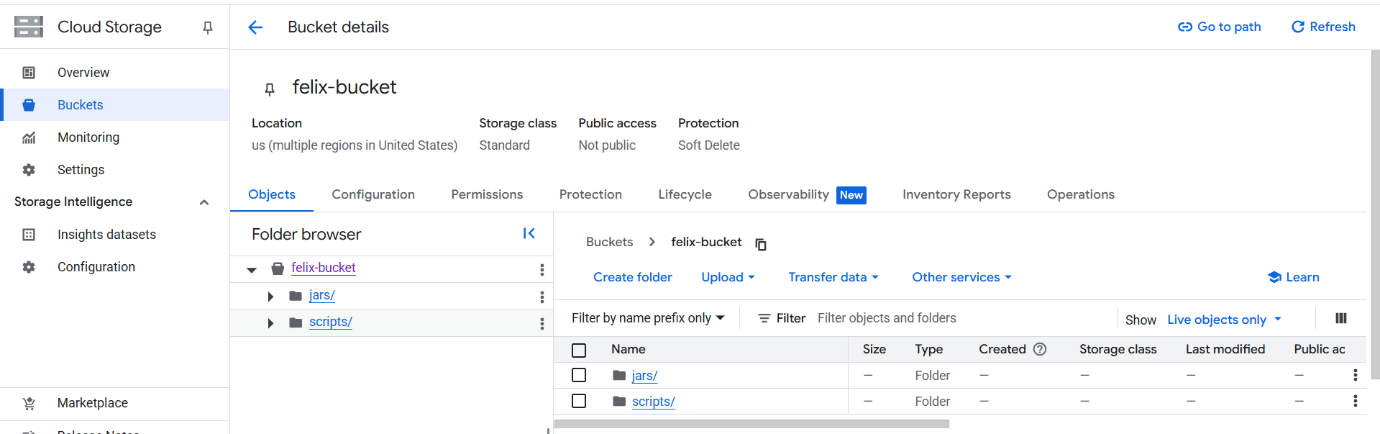
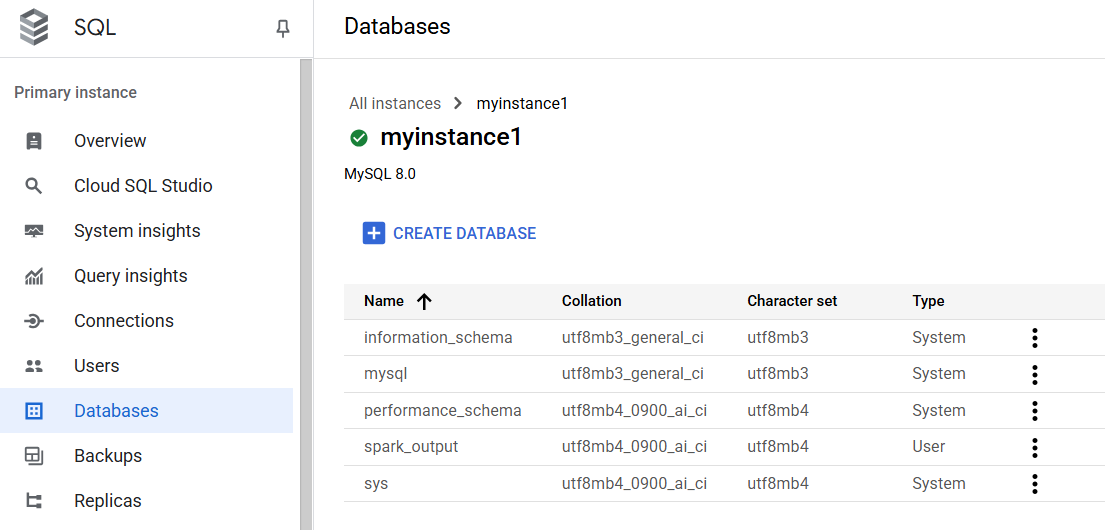
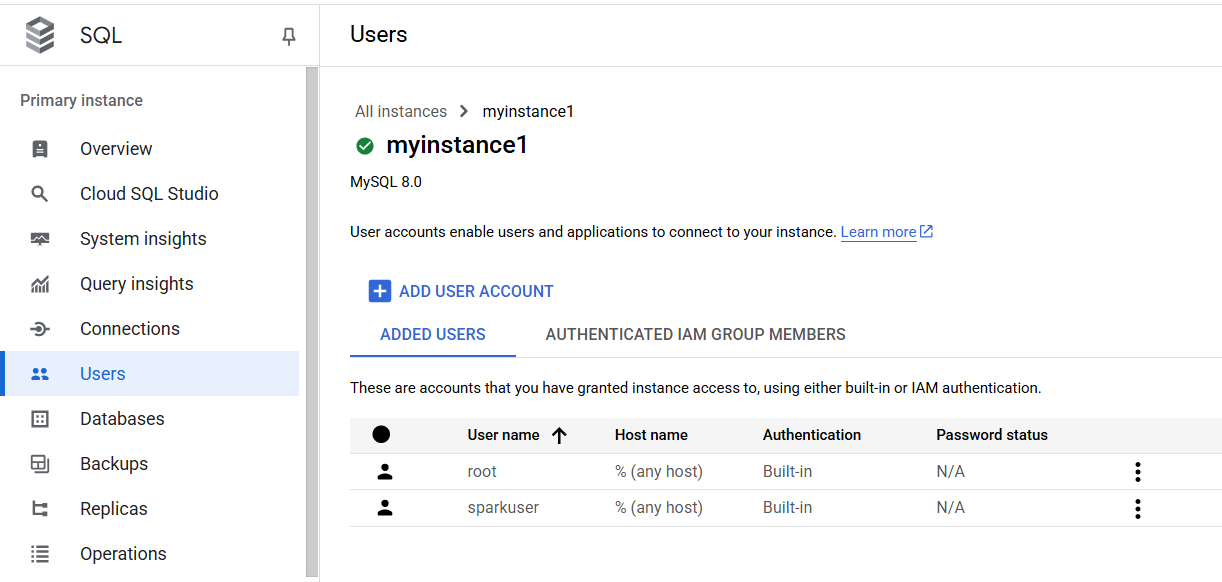
