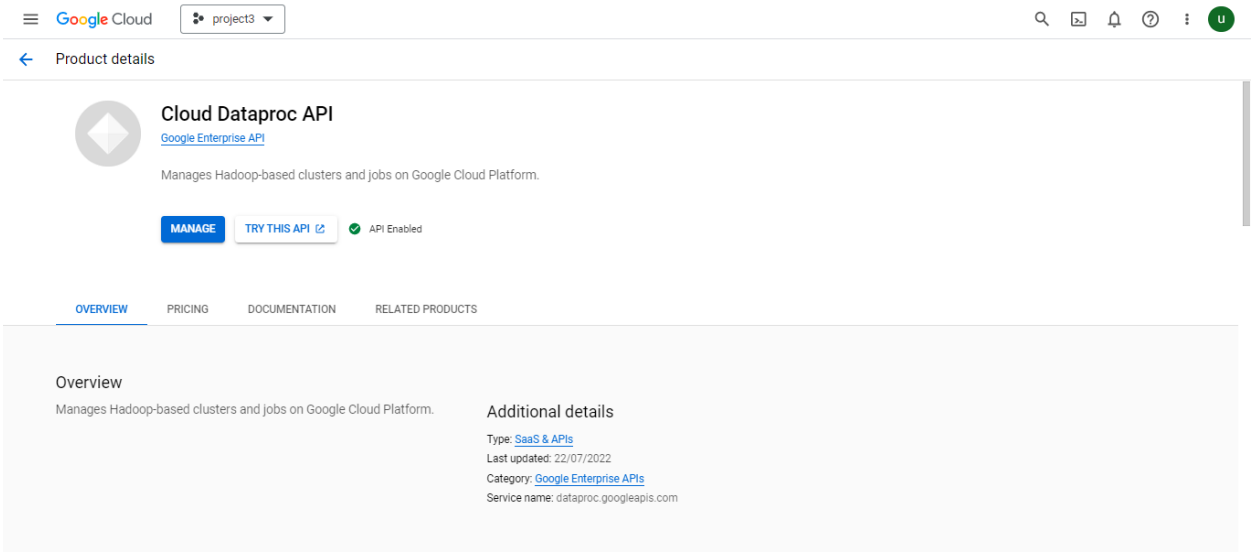
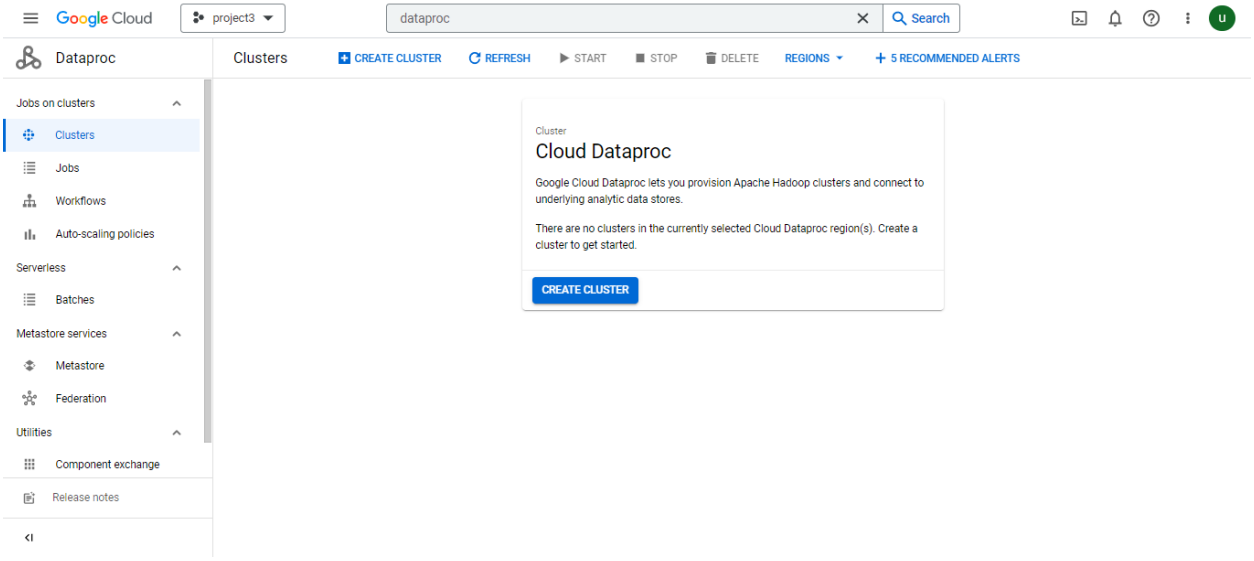
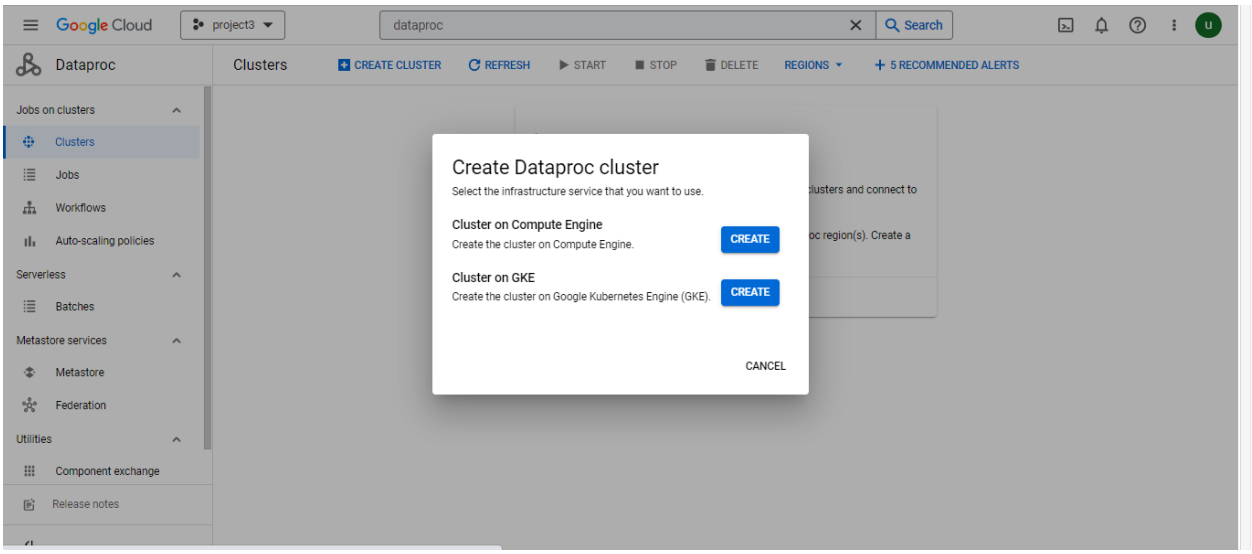
