudera uses the following resources in your Google Cloud account.

Use the following guidelines to ensure that your GCP account has all the necessary resources required by Cloudera and that Cloudera can access these resources:

GCP project

In order to deploy Cloudera resources, you must create a project on your Google Cloud account.

If you would like to use a shared VPC, the VPC from the host project needs to be shared with the newly created project.

If you need to create a project, refer to Creating a project in Google documentation.

GCP APIs

Review the following list of GCP APIs and make sure that they are enabled for the project used for Cloudera.

The following Google APIs must be enabled for the project:

- iamcredentials.googleapis.com
- iam.googleapis.com
- compute.googleapis.com
- storage.googleapis.com
- servicenetworking.googleapis.com
- sqladmin.googleapis.com

GCP region

When registering a GCP environment in Cloudera, you must specify the same GCP region as the one where your VPC and subnets are located. Cloudera recommends that the buckets used for storage and logs are also located in the same region as the VPC.

When registering a GCP environment in Cloudera, you should also select a specific availability zone within the selected region.

Supported GCP regions

Cloudera supports all Google Cloud regions.

Region name	Region ID	Environment	Cloudera Data Hub	Cloudera Operational Database
AMERICAS				
OREGON	us-west1	##	##	##
LOS ANGELES	us-west2	##	##	##
SALT LAKE CITY	us-west3	##	##	##
LAS VEGAS	us-west4	##	##	##
IOWA	us-central1	##	##	##
SOUTH CAROLINA	us-east1	##	##	##
N. VIRGINIA	us-east4	##	##	##
MONTRÉAL	northamerica-northeast1	##	##	##
SÃO PAULO	southamerica-east1	##	##	##
DALLAS	us-south-1	##	##	##
COLUMBUS	us-east5	##	##	##
EUROPE				
LONDON	europe-west2	##	##	##
BELGIUM	europe-west1	##	##	##
NETHERLANDS	europe-west4	##	##	##
ZURICH	europe-west6	##	##	##
FRANKFURT	europe-west3	##	##	##
FINLAND	europe-north1	##	##	##
MILAN	europe-west8	##	##	##
PARIS	europe-west9	##	##	##
MADRID	europe-southwest1	##	##	##
ASIA PACIFIC				
MUMBAI	asia-south1	##	##	##
SINGAPORE	asia-southeast1	##	##	##
JAKARTA	asia-southeast2	##	##	##
HONG KONG	asia-east2	##	##	##
TAIWAN	asia-east1	##	##	##
TOKYO	asia-northeast1	##	##	##

Region name	Region ID	Environment	Cloudera Data Hub	Cloudera Operational Database
OSAKA	asia-northeast2	##	##	##
SYDNEY	australia-southeast1	##	##	##
SEOUL	asia-northeast3	##	##	##
MELBOURNE	australia-southeast2	##	##	##
MIDDLE EAST				
Qatar Doha	me-central1	##	##	##
KSA	me-central2	##	##	##
TEL AVIV	me-west-1	##	##	##

When making a decision regarding what region to use, you should review Cloud locations in Google documentation and verify that the region that you would like to use has the compute instance types and storage options that meet your requirements.

VPC network and subnet

When registering a GCP environment in Cloudera, you must provide an existing VPC network. Cloudera will not create a VPC network for you.

The VPC must fulfill the following requirements:

- The VPC network must have one or more subnets in a single geographic region. All the subnets used for environment creation must be in the same geographic region.
- Your VPC should reside in the same region as the buckets used for storage and logs. This allows you to avoid latency and additional data transfer costs.
- A shared VPC can be used.
- If you would like to use a shared VPC, the VPC from the host project needs to be attached to the newly created project.
- When creating a VPC and subnets, there is an option to create subnets in a custom or automatic mode. Cloudera recommends using the custom mode.
- If you would like to use Public Endpoint Access Gateway, make sure that "Private Google Access" is disabled on at least one subnet in the VPC.

For instructions on how to create and manage VPC networks and subnets in GCP, refer to Using VPC networks in Google documentation.

Related Information

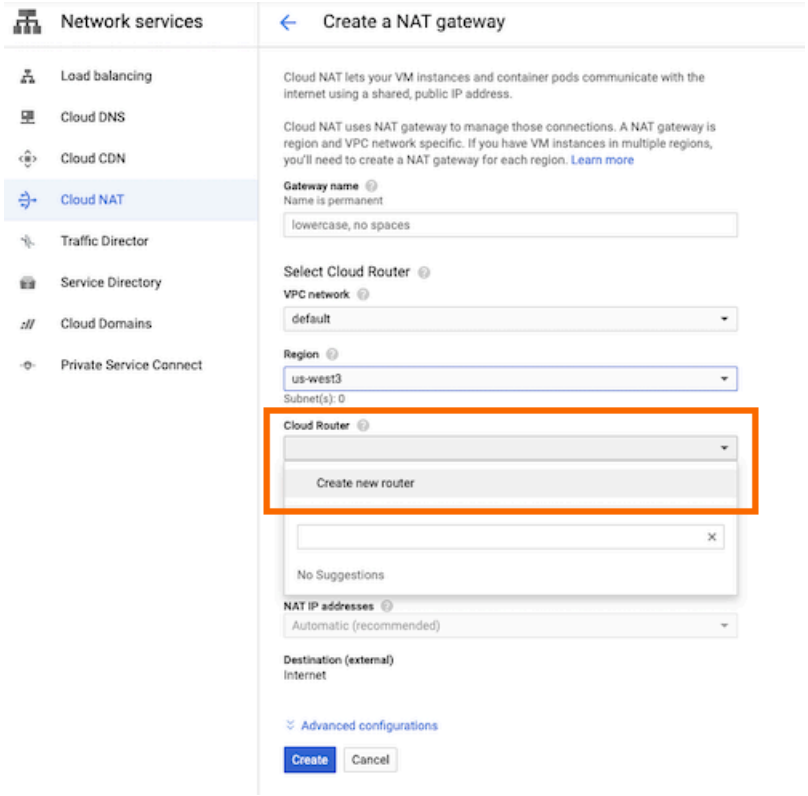
Provisioning Shared VPC

Internet connectivity

Depending on your network deployment scenario, Cloudera on GCP requires outbound internet connectivity and may additionally require inbound connectivity with public IP assignment.

The following is required, depending on whether or not you choose to use Cluster Connectivity Manager:

- With Cluster Connectivity Manager: Works with private IP address assignment and requires a NAT gateway to be provisioned. Note that NAT gateway creation will require a cloud router creation, as shown in the following screenshot:



- Without Cluster Connectivity Manager: Works with public IP address assignment and requires specific firewall rule configuration (as described in Firewall rules).

If you are planning to use Cluster Connectivity Manager, review the following GCP documentation for your reference:

- Cloud NAT Overview
- Using Cloud NAT

Related Information

Cluster Connectivity Manager

Firewall rules

Cloudera requires that you pre-create a set of firewall rules allowing your organization SSH and UI access to Cloudera and allowing internal communication between Cloudera components. Cloudera does not offer an option to create these firewall rules for you.

You have two options:

Option	VPC type supported for this option	What to do during environment registration
• You create all required firewall rules at the VPC level.	Per project VPC Shared VPC	In this case, you do not provide them to Cloudera during environment registration (That is, during environment registration you select "Do not create firewall rule").

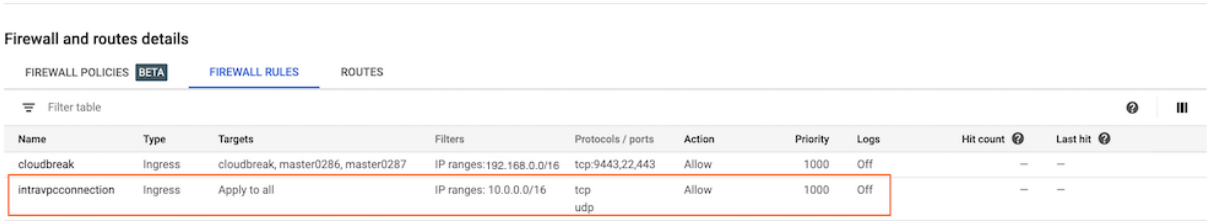
Option	VPC type supported for this option	What to do during environment registration
• You create the intravpc firewall rule at the VPC level. • Then, you create firewall rules for SSH and UI access via the security access mechanism in the Google Cloud UI. • If you need to create additional firewall rules (for example if you are not planning to use Cluster Connectivity Manager and you need to open ports 9443 and 443 for Cloudera), you should create these at the VPC level.	Per project VPC	In this case, you should select the firewall rules created for SSH and UI access during environment registration

Firewall rule requirements

You can add firewall rules in Google Cloud directly at the VPC level or via the security access control mechanism from VPC network > Firewall > Create firewall rule. For instructions on how to create and manage firewall rules in GCP, refer to Using firewall rules in Google documentation.

