Managed hadoop and spark service . big data analytic for huge data .

Dataproc > cluster > cluster mode ,

Cluster mode :



Master node and worker node both are to be configured . - specify compute option for both .

Network , subnet , storage ,

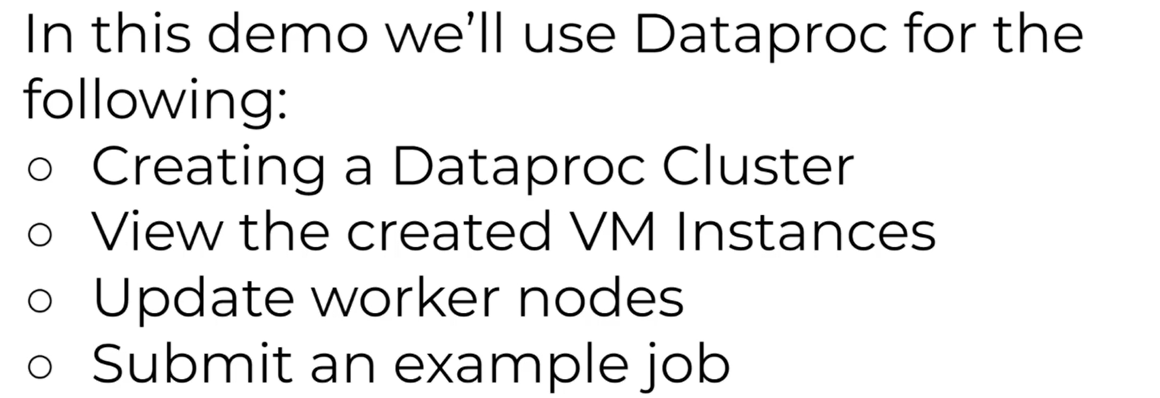
We can add preemptible worker nodes .

Create hadoop job in dataproc

Cloud Dataproc is a managed Hadoop and Spark service.

Cloud Dataproc is a managed Spark/Hadoop service that can be used to migrate Hadoop clusters GCP.

DATAPROC > CLUSTERS / JOBS / WORKFLOWS > Standard / single node / high availability cluster > MAchine type for master and worker node >



Dataproc > Create clusters > 1 master n workers >

VM instance > 1 master , 3 worker vm .

Go to cluster and edit worker nodes to 3 .

Create a spark job to run on the created cluster .



Arguments - 1000