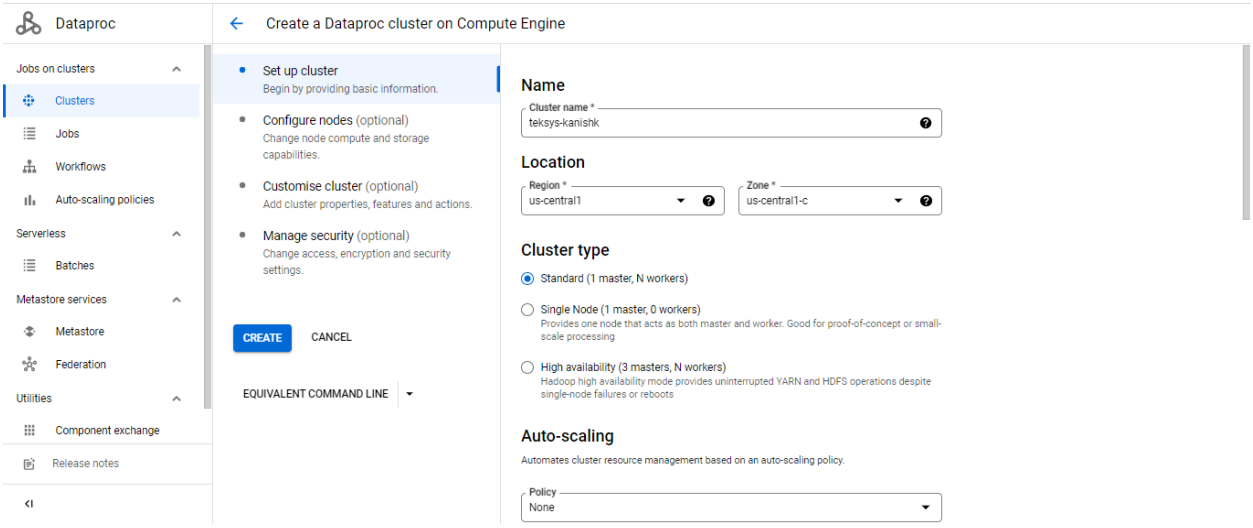
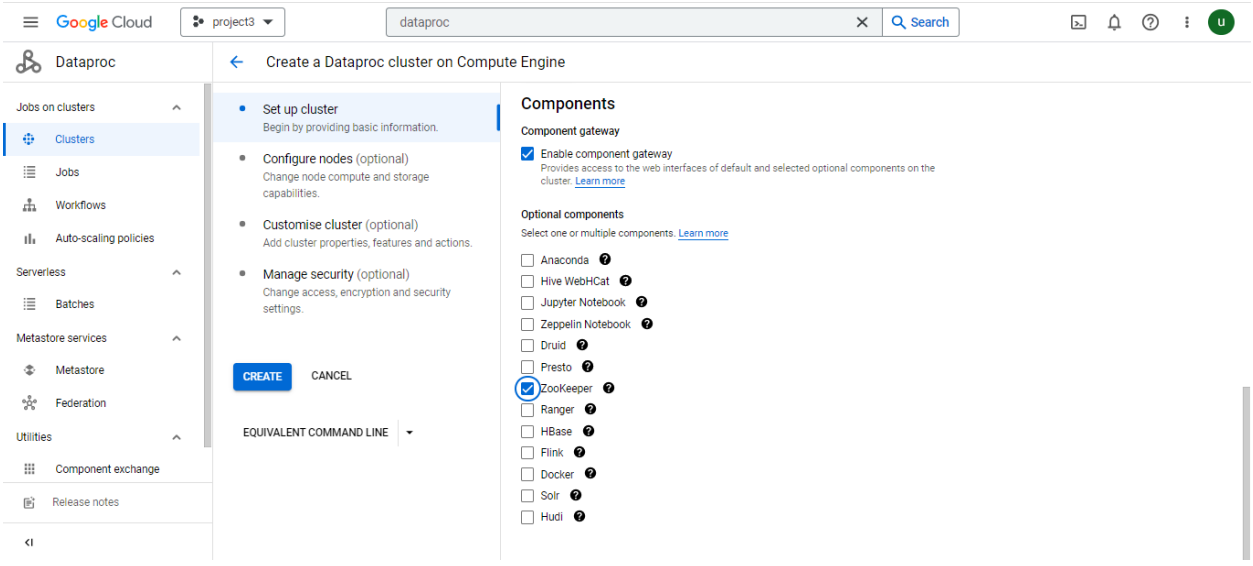
Cloud Dataproc