The firewall rules that you add should:

- Allow the instances in the VPC to connect with each other using TCP and UDP protocols on any port. To achieve this, add a TCP/UDP rule that is set to the subnet IP range. This is required for internal communication within the VPC. As an example, see the intravpcconnection firewall rule, which is set to the subnet IP range (10.0.0.0/16) in the following screenshot:



- Open TCP ports 22 and 443 to allow access from your organization's CIDR.
- If not using Cluster Connectivity Manager, also open TCP port 9443 to allow access from Cloudera CIDR.
- If not using Cluster Connectivity Manager, also open TCP port 443 to allow access from Cloudera CIDR. This is required for the gateway nodes.
- Open TCP/UDP ports 0-65535 to your VPC's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).
- Allow ICMP traffic from your internal VPC CIDR (for example 10.10.0.0/16).

Note: The communication via TCP/UDP 0-65535 and ICMP is essential for healthy operation of CDP environments, Data Hubs, and data services running within the , so ensure that you open these ports as described below. While some services only need well-known fixed ports, a majority of them depend on ephemeral (i.e. dynamically or randomly allocated) ports; This is why the wildcard 0-65535 TCP/UDP port range is used in the absence of a detailed breakdown of individual ports. Since overall access to the is typically secured by other means, the use of the wildcard rules does not pose a higher risk against external attacks.

Related Information

Cluster Connectivity Manager

Managed service network connection for CloudSQL

A CloudSQL database is created for the Data Lake cluster for external storage.

In order to use CloudSQL database with a private IP, your VPC needs to have private services access for CloudSQL. Private services access is implemented as a VPC peering connection between your VPC network and the underlying Google services VPC network where your CloudSQL instance resides. For more information about the setup and how to steps refer to Configuring private services access in Google documentation.

VM instances

Cloudera provisions VM instances as part of environment creation process (for Data Lake and FreeIPA) and for compute clusters.

Therefore, you should verify the limits on the number and type of VM instances in your GCP account to ensure that you are able to provision an environment and create clusters in Cloudera.

Custom images

By default Cloudera provides a set of default images that are used for all provisioned VMs, but you can optionally use custom images for Data Lake, FreeIPA, and Cloudera Data Hub.

You might require a custom image for compliance or security reasons (a “hardened” image), or to have your own packages pre-installed on the image, for example monitoring tools or software.

If you would like to use custom images instead of the default images, refer to Custom images and image catalogs.

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.

Permissions for the provisioning credential's service account

To allow Cloudera to access and provision resources in your Google Cloud project, you should create a service account in your Google Cloud project, assign the following roles or granular permissions. Next, you generate a JSON access key that can later be provided to Cloudera. Cloudera will assume this service account via the service account access key provided during credential creation for provisioning resources for your environment.

The service account must fulfill one of the following requirements (choose one of the options):

- Option 1: Assign the following IAM roles at the project level. This is a simpler option.
- Option 2: Alternatively, you can create custom IAM roles with the following granular IAM permissions assigned and then assign the role to the service account at the project level. This allows you to minimize the number of permissions granted to Cloudera.

Option 1: IAM roles

IAM role	Scope	Description
iam.serviceAccounts.list IAM permission	Project	This is required in order for Cloudera to be able to list service account names that you created in your GCP project. You need to create a custom role in order to assign this permission.
Compute Instance Admin (v1) roles/compute.instanceAdmin.v1	Project	This is required for provisioning of Compute Engine instances, disks, and images in your VPC.
Storage Admin roles/storage.admin	Project	This is required for the creation of a storage bucket to store the Cloudbreak image objects. Delete permissions are not required.
Compute Network Viewer roles/compute.networkViewer	Project	This is required for read-only access to all networking resources.
Compute Load Balancer Admin roles/compute.loadBalancerAdmin	Project	This role is required for load balancing between HA components of the Data Lake.
Cloud SQL Admin roles/cloudsql.admin	Project	This is required in order for Cloudera to have the permission for creating and deleting a Data Lake and and heavy duty flow management Cloudera Data Hub clusters cleanly.
Compute Network User roles/compute.networkUser	Project	Required for shared VPC only If you would like to use a shared VPC, you need this additional role in the scope of the host project of the VPC.
Compute Public IP Admin roles/compute.publicIpAdmin	Project	Required only when not using Cluster Connectivity Manager This additional role is only required if you are planning to disable Cluster Connectivity Manager for your environment.

Note: Additionally, once you create the Logger and IDBroker service accounts discussed in the minimum setup for cloud storage, you need to update each of these two service accounts to grant the provisioning service account the Service Account User (iam.serviceAccountUser) role. See instructions provided as part of Minimum setup for cloud storage.

Option 2: Granular permissions

You should create a custom IAM role to assign these permissions.

Granular IAM permissions	Scope	Description
Required for data encryption and decryption.		
cloudkms.cryptoKeys.list	Project	List all the keys inside any key ring.
cloudkms.keyRings.list	Project	List all the available key rings in the project.
cloudkms.cryptoKeyVersions.useToEncrypt	Project	Use a key to encrypt data.
cloudkms.cryptoKeyVersions.useToDecrypt	Project	Use a key to decrypt data.
Required to create, stop, start, and delete an external database from the Data Lake and Data Hub clusters.		
cloudsql.instances.create	Project	Create a new Cloud SQL instance.
cloudsql.instances.delete	Project	Delete a Cloud SQL instance.
cloudsql.instances.get	Project	View details of a Cloud SQL instance.
cloudsql.instances.list	Project	List all the Cloud SQL instances.

Granular IAM permissions	Scope	Description
cloudsql.instances.update	Project	Update a Cloud SQL instance.
cloudsql.users.create	Project	Create a new user on a Cloud SQL instance.
cloudsql.instances.startReplica	Project	Enables the creation of a read replica of an existing Cloud SQL instance.
cloudsql.instances.stopReplica	Project	Stop the replication process.
cloudsql.instances.restart	Project	Enable the restart process of a Cloud SQL instance.
Required to create VMs from images in your VPC.		
compute.addresses.create	Project	Create external IP addresses that can be assigned to Google Cloud resources like virtual machine (VM) instances, load balancers, etc.
compute.addresses.get	Project	Display both internal and external IP addresses.
compute.addresses.use	Project	Use both internal and external IP addresses.
compute.disks.create	Project	Create disks for the VM instances.
compute.disks.delete	Project	Clean up the disks.
compute.disks.setLabels	Project	Set or modify labels on disks (you or a service account).
compute.disks.use	Project	Use the disk.
compute.firewalls.list	Project	List the firewall rules associated with a project (you or a service account).
compute.forwardingRules.create	Project	Create forwarding rules.
compute.forwardingRules.delete	Project	Delete forwarding rules.
compute.forwardingRules.list	Project	List all forwarding rules.
compute.globalOperations.get	Project	View the status of global operations in Google Cloud, such as creating or deleting global resources.
compute.images.get	Project	View details of a specific image (you or a service account).
compute.images.useReadOnly	Project	Use an image, but only in read-only mode.
compute.instanceGroups.create	Project	Create instance groups.
compute.instanceGroups.delete	Project	Delete instance groups.
compute.instanceGroups.get	Project	Get information about a particular instance group.
compute.instanceGroups.update	Project	Update an instance group.
compute.instanceGroups.use	Project	Use an instance group.
compute.instances.create	Project	Create VM instances.
compute.instances.delete	Project	Delete VM instances,
compute.instances.get	Project	Get information about a particular instance.
compute.instances.setLabels	Project	Set or modify labels on a VM instance.
compute.instances.setMetadata	Project	Set or update the metadata of a VM instance.
compute.instances.setServiceAccount	Project	Set or update the service account associated with a VM.
compute.instances.setTags	Project	Set or modify tags on a VM instance.
compute.instances.start	Project	Start the instances.
compute.instances.stop	Project	Stop the instances.
compute.instances.update	Project	Modify the configuration of a specific instance. Useful for vertical scaling.