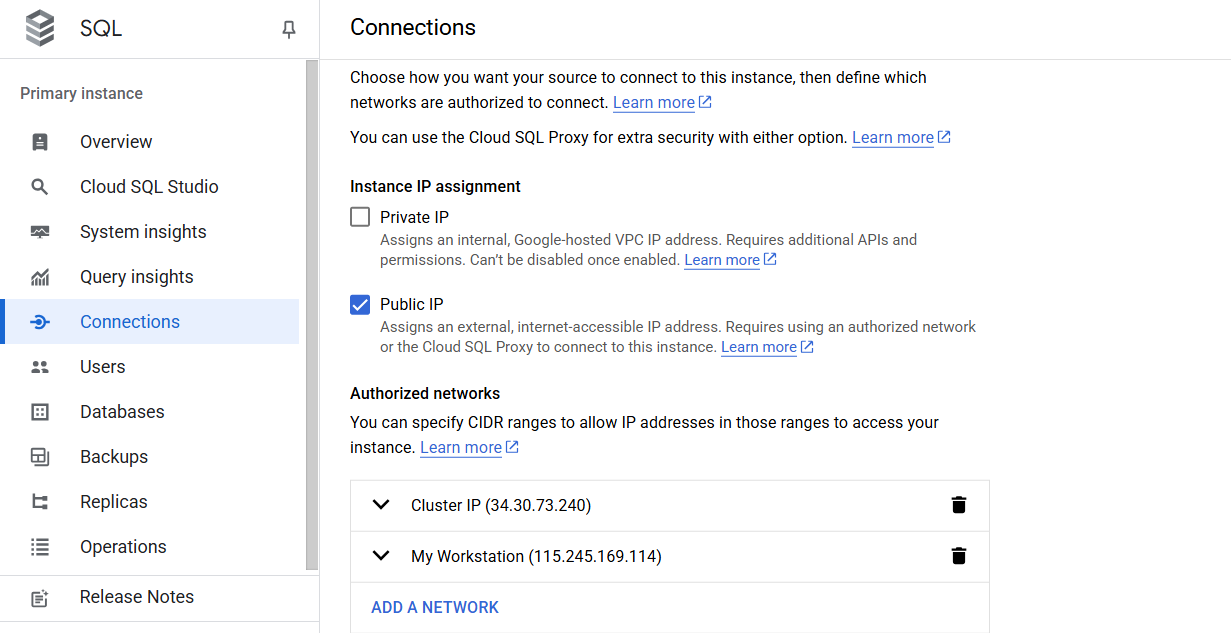
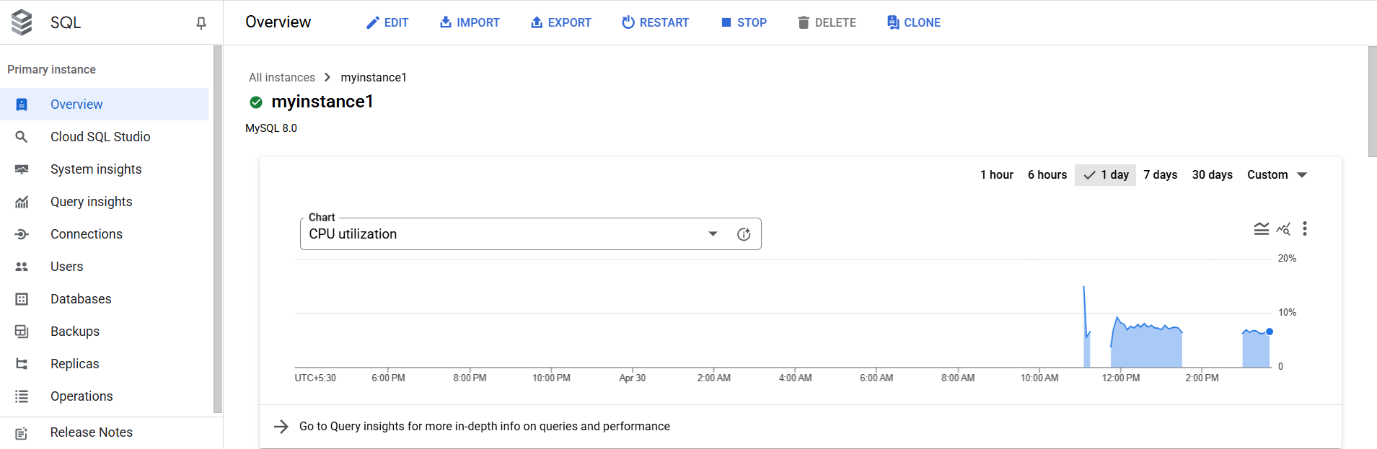
**PySpark job output from Dataproc into a MySQL database on GCP**