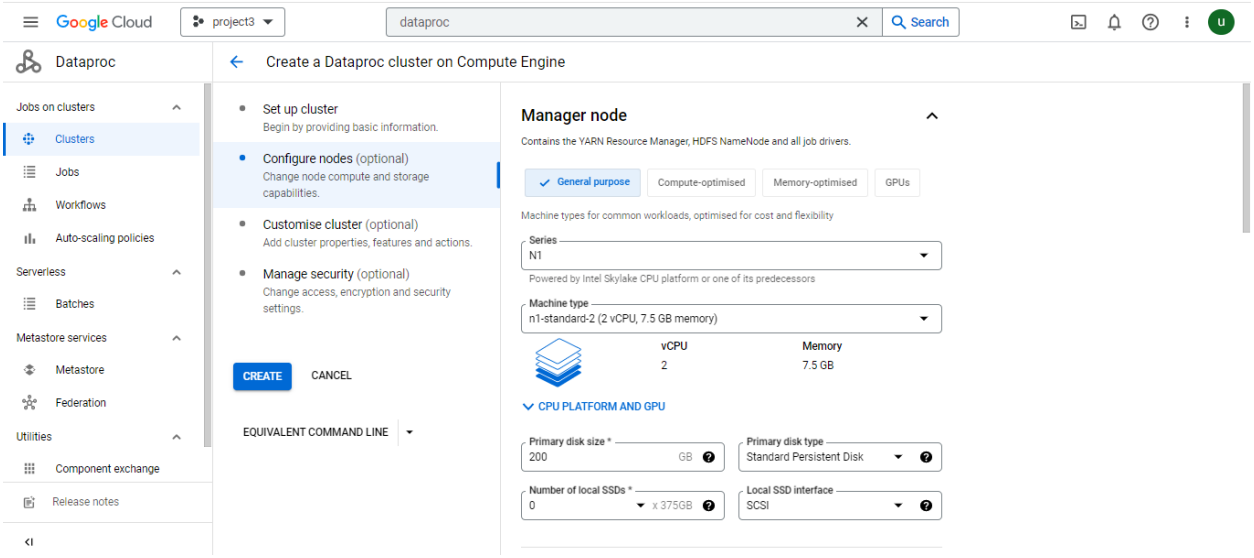


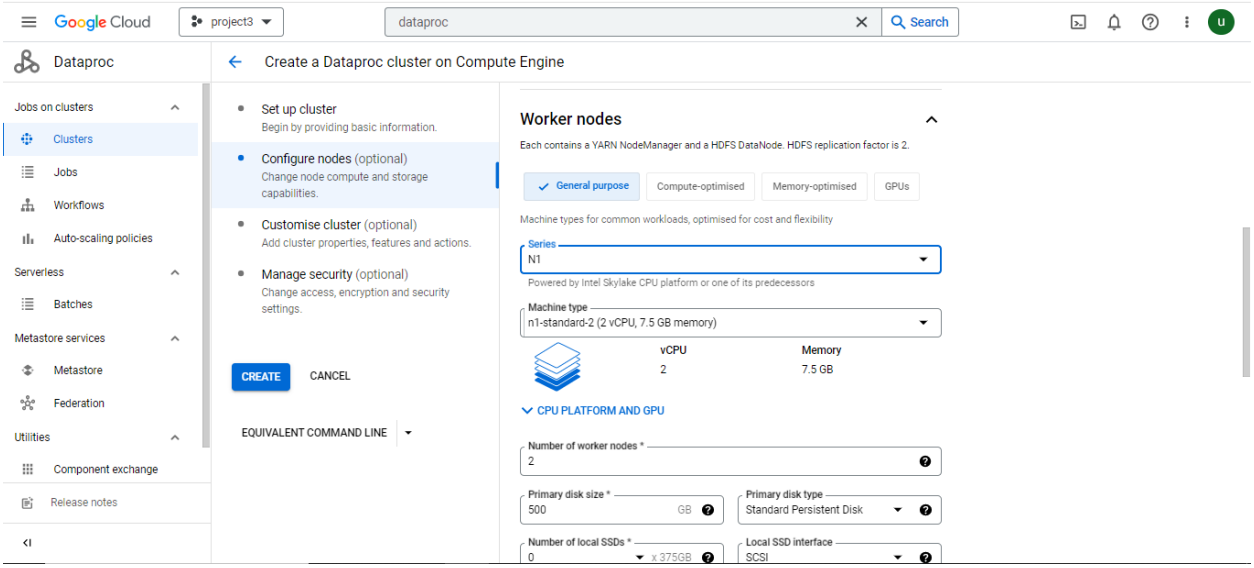


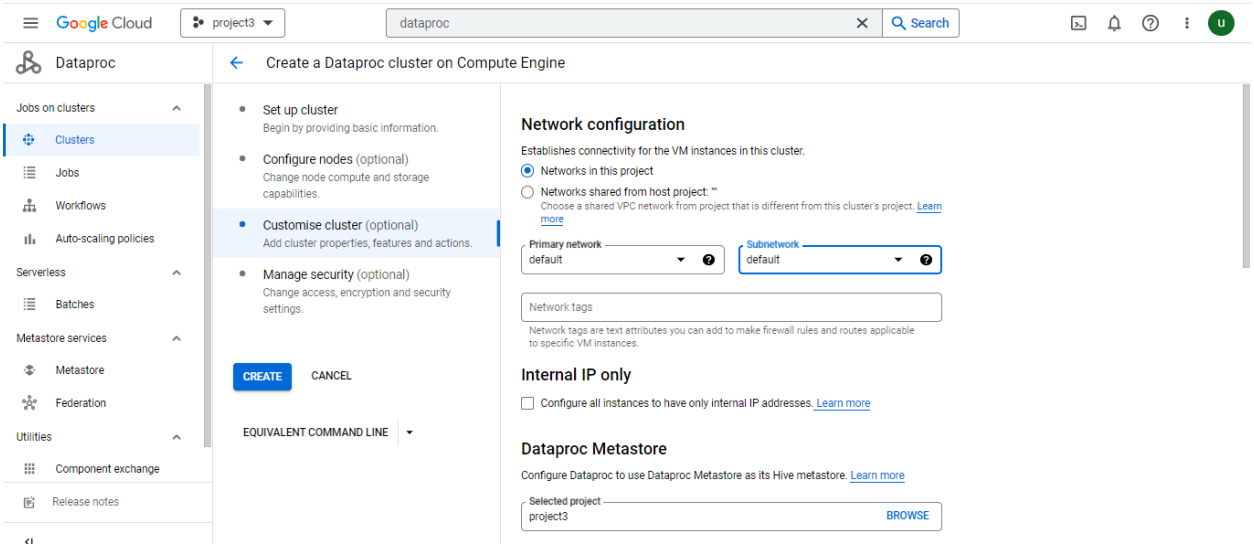

Google Cloud
project3
dataproc
Search
Dataproc
Clusters
CREATE CLUSTER
REFRESH
START
STOP
DELETE
REGIONS
5 RECOMMENDED ALERTS
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Create Dataproc cluster
Select the infrastructure service that you want to use.
Cluster on Compute Engine
Create the cluster on Compute Engine.
CREATE
Cluster on GKE
Create the cluster on Google Kubernetes Engine (GKE).
CREATE
CANCEL