Granular IAM permissions	Scope	Description
compute.instances.use	Project	Use a VM instance.
compute.machineTypes.list	Project	List the virtual hardware configuration for a VM.
compute.networks.list	Project	List all the available networks.
compute.regionBackendServices.create	Project	Create a regional backend service. A regional backend service is part of the infrastructure that routes traffic to backend instances or groups within a specific region. This is useful for load balancers.
compute.regionBackendServices.delete	Project	Delete a regional backend service.
compute.regionBackendServices.use	Project	Use a regional backend service.
compute.regionHealthChecks.create	Project	Create health checks that monitor the health of the backend services, an important feature for load balancers.
compute.regionHealthChecks.delete	Project	Delete health checks.
compute.regionHealthChecks.useReadOnly	Project	Use health checks restricted to read-only mode.
compute.regionOperations.get	Project	View the status of region operations in Google Cloud. This is required regardless of globalOperations get permissions because both have different scopes.
compute.regions.get	Project	Get information about a specific region.
compute.regions.list	Project	List all the regions in the project.
compute.subnetworks.list	Project	List all the subnets in the project.
compute.subnetworks.use	Project	Use subnets.
compute.subnetworks.useExternalIp	Project	Assign external IP addresses to VM instances. If this is enabled, attempting to create a VM instance with an external IP address will fail.
compute.zoneOperations.get	Project	View the status of zone operations in Google Cloud. This is required for both global and regional level permissions.
compute.images.create	Project	Create a new image. This is required if you want to use a custom image.
compute.addresses.delete	Project	Delete static IP addresses that are allocated within a specific region or globally in Google Cloud (you or a service account).
compute.forwardingRules.setLabels	Project	Set or update forwarding rules labels.
compute.forwardingRules.use	Project	Use forwarding rules, most often used by load balancers to balance the traffic to backend services.
compute.regionHealthChecks.update	Project	Update the health checks that monitor the backend services.
compute.addresses.createInternal	Project	Create internal IP addresses within a Virtual Private Cloud (VPC). These internal IP addresses are used for private communication within your network, not exposed to the Internet.
compute.firewalls.create	Project	Create firewall rules (you or a service account).
compute.firewalls.delete	Project	Delete the firewall rules (you or a service account).
compute.subnetworks.get	Project	Get details about a specific subnet.
compute.networks.get	Project	Get details about a specific network.
These are required for Cloudera to access the service accounts that you created.		
iam.serviceAccounts.actAs	Project	Allows you or a service account to impersonate a service account.

Granular IAM permissions	Scope	Description
iam.serviceAccounts.list	Project	List all the service accounts within a specific project.
(Optional) By default, Cloudera creates this bucket, but you can pre-create it. This is not required if you are planning to pre-create the GCS bucket for storing OS images for VMs. See Storage bucket for OS images.		
storage.buckets.get	Project	Get all the storage buckets from a specific project.
storage.objects.create	Project	Upload objects to the storage accounts.
storage.objects.delete	Project	Delete objects from the storage accounts.
storage.objects.get	Project	Retrieve objects from the storage account. This does not allow the listing of objects in the storage bucket.

Note: Additionally, once you create the Logger and IDBroker service accounts discussed in the minimum setup for cloud storage you need to update each of these two service accounts to grant the provisioning service account the Service Account User (iam.serviceAccountUser) role. See instructions provided as part of Minimum setup for cloud storage.

For instructions on how to create the service account, refer to the following documentation:

Create provisioning credential's service account and generate access key

Create a service account and generate a JSON access key.

Before you begin

Review the above permissions to learn what IAM permissions and IAM roles you need to assign to the service account that you will create.

Steps

1. Log in to your Google Cloud account.
2. Navigate to the project used for Cloudera.
3. Navigate to the IAM & Admin.
4. To create a custom role:
   a. Navigate to the Roles page.
   b. Click +Create Role.
   c. Specify a Title.
   d. Specify an ID.
   e. Click +Add Permissions.
   f. Add the required granular permission(s).
   g. Use the same steps to add all the required permissions.

      Note: If you are using the Option 1: IAM roles, you only need to assign the iam.serviceAccounts.list permission. If you are using the Option 2: Granular permissions, you need to assign all the permissions listed in the table listing the permissions.

   h. Click Create.
5. To create a service account:
   a. Navigate to the Service accounts page.
   b. Click Create service account.
   c. Enter a service account name.
   d. Click Create.
   e. Under Grant this service account access to project, choose the IAM roles to grant to the service account on the project. You need to assign all of the roles listed in the table.
   f. When you are done adding all the required roles, click Done to finish creating the service account.

6. To generate an access key:
   a. Once your account has been created, find the row of the service account that you want to create a key for. In that row, click the (context menu) button, and then click Create key.
   b. Under Key type, select JSON and click Create.
   c. Clicking Create downloads the service account key file. You will use the JSON access key to register the service account as a credential in Cloudera.

   Warning: After you download the key file, you cannot download it again.

7. Additionally, once you create the Logger and IDBroker service accounts, you need to update each of these two service accounts to grant the provisioning service account the Service Account User (iam.serviceAccountUser) role. The instructions are provided as part of Minimum setup for cloud storage.

What to do next

Once you have this setup ready, you can Register a GCP credential in Cloudera.

Storage buckets and service accounts for logs, backups, and data storage

Cloudera requires that you pre-create and provide buckets for logs and data storage and create service accounts controlling access to them.

You should create two Google storage buckets:

- One for data storage
- One for logs
- Optionally, you can also create a third bucket for storing FreeIPA and Data Lake backups

Note: It is possible to use a single bucket. If you choose to do so, you must adjust service account permissions accordingly. This scenario is not covered in this documentation.

The buckets should fulfill the following requirements:

- For best performance, create the buckets in the same region as the VPC.
- If you would like to use encryption, use a Google-managed key.

In addition to the two Google storage buckets, you should create multiple service accounts and assign roles as described in the following documentation:

Minimum setup for GCP cloud storage

The minimal setup recommended for production includes two GCS buckets (one for storing workload data and another for storing logs) and four service accounts. Additionally, you can create a third bucket for storing FreeIPA and Data Lake backup data separately. If the third bucket is not provided, FreeIPA and Data Lake backup data is stored in the Logs bucket.

Note: You may choose a different setup. For example, for getting started with a test environment you may want to use a single GCS bucket. Just note that such setup is not covered in this documentation.

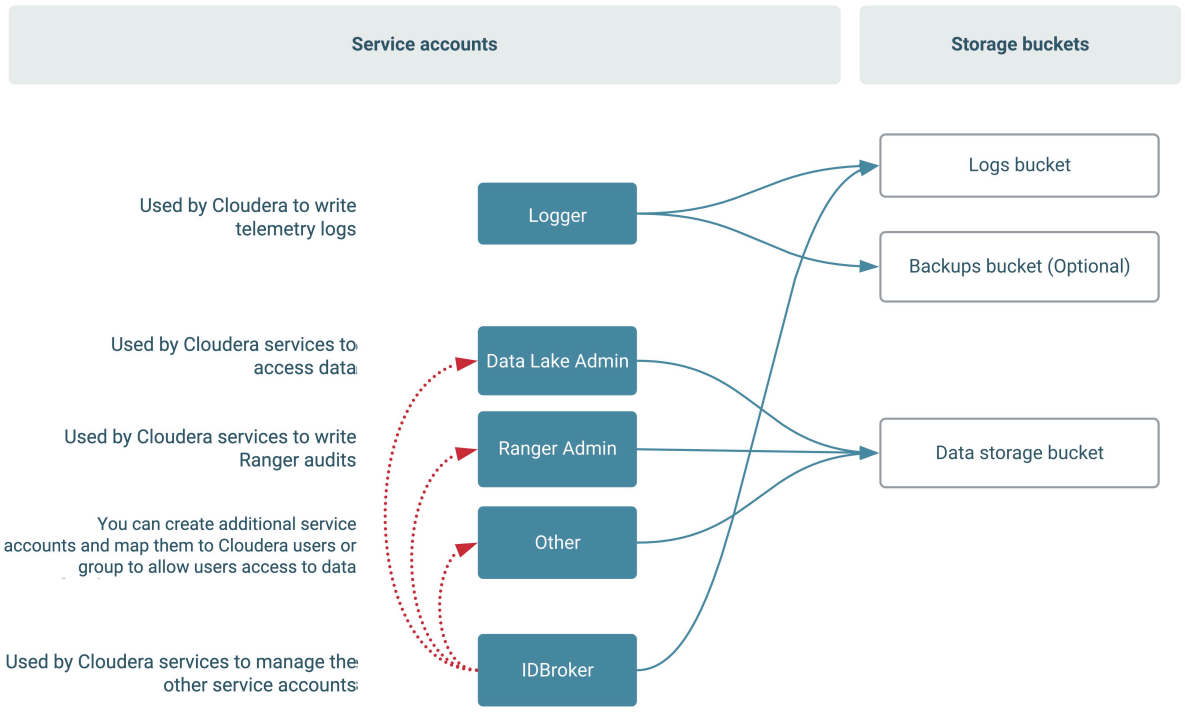
You need to create service accounts mentioned in the table and while creating them:

- The Service account name column lists all the service accounts that need to be created. You may choose different service account names. The ones provided here follow the same terminology as Cloudera web interface and CDP CLI making it easier to understand where to provide them to Cloudera.
- The Required IAM roles column explains what IAM role each service account needs over the item listed in the Scope column. For example, the Logger service account requires that you create a custom role with storage.buckets.get and storage.objects.create permissions. Next, you navigate to the Logs bucket permissions and add the Logger service account as a member with the custom role that you created earlier.