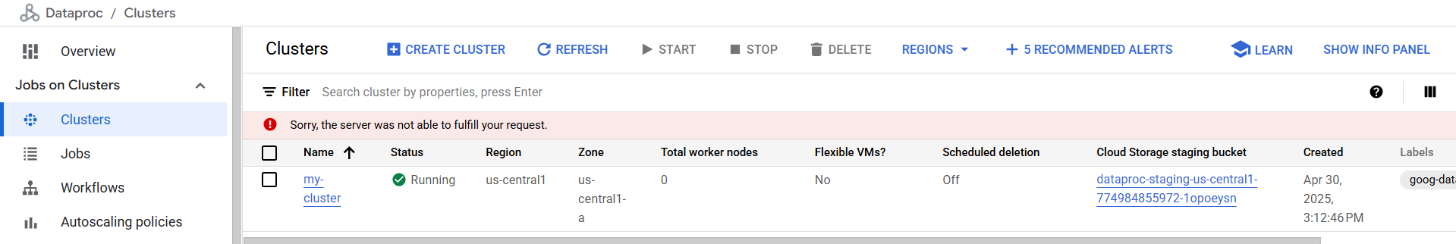
1. create bucket

* upload python script file in bucket/scripts/
  + In the script file,near the port, you need to give the IP of public IP in cloud SQL.
* upload the jar file in the bucket/jars/



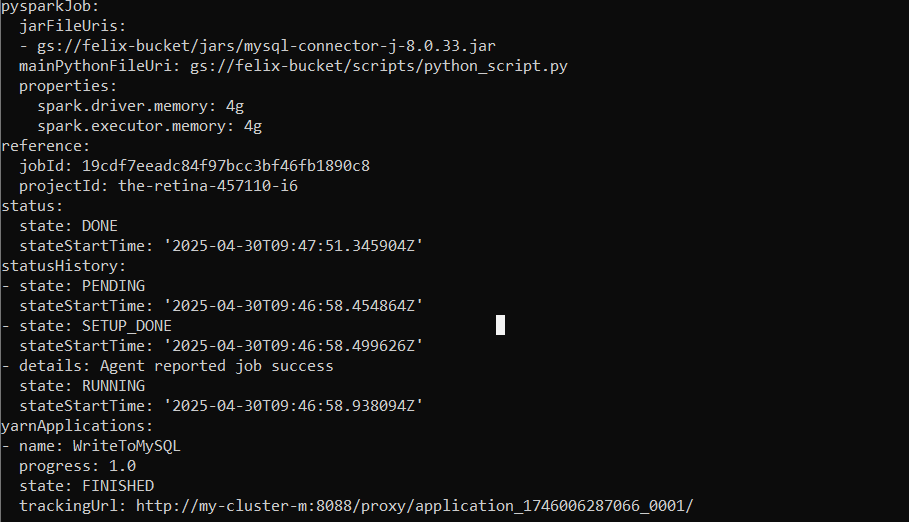
2. create instance in cloud sql

* after creating, give the IP in the connections --> networking --> public IP --> add network.
  + In add a network, you need to give 2 IP's, one is your system IP (what is my IP) and another IP is from the external IP in cluster(to find this, search for IP in the searchbox in gcp and select the IP Address --> external IP --> copy the IP).
* then, create a user with password.
* create a database named spark\_output.



3. Create a cluster using this command

* gcloud dataproc clusters create training-cluster --region=us-central1 --zone=us-central1-a --single-node --master-machine-type=n1-standard-2 --image-version=2.0-debian10



4. At last, submit the jobs using this command.

* gcloud dataproc jobs submit pyspark gs://felix-bucket/scripts/python\_script.py --cluster=my-cluster --region=us-central1 --jars=gs://felix-bucket/jars/mysql-connector-j-8.0.33.jar --properties=spark.driver.memory=4g,spark.executor.memory=4g

Result