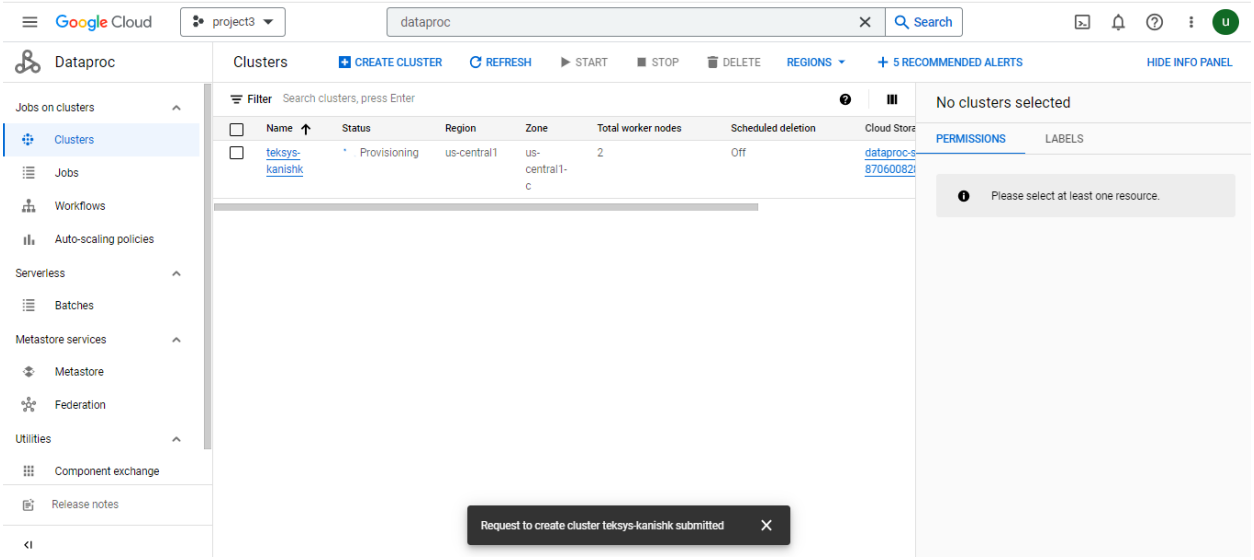

Dataproc
Create a Dataproc cluster on Compute Engine
Set up cluster
Begin by providing basic information.
Configure nodes (optional)
Change node compute and storage capabilities.
Customise cluster (optional)
Add cluster properties, features and actions.
Manage security (optional)
Change access, encryption and security settings.
CREATE
CANCEL
EQUIVALENT COMMAND LINE
Name
Cluster name *
teksys-kanishk
Location
Region *
us-central1
Zone *
us-central1-c
Cluster type
Standard (1 master, N workers)
Single Node (1 master, 0 workers)
Provides one node that acts as both master and worker. Good for proof-of-concept or small-scale processing
High availability (3 masters, N workers)
Hadoop high availability mode provides uninterrupted YARN and HDFS operations despite single-node failures or reboots
Auto-scaling
Automates cluster resource management based on an auto-scaling policy.
Policy
None

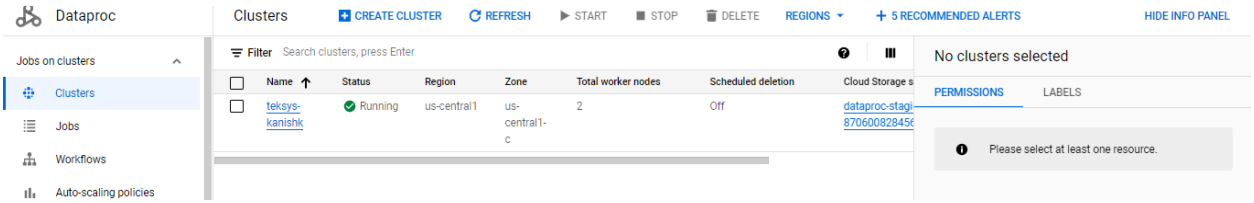

Google Cloud
project3
dataproc
Search
Dataproc
Create a Dataproc cluster on Compute Engine
Set up cluster
Begin by providing basic information.
Configure nodes (optional)
Change node compute and storage capabilities.
Customise cluster (optional)
Add cluster properties, features and actions.
Manage security (optional)
Change access, encryption and security settings.
CREATE
CANCEL
EQUIVALENT COMMAND LINE
Components
Component gateway
Enable component gateway
Provides access to the web interfaces of default and selected optional components on the cluster. Learn more
Optional components
Select one or multiple components. Learn more
Anaconda
Hive WebHCat
Jupyter Notebook
Zeppelin Notebook
Druid
Presto
ZooKeeper
Ranger
HBase
Flink
Docker
Solr
Hudi