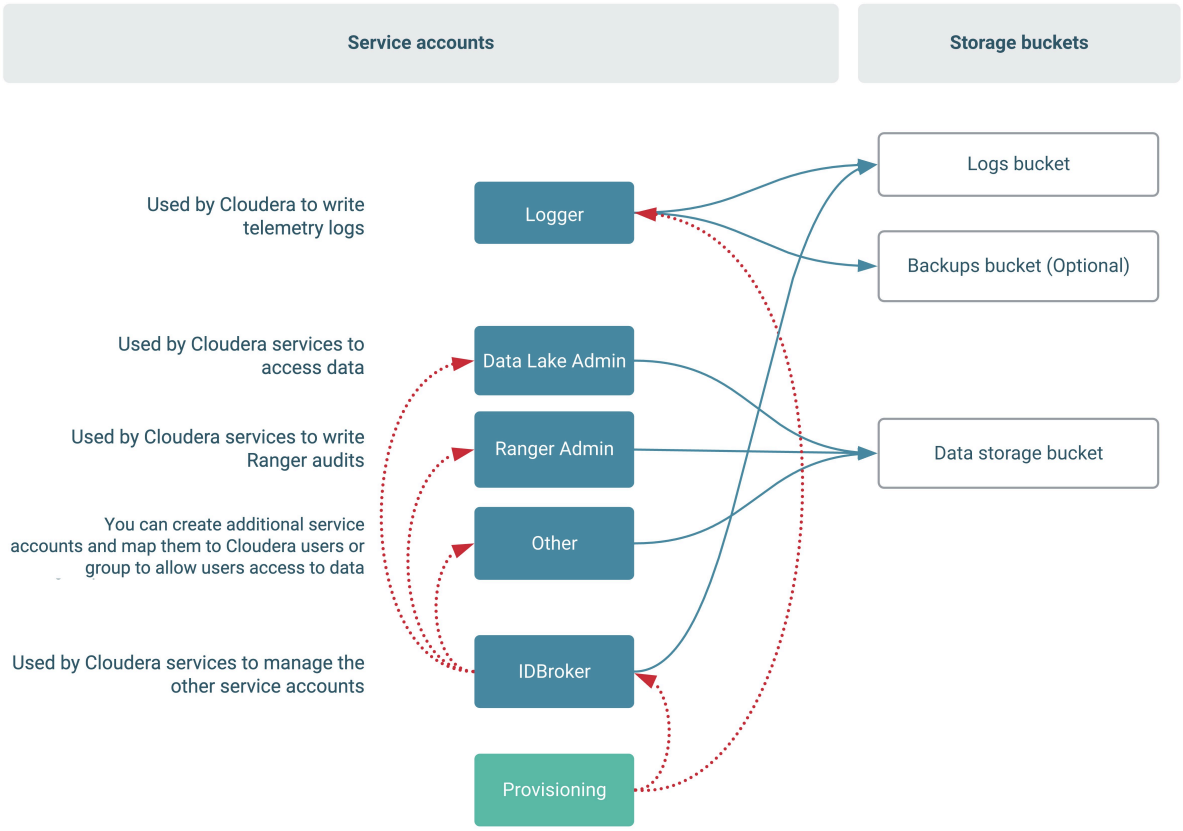
Service account name	Description	Required IAM roles	Scope
Logger	This service account will be assigned to all the workload instances in Cloudera. It will be used by Cloudera to: • Write telemetry logs to the Logs bucket. • Write FreeIPA backups to the Backups bucket or, if there is no designated bucket provided for backups, write to the Logs bucket.	A custom role with the following permissions: • storage.buckets.get • storage.objects.create • If you would like to use a bucket path (gs://<bucket>/<path>) instead of a bucket (gs://<bucket>) for the Logs or Backups bucket, you should also assign the storage.objects.list permission. When using lower versions than Cloudera Public Cloud 7.2.17, you need to create a custom role with the following permissions: • storage.buckets.get • storage.objects.create • storage.objects.get • storage.objects.list When using Cloudera Public Cloud 7.2.17 or higher versions, you need to create a custom role with the following permissions: • storage.buckets.get • storage.objects.create	Logs bucket and Backups bucket (if created)
Data Lake Admin	This service account will be used by Cloudera services to access workload data. It provides full access to the data storage location.	Storage Admin (roles/storage.admin) IAM role Alternatively, you can create a custom role and assign the following permissions: • storage.buckets.get • storage.objects.create • storage.objects.delete • storage.objects.get • storage.objects.list	Data storage bucket For Data Lake backup and restore: Backups bucket, if different from the main data storage bucket

Service account name	Description	Required IAM roles	Scope
Ranger Audit	This service account will be used by Cloudera to write Ranger audits to the storage bucket.	Storage Object Admin (roles/storage.objectAdmin) IAM role Alternatively, you can create a custom role and assign the following permissions: • storage.buckets.get • storage.objects.create • storage.objects.delete • storage.objects.get • storage.objects.list For Data Lake backup and restore, create a custom role and assign the following permissions: • storage.buckets.get • storage.objects.create • storage.objects.delete • storage.objects.get • storage.objects.list • storage.objects.getIamPolicy • storage.objects.setIamPolicy • storage.objects.update • resourcemanager.projects.get	Data storage bucket For Data Lake backup and restore: Backups bucket, if different from the main data storage bucket
Other service account for data access by users	Depending on your requirements, you may want to create a set of service accounts for data access by different user groups. For example, you may want to have one service account to assign to data science users and another service account for data engineering users.	Depending on your requirements, you should assign a custom role or a predefined role from Cloud Storage roles > Predefined roles on the bucket used for data storage.	Data storage bucket
IDBroker	This service account will be used by Cloudera to assume the other service accounts.	Workload Identity User (roles/iam.workloadIdentityUser) IAM role Alternatively, you can create a custom role and assign the following permissions: • iam.serviceAccounts.getAccessToken • iam.serviceAccounts.actAs	Service accounts (All of the above service accounts except Logger)
		Additionally, assign the same permissions as those assigned to the Logger service account.	Logs bucket

The following diagram illustrates the relationships between service accounts and buckets and between the IDBroker service account and other service accounts. The dotted arrows signify which entity needs access to what. For example, the Data Lake admin role must be able to access the Logs bucket:



In addition, Cloudera provisioning credential's service account (that you create as part of Create provisioning service account and generate access key) needs to have the Service Account User role to access to the Logger and IDBroker service accounts:



You need to perform the following high-level steps in order to create the required resources:

1. You should create the required Logs and Data storage GCS buckets. You can also create a separate Backups bucket.
2. Create the required service accounts.
3. Create the required custom roles.
4. Add service accounts as members to the Logs and Data Storage buckets.
5. Add the IDBroker service account as a member to other service accounts.
6. Add the provisioning service account as a service account user to the Logger and IDBroker service accounts.
7. Once you have met all of the GCS prerequisites, you can register a GCP environment in Cloudera.

The instructions for performing these steps are mentioned below.

Create the GCS buckets

Use these steps to create the two required GCS buckets.

Steps

1. In Google Cloud console, navigate to Cloud Storage > Browser.
2. Click on +Create bucket.
3. Name your bucket.
4. Click Create.

Repeat these steps for both buckets. Note the bucket names. You will need to provide them to Cloudera later.

For more information, see GCP docs linked below.

Related Information

Creating storage buckets

Create the service accounts

Use these steps to create the required service account.

Steps

1. In Google Cloud console, navigate to the project used for Cloudera.
2. Navigate to IAM > Service accounts.
3. Click on +Create service account.
4. Provide a name.
5. Click Create.

Repeat these steps for all service accounts. Copy and save the email addresses identifying the created service accounts. You will need to provide them to Cloudera. Service account naming convention is <service-account-name>@<project-id>.iam.gserviceaccount.com.

For more information, see GCP docs linked below.

Related Information

Creating a service account

Create the custom role for the Logger

Use these steps to create the custom role for the Logger service account.

Steps

1. In Google Cloud console, in the same project, navigate to IAM > Roles
2. Click on +Create Role
3. Enter a name
4. Click +Add permissions and add all the permissions mentioned in the respective entry in the above table.
5. When done adding permissions, click Create.

For more information, see GCP docs linked below.

