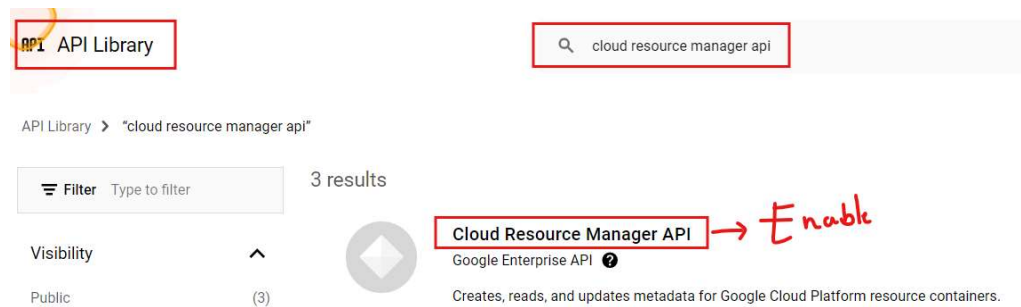


Pre-requisites

Wednesday, October 2, 2024 9:30 PM

Steps to create cluster in Dataproc :

- Make sure Compute Engine API enabled
- Make sure Dataproc API enabled
- Enable Cloud Resource Manager API



○ Assign Storage Admin Role to Default Service account

VIEW BY PRINCIPALS

VIEW BY ROLES

+ GRANT ACCESS

- REMOVE ACCESS

Filter Enter property name or value

<input type="checkbox"/>	Type	Principal ↑	Name	Role	Security insights ?
<input type="checkbox"/>		897233386692-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor	
				Storage Admin	
<input type="checkbox"/>		gcpdemopractice13@gmail.com	gcpdemopractice13	Owner	

○ VPC networks ==> default ==> subnets ==> default (us-east1)

← Subnet details EDIT DELETE

default

VPC Network

[default](#)

Region

us-east1

IP stack type

IPv4 (single-stack)

Primary IPv4 range

Primary IPv4 range	Access type	Reserved internal range
10.142.0.0/20	Internal	None

Gateway

10.142.0.1

Private Google Access

On

○ Configuration for both master and worker Nodes

✓ General purpose

Compute-optimised

Memory-optimised

GPUs

Machine types for common workloads, optimised for cost and flexibility


Series

N2

Powered by Intel Cascade Lake and Ice Lake CPU platforms

Machine type

n2-standard-2 (2 vCPU, 1 core, 8 GB memory)



vCPU

2

Memory

8 GB

✓ CPU PLATFORM AND GPU

Primary disk size *

50

GB

?

Primary disk type *

Balanced Persistent Disk

?

Number of local SSDs

▼ x 375GB

?

Local SSD interface

▼

?

○ Possible Errors while Creating Cluster :

The screenshot shows the 'Cluster details' page for a cluster named 'my-cluster'. At the top, there are navigation buttons: SUBMIT JOB, REFRESH, START, STOP, DELETE, and VIEW LOGS. Below these, an error message is displayed in a grey box with a red icon: 'Failed to validate permissions required for default service account: 774436673950-compute@developer.gserviceaccount.com'. The message explains that cluster creation could still be successful if required permissions have been granted to the respective service accounts as mentioned in the document https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/service-accounts#dataproc_service_accounts_2. It suggests that this could be due to the Cloud Resource Manager API not being enabled in the project '774436673950' before or if it is disabled, and provides a link to enable it: <https://console.developers.google.com/apis/api/cloudresourcemanager.googleapis.com/overview?project=774436673950>. A 'MORE' link is also present.

Name	my-cluster
Cluster UUID	8c65d11e-6099-4f32-8eb0-7a188d5157d1
Type	Dataproc Cluster
Status	Error

Below the table, there are tabs for MONITORING, JOBS, VM INSTANCES, CONFIGURATION, and WEB INTERFACES. Under the MONITORING tab, there are buttons for SAVE AS DASHBOARD and RESET ZOOM. A time range selector is set to '1 hour'. The main content area shows two graphs: 'YARN memory' and 'YARN pending memory', both with zoom and refresh icons.

Multiple validation errors: **Permissions are missing for the default service account** '774436673950-compute@developer.gserviceaccount.com', missing permissions: [storage.buckets.get, storage.objects.create, storage.objects.delete, storage.objects.get, storage.objects.list, storage.objects.update] on the project 'projects/caramel-medley-435805-c7'. This usually happens when a custom resource (ex: custom staging bucket) or a user-managed VM Service account has been provided and the default/user-managed service account hasn't been granted enough permissions on the resource. See https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/service-accounts#VM_service_account. - Permissions are missing for the default service account '774436673950-compute@developer.gserviceaccount.com', missing permissions: [storage.objects.get, storage.objects.update] on the staging_bucket 'projects/_/buckets/dataproc-staging-us-central1-774436673950-mzfxquq2'. This usually happens when a custom resource (ex: custom staging bucket) or a user-managed VM Service account has been provided and the default/user-managed service account hasn't been granted enough permissions on the resource. See https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/service-accounts#VM_service_account.

Request ID: 4492064043386838551

Cluster warnings

- Subnetwork 'default' does not support Private Google Access which is required for Dataproc clusters when 'internal_ip_only' is set to 'true'. Enable Private Google Access on subnetwork 'default' or set 'internal_ip_only' to 'false'.
- The firewall rules for specified network or subnetwork would allow ingress

traffic from 0.0.0.0/0, which could be a security risk.