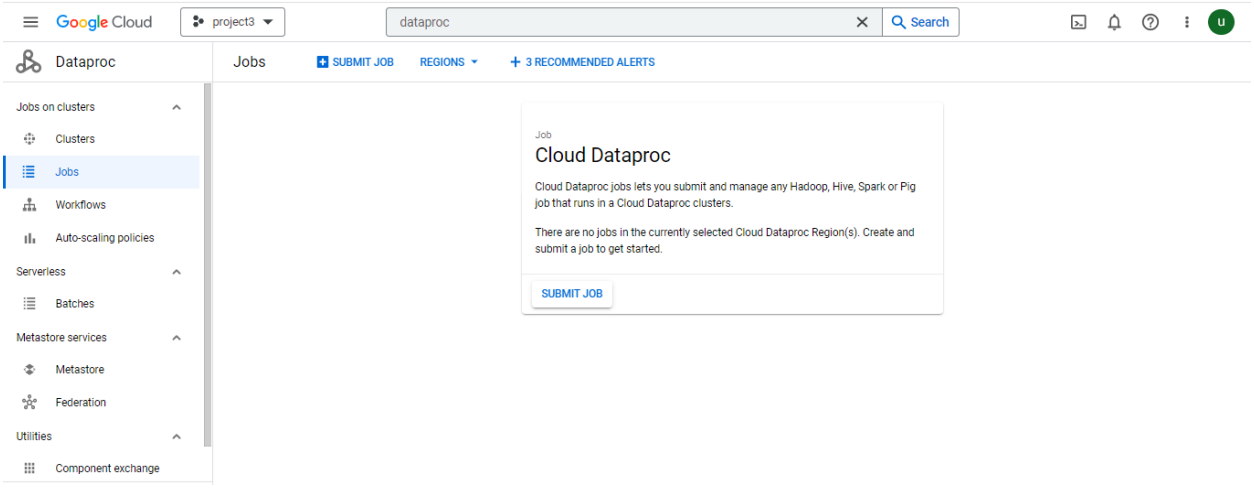

Google Cloud
project3
dataproc
Search
Dataproc
Create a Dataproc cluster on Compute Engine
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Set up cluster
Begin by providing basic information.
Configure nodes (optional)
Change node compute and storage capabilities.
Customise cluster (optional)
Add cluster properties, features and actions.
Manage security (optional)
Change access, encryption and security settings.
CREATE
CANCEL
EQUIVALENT COMMAND LINE
Manager node
Contains the YARN Resource Manager, HDFS NameNode and all job drivers.
General purpose
Compute-optimised
Memory-optimised
GPUs
Machine types for common workloads, optimised for cost and flexibility
Series
N1
Powered by Intel Skylake CPU platform or one of its predecessors
Machine type
n1-standard-2 (2 vCPU, 7.5 GB memory)
vCPU
2
Memory
7.5 GB
CPU PLATFORM AND GPU
Primary disk size *
200
GB
Primary disk type
Standard Persistent Disk
Number of local SSDs *
0
x 375GB
Local SSD interface
SCSI

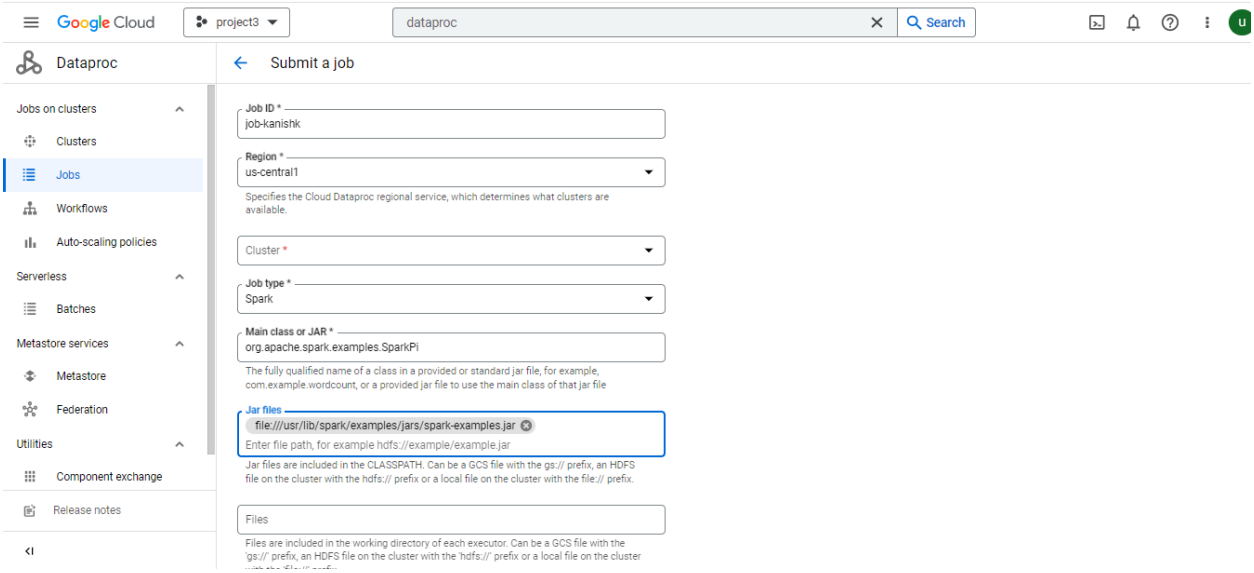

Google Cloud
project3
dataproc
Search
Dataproc
Create a Dataproc cluster on Compute Engine
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Set up cluster
Begin by providing basic information.
Configure nodes (optional)
Change node compute and storage capabilities.
Customise cluster (optional)
Add cluster properties, features and actions.
Manage security (optional)
Change access, encryption and security settings.
CREATE
CANCEL
EQUIVALENT COMMAND LINE
Worker nodes
Each contains a YARN NodeManager and a HDFS DataNode. HDFS replication factor is 2.
General purpose
Compute-optimised
Memory-optimised
GPUs
Machine types for common workloads, optimised for cost and flexibility
Series
N1
Powered by Intel Skylake CPU platform or one of its predecessors
Machine type
n1-standard-2 (2 vCPU, 7.5 GB memory)
vCPU
2
Memory
7.5 GB
CPU PLATFORM AND GPU
Number of worker nodes *
2
Primary disk size *
500
GB
Primary disk type
Standard Persistent Disk
Number of local SSDs *
0
x 375GB
Local SSD interface
SCSI