Related Information

Creating a custom role

(Optional) Create other custom roles

If you would like to create custom roles for other service accounts, follow the same instructions as above. You only need to do this if you don't want to use the predefined roles listed in the table.

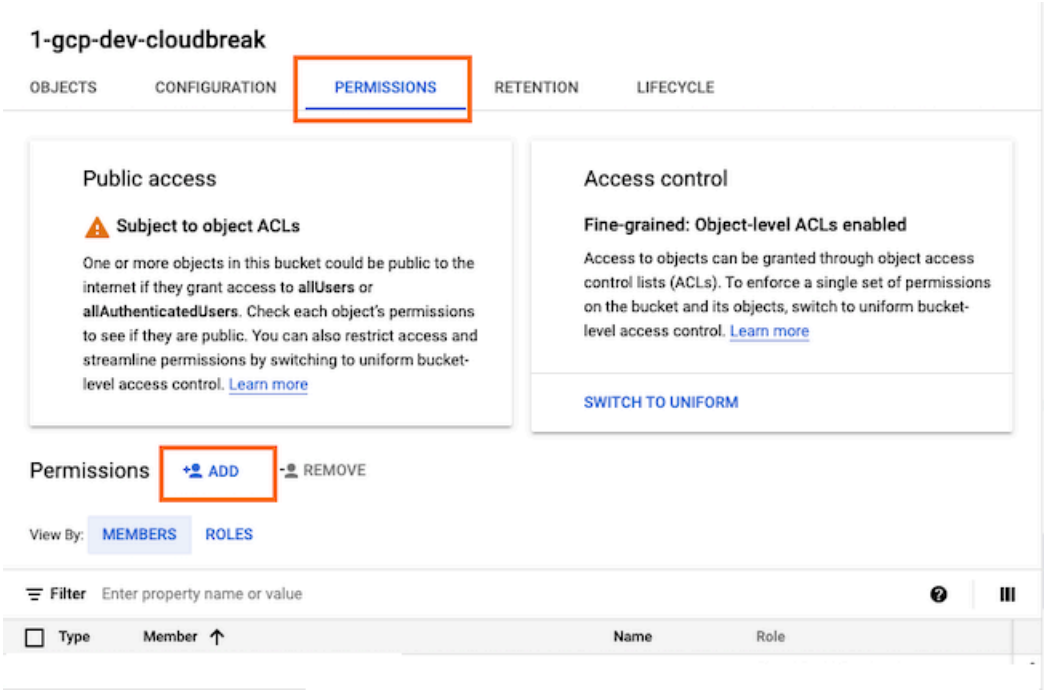
Add service accounts as members to buckets

Use the following steps to a add service account as a member to a bucket.

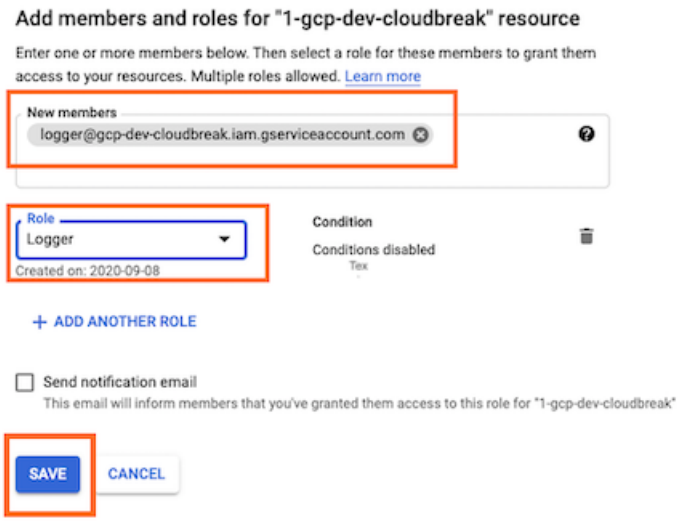
Steps

1. In Google Cloud console, in the same project, navigate to Cloud Storage > Browser.

2. Perform membership and role assignment for the Logs bucket:
   a. Find your Logs bucket.
   b. Click on ⋮ > Edit bucket permissions or double-click on the bucket and then click on the Permissions tab.
   c. Next to Permissions, click +Add:



   d. Under New members, select the Logger service account and the IDBroker service account. Under Role, select the custom role created earlier (in the screenshot the role is called Logger):



   e. Click Save.
3. If you created the separate Backups bucket for FreeIPA backups, repeat the above steps for the Backups bucket. Use the same Logger service account, IDBroker service account and the custom Logger role created earlier.
4. Repeat the above steps multiple times for the Data storage bucket. You need to add the Data Lake Admin, Ranger Audit, and other service accounts (if created) as members of the Data storage bucket and assign the respective roles mentioned in the above table. Each of these role assignments requires a separate set of steps, so you need to repeat the steps as many times as you have service accounts.

For more information, see GCP docs linked below.

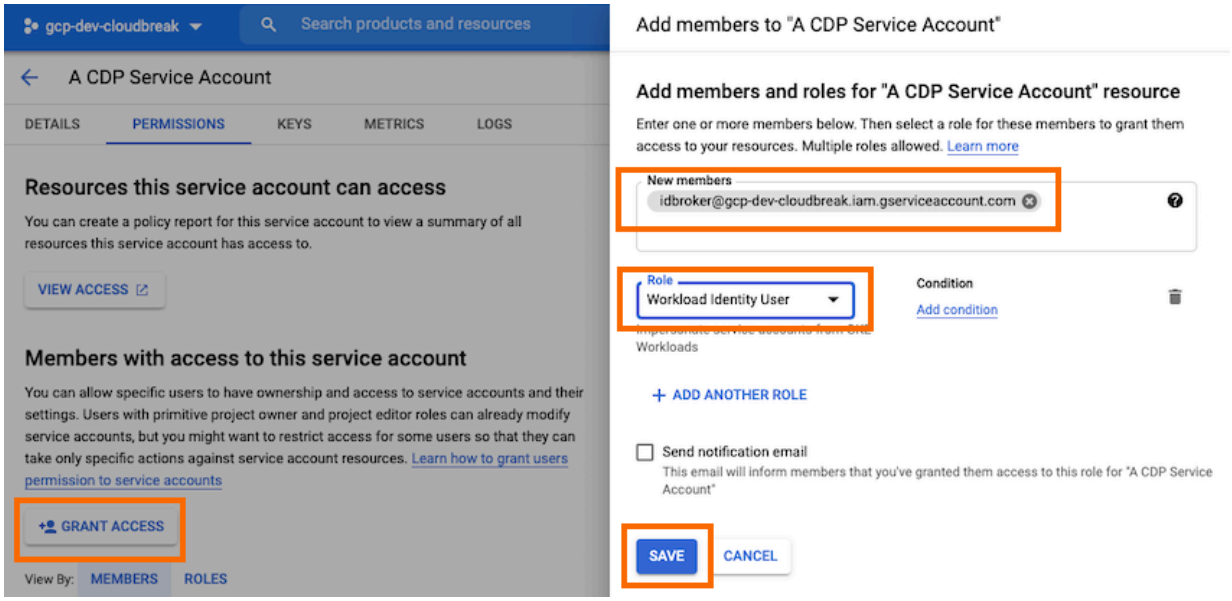
Related Information
Adding a member to a bucket-level policy

Add IDBroker as a member to other service accounts
Similarly as with the buckets, you need to navigate to service account permissions and add the IDBroker service account as a member with the specified role.

Steps

You need to do this for all the service accounts except Logger (and except IDBroker itself):

1. In Google Cloud console, in the same project, navigate to IAM > Service accounts.
2. Double-click on the service account entry.
3. Navigate to the Permissions tab.
4. Click Grant Access and then add the IDBroker service account as a member with the role specified in the above table:



5. Click Save.

Repeat these steps for all service accounts except Logger.

For more information, see GCP docs linked below.

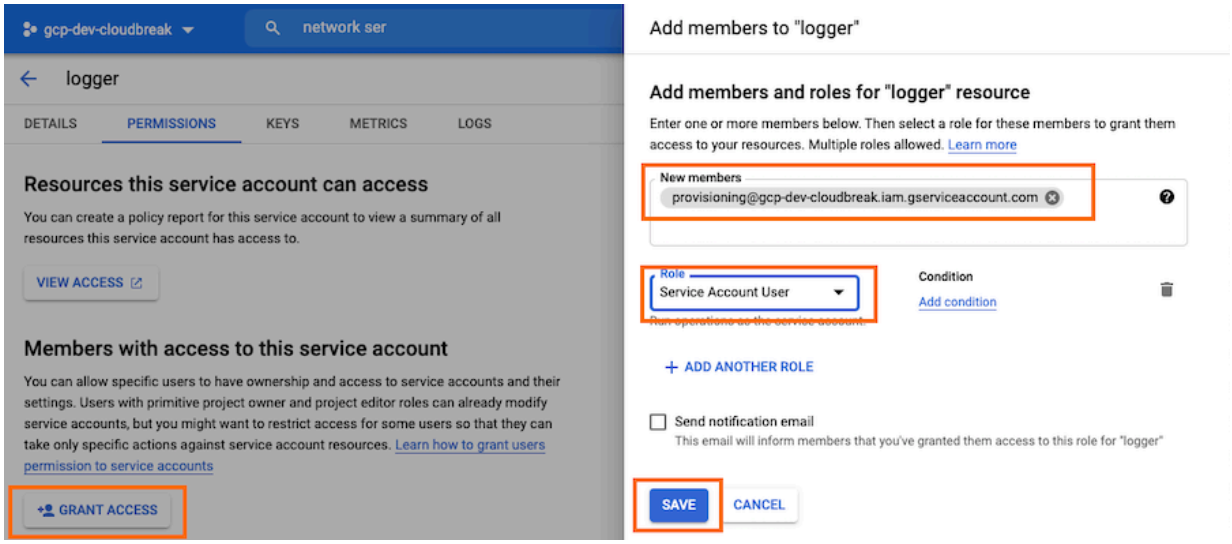
Related Information
Allowing a member to impersonate a single service account

Add provisioning service account as a service account user
To complete the setup, you need to update the permissions of the Logger and IDBroker service accounts, granting the provisioning service account the Service Account User role.

Steps

1. In GCP IAM console, navigate to Service Accounts.
2. Find your Logger service account.
3. Click Manage Permissions to access the Permissions tab.

4. Click Grant Access and then add the provisioning service account as a member with role Service Account User:



5. Click Save.
6. Repeat the steps for the IDBroker service account.

Providing the parameters in Cloudera

Once you've created the bucket and instance profiles, provide the information related to these resources in the Register Environment wizard as follows:

Data Access and Data Lake Scaling > Data Access:

UI parameter	What to provide
Assumer Service Account	Select the IDBroker service account created earlier.
Storage Location Base	Enter the name of your Data storage bucket created earlier for the storage location base.
Data Access Service Account	Select the Data Lake Admin service account created earlier.
Ranger Audit Service Account	Enter the email address for the Ranger Audit service account created earlier. The service account email address uses the following format: <service-account-name>@<project-id>.iam.gserviceaccount.com.
Backup Location Base (Optional)	If you created it, enter the name of your Backups bucket for storing FreeIPA and Data Lake backups. This is optional. If you don't provide this, FreeIPA and Data Lake backups will be stored in the Logger bucket.

Storage and Audit page > Logs

UI parameter	What to provide
Logger Service Account - Logger	Select the Logger service account created earlier.
Logs Location Base	Enter the name of your Logger bucket created earlier for logs location base.

Data Access and Data Lake Scaling > Mappings

You can use this section to set service account to Cloudera user/group mappings for the additional service accounts created for user access to data. Or you can do this once your environment is running, as part of Onboarding Cloudera users and groups for cloud storage.

Onboarding Cloudera users and groups for GCP cloud storage

The minimal setup defined earlier spins up a Cloudera environment and Data Lake with no end user access to cloud storage. Adding users and groups to a Cloudera environment involves ensuring they are properly mapped to service accounts to access cloud storage.

In general, to have new users or groups onboarded, you need to do the following:

1. Create a new service account and assign appropriate IAMroles or granular permissions on the scope of the Storage Location Base or its specific sub-directory. You might have already performed this step earlier during setting up the Minimum setup for cloud storage.
2. In order to use these storage accounts in Cloudera, create a user/group to service account mapping in Cloudera.

This needs to be done for each user type. For example, you can create two service accounts in GCP, one for Data Scientists and another for Data Engineers, and then you map each of them to a group of users in Cloudera.

The onboarding of users can either happen as part of environment registration, or you can do it once an environment is running. The steps below show you how to onboard users once an environment is running.

Creating Cloudera user/group to service account mappings

After creating the additional service accounts, map each of them to a specific user or group.

Before you begin

The steps below show how to add the mappings to an existing environment. Alternatively, you can add them during environment registration, as mentioned in the Minimum setup for cloud storage.

Note: If a user is mapped to multiple roles via group membership, the specific role to be used needs to be provided at runtime. If the user is mapped directly to a role, the direct mapping takes precedence over mapping via group membership. For information on how to specify the role, refer to Specifying a group when user belongs to multiple groups.

Required role: DataSteward, EnvironmentAdmin, or Owner

Steps

For Cloudera UI

1. The option to add/modify the service account to user/group mappings is available from the Cloudera Management Console under Environments > click on an environment > Actions > Manage Access > IDBroker Mappings.
2. Under Current Mappings, click Edit.
3. Click + to display a new field for adding a mapping.
4. Provide the following:
   a. The User or Group dropdown is pre-populated with Cloudera users and groups. Select the user or group that you would like to map.
   b. Under Role, specify the resource ID of a service account (copied from Google Cloud IAM). For example “datascientists@gcp-cdpdev.iam.gserviceaccount.com”.
5. Repeat the previous two steps if you would like to add additional mappings.
6. Click Save and Sync.

For example, in the example setup we created the following roles:

- DATAENG_ROLE - We created this role while onboarding users and we assume that there is a DataEngineers group that was created in Cloudera.
- DATASCI_ROLE - We created this role while onboarding users and we assume that there is a DataScientists group that was created in Cloudera.

For CDP CLI

If you would like to create the mappings via CDP CLI, you can:

1. Use the cdp environments get-id-broker-mappings command to obtain your current mappings.

2. Use the cdp environments set-id-broker-mappings command to set additional mappings. The only way to use this command is to:
   a. Pass all the current mappings
   b. Add the new mappings.
3. Next, sync IDBroker mappings to the environment:

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

4. Finally, check the sync status:

```
cdp environments get-id-broker-mappings-sync-status --environment-name d
emo3
```

Storage bucket for OS images

By default, Cloudera creates a GCS bucket for storing OS images used for Data Lake and Cloudera Data Hub VMs, but you can optionally pre-create it if your organization requires it.

- It must have a name using the following convention: <ClouderatenantID>-<ProjectID>

You can your Cloudera tenant ID as described in Obtain Cloudera tenant ID.

Supported GCP block storage

Review the list of supported GCP disk types for Cloudera.

Cloudera supports the following GCP disk types:

Cloudera name	GCP name
pd-ssd	SSD Persistent Disk
pd-balanced	Balanced Persistent Disk
pd-extreme	Extreme Persistent Disk
pd-standard	Standard Persistent Disk
local-ssd	Local SSDs

Related Information

Google Cloud storage options

SSH key pair

When registering an environment, you will be asked to provide a BD FIPS SSH key pair for admin access to Cloudera. The minimum SSH key size is 4096 bits.

You will need to paste the public SSH key in Cloudera during environment registration. If you need help generating an SSH key, refer to https://www.ssh.com/ssh/keygen/.

Customer managed encryption keys

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

When a CMEK is provided during environment registration, all the Data Lake, FreeIPA, and Cloudera Data Hub disks and the Cloud SQL instances are encrypted using that key.

To set up a CMEK, perform the following tasks:

1. Review the CMEK requirements.
2. Create a key ring and an encryption key.
3. Assign the required permissions to the encryption key.

This document guides you through all the required steps performed using the GCP console and Google Cloud Shell. Once you've met the prerequisites, pass the encryption key when creating a Cloudera environment via Cloudera web interface or Cloudera CLI.

For general information about customer managed encryption keys, see Customer-managed encryption keys (CMEK).

Cloud HSM and Hosted Private HSM encryption

Cloudera supports encryption using Cloud HSM encryption keys. The overall requirements and steps are the same as usual, just when you create the encryption key, you need to select the protection level to be "HSM". The instructions provided here consider the scenario of a Cloud HSM encryption key.

Cloudera also supports encryption using encryption keys from Hosted Private HSM. If you wish to use Hosted Private HSM, the GCP support helps you configure the setup and then provides you with the encryption key ARN. Cloudera can then use the key ARN.

For more information, see Cloud HSM and Hosted Private HSM.

CMEK requirements

Ensure that the CMEK that you are planning to use for encrypting your Cloudera environment meets the requirements.

The CMEK must meet the following requirements:

- CMEK needs to be in the same region as the environment.
- The key should have the following permissions for the compute and cloud sql service agents:

  Cloud KMS CryptoKey Encrypter/Decrypter

The instructions below show you how to create a CMEK that meets these requirements.

Create key ring and encryption key

Use the following instructions to create a key ring and an encryption key in Google Cloud.