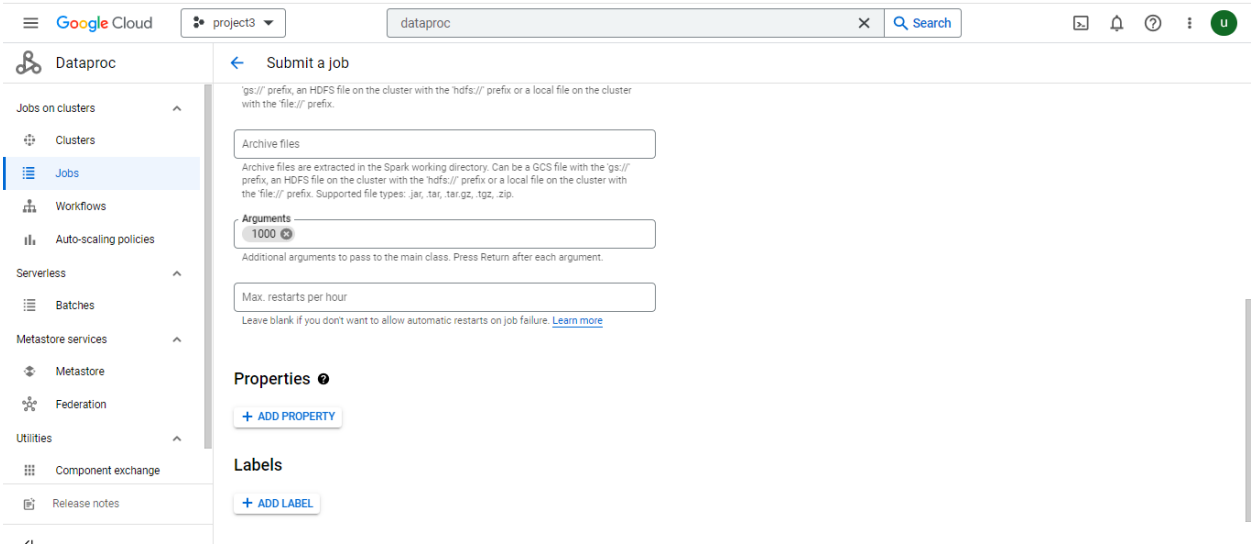

Google Cloud
project3
dataproc
Search
Dataproc
Create a Dataproc cluster on Compute Engine
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Set up cluster
Begin by providing basic information.
Configure nodes (optional)
Change node compute and storage capabilities.
Customise cluster (optional)
Add cluster properties, features and actions.
Manage security (optional)
Change access, encryption and security settings.
CREATE
CANCEL
EQUIVALENT COMMAND LINE
Network configuration
Establishes connectivity for the VM instances in this cluster.
Networks in this project
Networks shared from host project: ""
Choose a shared VPC network from project that is different from this cluster's project. Learn more
Primary network
default
Subnetwork
default
Network tags
Network tags are text attributes you can add to make firewall rules and routes applicable to specific VM instances.
Internal IP only
Configure all instances to have only internal IP addresses. Learn more
Dataproc Metastore
Configure Dataproc to use Dataproc Metastore as its Hive metastore. Learn more
Selected project
project3
BROWSE

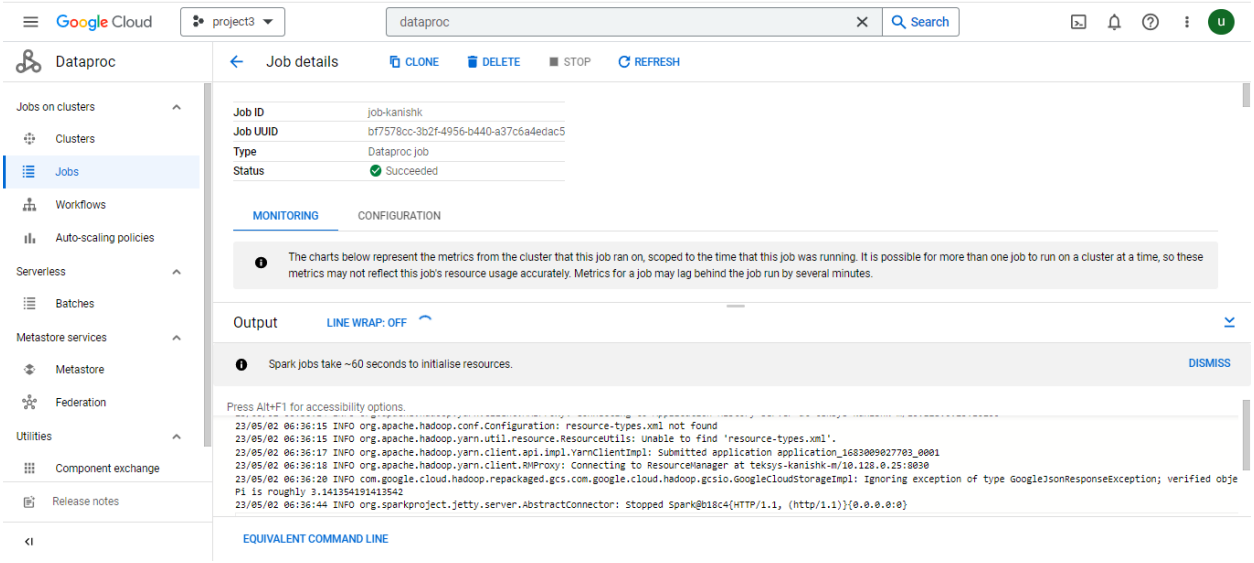

Google Cloud
project3
dataproc
Search
Dataproc
Clusters
CREATE CLUSTER
REFRESH
START
STOP
DELETE
REGIONS
5 RECOMMENDED ALERTS
HIDE INFO PANEL
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Filter
Search clusters, press Enter
Name
Status
Region
Zone
Total worker nodes
Scheduled deletion
Cloud Stora
teksys-kanishk
Provisioning
us-central1
us-central1-c
2
Off
No clusters selected
PERMISSIONS
LABELS
Please select at least one resource.
Request to create cluster teksys-kanishk submitted


Dataproc
Clusters
CREATE CLUSTER
REFRESH
START
STOP
DELETE
REGIONS
5 RECOMMENDED ALERTS
HIDE INFO PANEL
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Filter
Search clusters, press Enter
Name
Status
Region
Zone
Total worker nodes
Scheduled deletion
Cloud Storage s
teksys-kanishk
Running
us-central1
us-central1-c
2
Off
No clusters selected
PERMISSIONS
LABELS
Please select at least one resource.

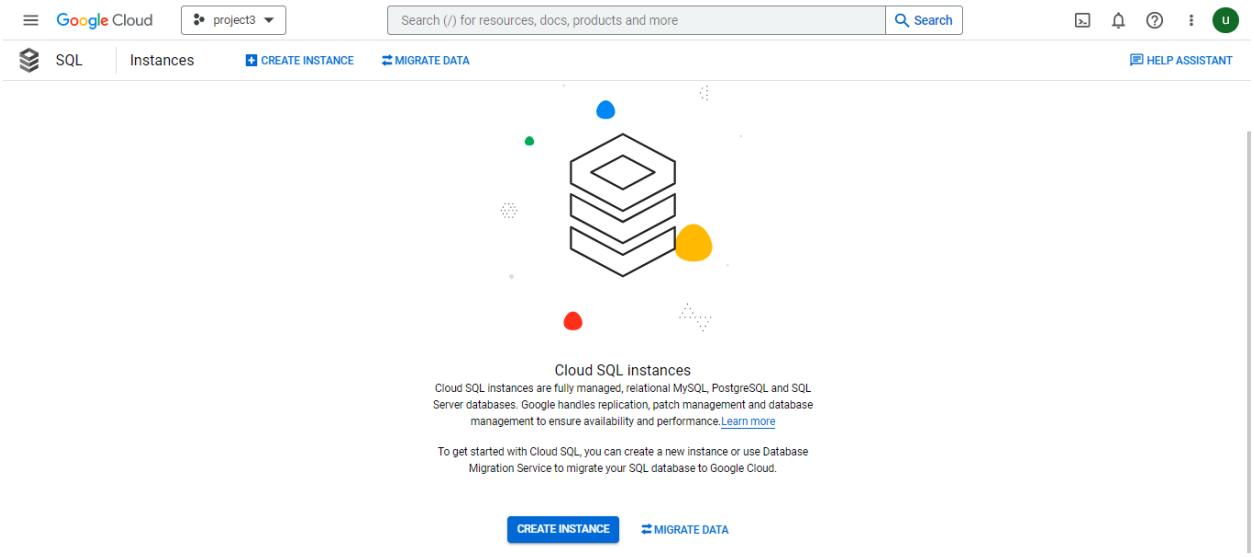

Google Cloud
project3
dataproc
Search
Dataproc
Jobs
SUBMIT JOB
REGIONS
3 RECOMMENDED ALERTS
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Job
Cloud Dataproc
Cloud Dataproc jobs lets you submit and manage any Hadoop, Hive, Spark or Pig job that runs in a Cloud Dataproc clusters.
There are no jobs in the currently selected Cloud Dataproc Region(s). Create and submit a job to get started.
SUBMIT JOB

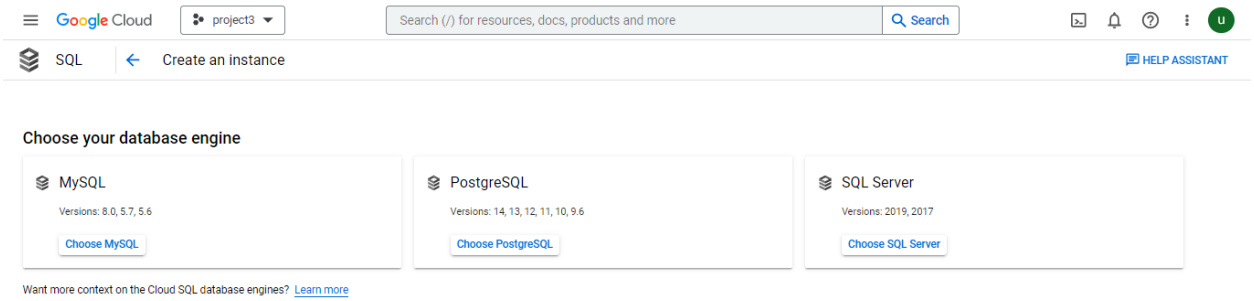

Google Cloud
project3
dataproc
Search
Dataproc
Submit a job
Jobs on clusters
Clusters
Jobs
Workflows
Auto-scaling policies
Serverless
Batches
Metastore services
Metastore
Federation
Utilities
Component exchange
Release notes
Job ID *
job-kanishk
Region *
us-central1
Specifies the Cloud Dataproc regional service, which determines what clusters are available.
Cluster *
Job type *
Spark
Main class or JAR *
org.apache.spark.examples.SparkPi
The fully qualified name of a class in a provided or standard jar file, for example, com.example.wordcount, or a provided jar file to use the main class of that jar file
Jar files
file:///usr/lib/spark/examples/jars/spark-examples.jar
Enter file path, for example hdfs://example/example.jar
Jar files are included in the CLASSPATH. Can be a GCS file with the gs:// prefix, an HDFS file on the cluster with the hdfs:// prefix or a local file on the cluster with the file:// prefix.
Files
Files are included in the working directory of each executor. Can be a GCS file with the 'gs://' prefix, an HDFS file on the cluster with the 'hdfs://' prefix or a local file on the cluster