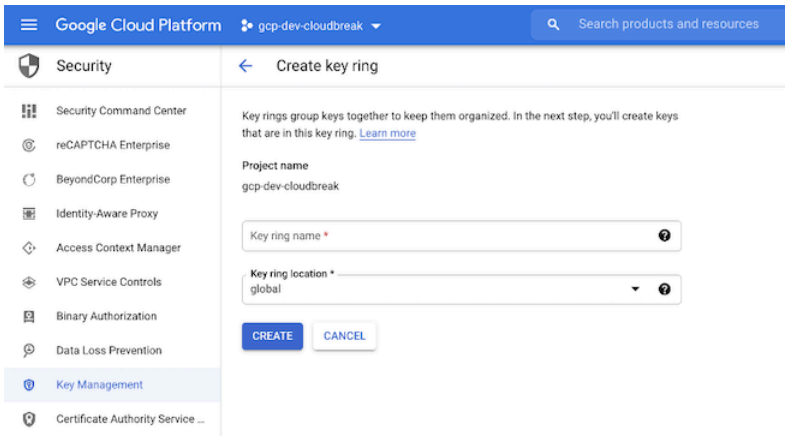
Note:

Key rotation and storage are Google-managed.

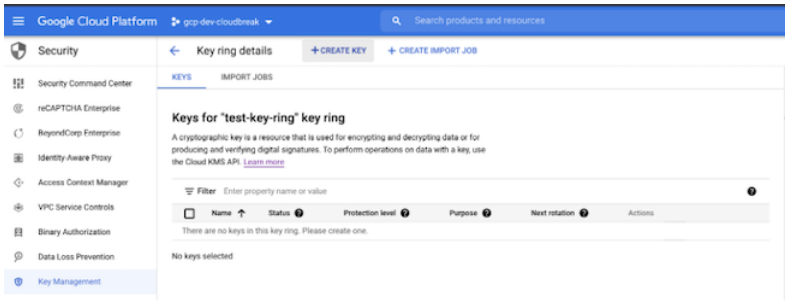
Steps

1. In the GCP console, navigate to Security > Key Management.

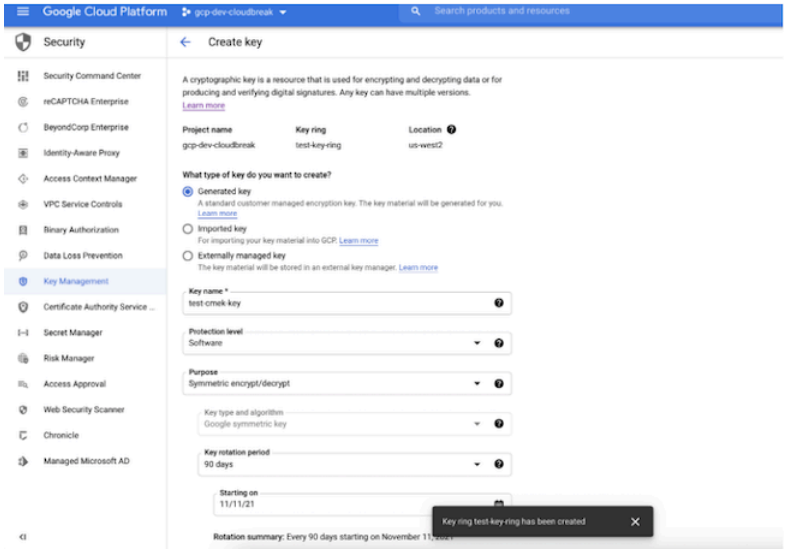
2. Create a key ring or use any existing one. Ensure that the key ring location and the location of the resources you create for the Cloudera environment are the same.



3. Navigate to the key ring that you have previously created.
4. Click on +Create Key to create a new key inside the key ring.



5. Under What type of key do you want to create?, select Generated key.
6. Under Key name, enter the name for your key.
7. From the Protection level dropdown:
   a. If you are using a standard CMEK, select Software.
   b. If you are using a Cloud HSM key, select HSM.
8. From the Purpose dropdown, select Symmetric encrypt/decrypt.
9. Use the default values for Rotation period and Starting on.
10. Click Create.



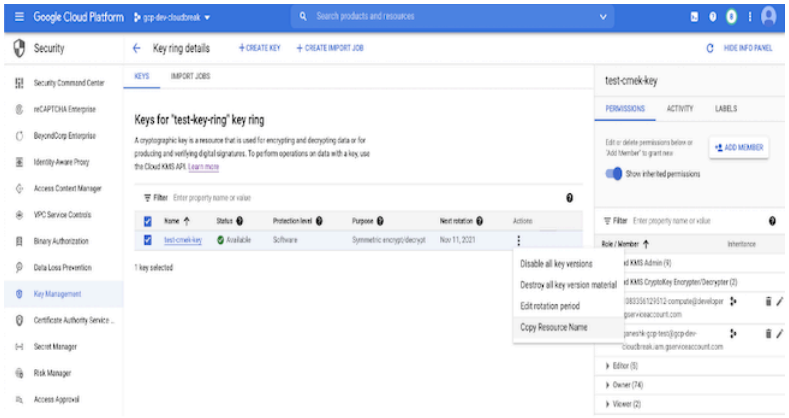
Assign the required permissions to the encryption key

Once the key has been created, you need to assign the required permissions to it. The following commands can be used to set it up using Google Cloud Shell.

Prerequisites

Make sure that you have the following available:

- The project number of the Google project number where the compute and SQL resources are created (PROJECT_NUMBER).
- Copy the key ring resource name (KEYRING_RESOURCE_NAME) and the key resource name (KEY_RESOURCE_NAME) from the Google cloud console. You can copy the key ring resource name from the dropdown after clicking three vertical dots next to the key ring. You can copy the key resource name in a similar manner.



Steps

1. If the cloud sql service agent does not exist in the project, create it using:

```
gcloud beta services identity create --service=sqladmin.googleapis.com --
project=<project_name>
```

   This command creates a service identity in the following format:

   service-108335612.5..@gcp-sa-cloud-sql.iam.gserviceaccount.com

   where "108335612.5.." is the PROJECT_NUMBER to be used in step 2 and 3.
2. Assign the IAM policy to encrypt and decrypt KMS keys. Replace the variables in caps with the values obtained earlier:

```
gcloud kms keys add-iam-policy-binding KEY_RESOURCE_NAME \
 --location=GCP_REGION \
 --keyring=KEYRING_RESOURCE_NAME \
 --member=serviceAccount:service-PROJECT_NUMBER@gcp-sa-cloud-sql.iam.gser
viceaccount.com \
 --role=roles/cloudkms.cryptoKeyEncrypterDecrypter
```

3. Assign the IAM policy to the compute service agent. Replace the variables in caps with the values obtained earlier:

```
gcloud kms keys add-iam-policy-binding KEY_RESOURCE_NAME \
 --location=GCP_REGION \
 --keyring=KEYRING_RESOURCE_NAME \
 --member=serviceAccount:service-PROJECT_NUMBER@compute-system.iam.gservi
ceaccount.com \
 --role=roles/cloudkms.cryptoKeyEncrypterDecrypter
```

GCP limits

When you create your Google Cloud project, Google sets limits to the resources available to you. In some cases, the limits are insufficient for Cloudera and you need to request to have them increased.

For example, in a newly created project, you may need to increase the limits for:

- Public IPs
- Virtual CPU (vCPU)
- Disk storage

For a full list of resources that Cloudera provisions, refer to GCP resources used by Cloudera.

For information on checking and increasing your resources quotas, refer to Resource quotas.

Overview of GCP resources used by Cloudera

The following Google Cloud resources are used by Cloudera and Cloudera services.

GCP resources created for a Cloudera environment

When a Cloudera environment is created, a FreeIPA cluster and a Data Lake cluster are created.

The following Google Cloud resources are created for FreeIPA (one per environment):

Resource	Description
Service account for credential	To allow Cloudera to access and provision resources in your Google Cloud project, you must create a service account in your Google Cloud project, assign required roles, and generate a JSON access key that can later be provided to Cloudera.
VPC network and subnets	During environment creation you provide your own existing VPC network and subnets. All compute resources that Cloudera provisions for the environment and Cloudera services are provisioned into the VPC network specified during environment creation.
Firewall rules	Firewall rules define inbound and outbound access to the instances. If during environment creation you choose to have new firewall rules created, then they are created on your GCP account. Alternatively, you can provide your own existing firewall rules.
VM instances	During environment creation, two or three e2-standard-2 VM instances are provisioned for the FreeIPA HA server. The number of VMs depends on the selected Data Lake type.
OS disk	An OS disk is provisioned for the FreeIPA VM.
Attached disk	An attached disk (pd-standard) is provisioned for each VM.
Public IP address (if required)	If you choose to use public IPs, your VM is assigned a public IP address.
GCS bucket for storing operating system images	By default, Cloudera creates a storage bucket that is used solely for storing operating system images. If required, you can optionally pre-create this account and copy the required images.

In addition, the following resources are created for each Data Lake (one per environment):

Resource	Description
VM instances	VM instances are provisioned for the Data Lake nodes. • Light duty: Two VM instances are provisioned: One e2-standard-2 VM instance (for IDBroker) and one e2-standard-8 VM instance (for master) are created. • Medium duty: Ten VM instances are provisioned: Two e2-standard-2 (IDBroker), three e2-standard-4 (two Data Lake Master nodes and one Auxiliary node), and five e2-standard-8 (three DataLake Core nodes and two Gateway nodes).
Attached disk	An attached disk (pd-standard) is provisioned for each VM.
OS disk	An OS disk is provisioned for each VM.
PostgreSQL database	A custom PostgreSQL database instance (100GB SSD, 2vCPU, 13 GB RAM) is provisioned for the Data Lake. This databse instance is used for Cloudera Manager, Ranger, and Hive MetaStore.
Firewall rules	Firewall rules define inbound and outbound access to VM instances. If during environment creation you choose to have new firewall rules created, then they are created on your GCP project.
Google storage buckets	The existing Google Storage bucket that you provide during environment creation for the Data Lake is used for Data Lake log storage and workload data storage.
Service accounts	Prior to registering your environment in Cloudera, during Google storage setup, you should create service accounts and assign roles to them as instructed in Cloudera documentation.
Public IP address (if required)	If you choose to use public IPs, your VM is assigned a public IP address.

GCP resources created for Cloudera Data Hub

The following Google Cloud resources are created for each Cloudera Data Hub cluster:

Resource	Description
VM instances and attached storage	A VM is created for each cluster node. The VM type varies depending on what you selected during Cloudera Data Hub cluster creation. For a list of supported VM types, refer to Cloudera on cloud service rates.
Firewall rules	Firewall rules define inbound and outbound access to VM instances. If during environment creation you choose to have new firewall rules created, then they are created on your GCP project.
OS disk	An OS disk is provisioned for each VM.
Attached disk	An attached disk (pd-standard) is provisioned for each VM, as specified during Cloudera Data Hub cluster creation. The disk size is selected during cluster creation.
Public IP address (if required)	If you choose to use public IPs, each of the VMs is assigned a public IP address.

GCP outbound network access destinations

If you have limited outbound internet access (for example due to using a firewall or proxy), review this content to learn which specific outbound destinations must be available in order to register a Cloudera environment.

We recommend hostname-based policies, as some of the destination services do not have static IP addresses. IP address details in CIDR notation have been provided where static IPs are in-use.

Warning: On 30 September 2025, Microsoft will retire the default outbound access connectivity for virtual machines in Azure. You must configure an explicit outbound connectivity method. For more information, see Cloudera Customer Advisory-866 and Transitioning from Azure default outbound access documentation.

Note:

If the cloud provider network that you would like to use for registering a Cloudera environment uses a custom DNS server that does not allow name resolution for public domain, you should add all the domains listed in the below tables to the DNS forwarder for name resolution.

The following list includes general destinations as well as GCP-specific destinations.

General endpoints

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Control Plane API	All services	US-based Control Plane: api.us-west-1.cdp.cloudera.com EU-based Control Plane: api.eu-1.cdp.cloudera.com AP-based Control Plane: api.ap-1.cdp.cloudera.com	HTTPS with Cloudera-generated access key	TCP/443	Cloudera Control Plane REST API.
Cloudera CCMv1 Persistent Control Plane connection	All services	*.ccm.cdp.cloudera.com 44.234.52.96/27	SSH public/ private key authentication	TCP/6000-6049	One connection per cluster configured; persistent
Cloudera CCMv2 Persistent Control Plane connection	All services	US-based Control Plane: *.v2.us-west-1.ccm.cdp.cloudera.com 35.80.24.128/27 EU-based Control Plane: *.v2.ccm.eu-1.cdp.cloudera.com 3.65.246.128/27 AP-based Control Plane: *.v2.ccm.ap-1.cdp.cloudera.com 3.26.127.64/27	HTTPS with mutual authentication	TCP/443	Multiple long-lived/persistent connections

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Cloudera Databus Telemetry, billing and metering data	All services	US-based Control Plane: dbusapi.us-west-1.sigma.altus.cloudera.com api.us-west-1.cdp.cloudera.com https://cloudera-dbus-prod.s3.amazonaws.com EU-based Control Plane: api.eu-1.cdp.cloudera.com https://mow-prod-eu-central-1-sigmadbus-dbus.s3.eu-central-1.amazonaws.com https://mow-prod-eu-central-1-sigmadbus-dbus.s3.amazonaws.com AP-based Control Plane: api.ap-1.cdp.cloudera.com https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.ap-southeast-2.amazonaws.com https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.amazonaws.com	HTTPS with Cloudera-generated access key for dbus HTTPS for S3	TCP/443	Regular interval for telemetry, billing, metering services, and used for Cloudera Observability if enabled. Larger payloads are sent to a Cloudera managed S3 bucket.
Cloudera Observability Metrics System metrics collection	All services	US-based Control Plane: *.api.monitoring.us-west-1.cdp.cloudera.com EU-based Control Plane: *.api.monitoring.eu-1.cdp.cloudera.com AP-based Control Plane: *.api.monitoring.ap-1.cdp.cloudera.com	HTTPS	TCP/443	New as of March 2024
Cloudera Manager parcels Software distribution	All services	archive.cloudera.com	HTTPS	TCP/443	Cloudera's public software repository. CDN backed service; IP range not predictable.
RPMs Cloudera RPMs for workload agents	All services	cloudera-service-delivery-cache.s3.amazonaws.com	HTTPS	TPC/443	RPM packages for some workload components
AI Agent Studios	Cloudera AI	huggingface.co github.com github.infra.cloudera.com nodejs.org iojs.org container.repo.cloudera.com	HTTPS	TCP/443	Required for AI Agent Studios
Cloudera Copilot	Cloudera AI	bedrock-runtime.<region>.amazonaws.com api.gradio.app	HTTPS	TCP/443	Required for Cloudera Copilot AMP when deploying an LLM on AWS Bedrock.

GCP-specific endpoints

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
APIs	All services	storage.googleapis.com iamcredentials.googleapis.com	HTTPS	TCP/443	In addition to adding the listed destinations, you need to configure Private Service Connect. Private Service Connect lets you send traffic to Google APIs using a Private Service Connect endpoint that is private to your VPC network. To configure Private Service Connect, refer to Configuring Private Service Connect. **Note:** This is not optional. If you don't configure this, environment registration will fail.

Access to workload UIs

If you have restricted DNS or networking setup, make sure that *.cloudera.site is resolvable from your network so that members of your organization can access workload UIs.

Cloudera workloads (including Data Lake) use subdomains under cloudera.site to host various UI endpoints (Cloudera Manager, Ranger, Knox, Hue and so on). Cloudera automatically provisions these endpoints whenever a Data Lake, Cloudera Data Hub or another type of workload (for example, Virtual Warehouse in Cloudera Data Warehouse) is created, and routing is set up so that you can access these endpoints from your network.

The subdomains are assigned under cloudera.site using the following convention:

```
<endpoint-name>.<env-truncated-name>.<customer-workload-subdomain>.<regional
-subdomain>.cloudera.site
```

Supported browsers

Cloudera validates and tests against the latest version and supports recent versions of the following browsers:

- Google Chrome

- Mozilla Firefox

  Note:

  - Mozilla Firefox is not supported by Cloudera Data Engineering.
  - Certain accessibility features in Cloudera DataFlow do not work in Mozilla Firefox.
- Safari
- Microsoft Edge

Cloudera CIDR

List of Cloudera Control Plane source IP addresses

In cases where the Cloudera Control Plane is communicating directly with resources in the Cloudera environment, the Cloudera Control Plane source IP address will be from the ranges below.

Security groups deployed/managed by Cloudera Control Plane will automatically include the required rules. These source IP ranges will be included if the Cloudera Control Plane requires direct inbound access to the resource.

Cloudera Control Plane Region	IP Ranges
us-west-1	35.80.24.128/27, 35.166.86.177/32, 52.36.110.208/32, 52.40.165.49/32
eu-1	3.65.246.128/27
ap-1	3.26.127.64/27

If the security groups (or other network access policies) in the network path are not managed by the Cloudera Control Plane, ensure that inbound access from the appropriate range(s) are permitted.