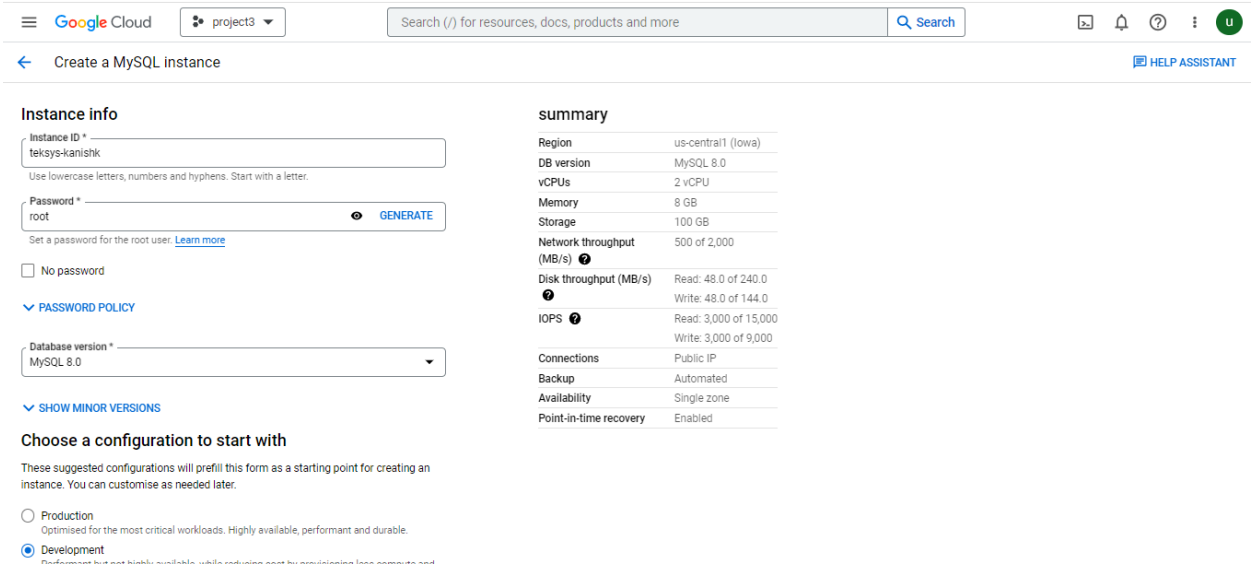

Google Cloud
project3
dataproc
Search
Dataproc
Submit a job
'gs://' prefix, an HDFS file on the cluster with the 'hdfs://' prefix or a local file on the cluster with the 'file://' prefix.
Archive files
Archive files are extracted in the Spark working directory. Can be a GCS file with the 'gs://' prefix, an HDFS file on the cluster with the 'hdfs://' prefix or a local file on the cluster with the 'file://' prefix. Supported file types: .jar, .tar, .tar.gz, .tgz, .zip.
Arguments
1000
Additional arguments to pass to the main class. Press Return after each argument.
Max. restarts per hour
Leave blank if you don't want to allow automatic restarts on job failure. Learn more
Properties
+ ADD PROPERTY
Labels
+ ADD LABEL


Google Cloud
project3
dataproc
Search
Dataproc
Job details
CLONE
DELETE
STOP
REFRESH
Job ID job-kanishk
Job UUID bf7578cc-3b2f-4956-b440-a37c6a4edac5
Type Dataproc job
Status Succeeded
MONITORING
CONFIGURATION
The charts below represent the metrics from the cluster that this job ran on, scoped to the time that this job was running. It is possible for more than one job to run on a cluster at a time, so these metrics may not reflect this job's resource usage accurately. Metrics for a job may lag behind the job run by several minutes.
Output
LINE WRAP: OFF
Spark jobs take ~60 seconds to initialise resources.
DISMISS
Press Alt+F1 for accessibility options.
23/05/02 06:36:15 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
23/05/02 06:36:15 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/05/02 06:36:17 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_1683009027703_0001
23/05/02 06:36:18 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at teksys-kanishk-m/10.128.0.25:8030
23/05/02 06:36:20 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageImpl: Ignoring exception of type GoogleJsonResponseException; verified obje
Pi is roughly 3.141354191413542
23/05/02 06:36:44 INFO org.sparkproject.jetty.server.AbstractConnector: Stopped Spark@b18c4{HTTP/1.1, (http/1.1)}{0.0.0.0:0}
EQUIVALENT COMMAND LINE

SQL


Google Cloud
project3
Search (/) for resources, docs, products and more
Search
SQL
Instances
CREATE INSTANCE
MIGRATE DATA
HELP ASSISTANT
Cloud SQL instances
Cloud SQL instances are fully managed, relational MySQL, PostgreSQL and SQL Server databases. Google handles replication, patch management and database management to ensure availability and performance.Learn more
To get started with Cloud SQL, you can create a new instance or use Database Migration Service to migrate your SQL database to Google Cloud.
CREATE INSTANCE
MIGRATE DATA


Google Cloud
project3
Search (/) for resources, docs, products and more
Search
SQL
Create an instance
HELP ASSISTANT
Choose your database engine
MySQL
Versions: 8.0, 5.7, 5.6
Choose MySQL
PostgreSQL
Versions: 14, 13, 12, 11, 10, 9.6
Choose PostgreSQL
SQL Server
Versions: 2019, 2017
Choose SQL Server
Want more context on the Cloud SQL database engines? Learn more


Google Cloud
project3
Search (/) for resources, docs, products and more
Search
Create a MySQL instance
HELP ASSISTANT
Instance info
Instance ID *
teksys-kanishk
Use lowercase letters, numbers and hyphens. Start with a letter.
Password *
root
GENERATE
Set a password for the root user. Learn more
No password
PASSWORD POLICY
Database version *
MySQL 8.0
SHOW MINOR VERSIONS
Choose a configuration to start with
These suggested configurations will prefill this form as a starting point for creating an instance. You can customise as needed later.
Production
Optimised for the most critical workloads. Highly available, performant and durable.
Development
summary
Region us-central1 (Iowa)
DB version MySQL 8.0
vCPUs 2 vCPU
Memory 8 GB
Storage 100 GB
Network throughput (MB/s) 500 of 2,000
Disk throughput (MB/s) Read: 48.0 of 240.0 Write: 48.0 of 144.0
IOPS Read: 3,000 of 15,000 Write: 3,000 of 9,000
Connections Public IP
Backup Automated
Availability Single zone
Point-in-time recovery Enabled