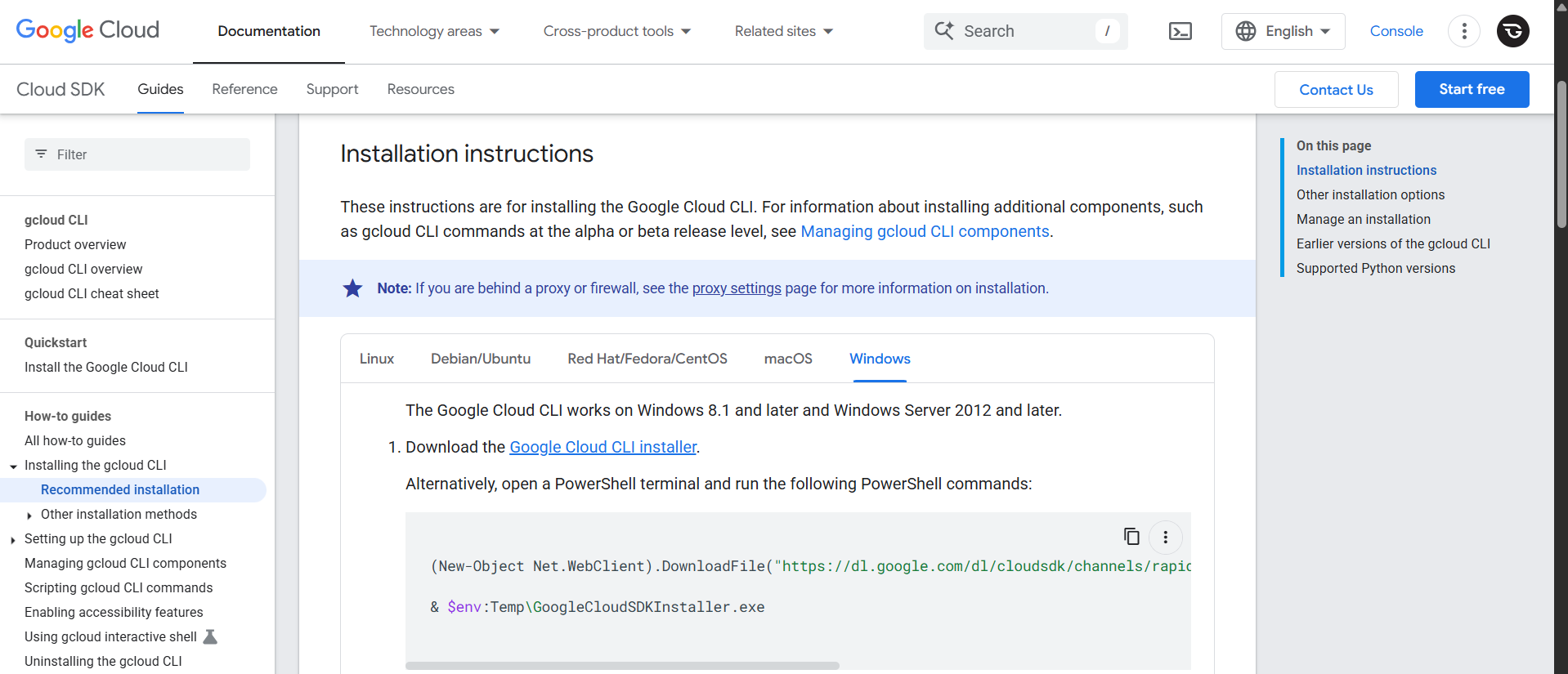
GCP Dataproc Procedures

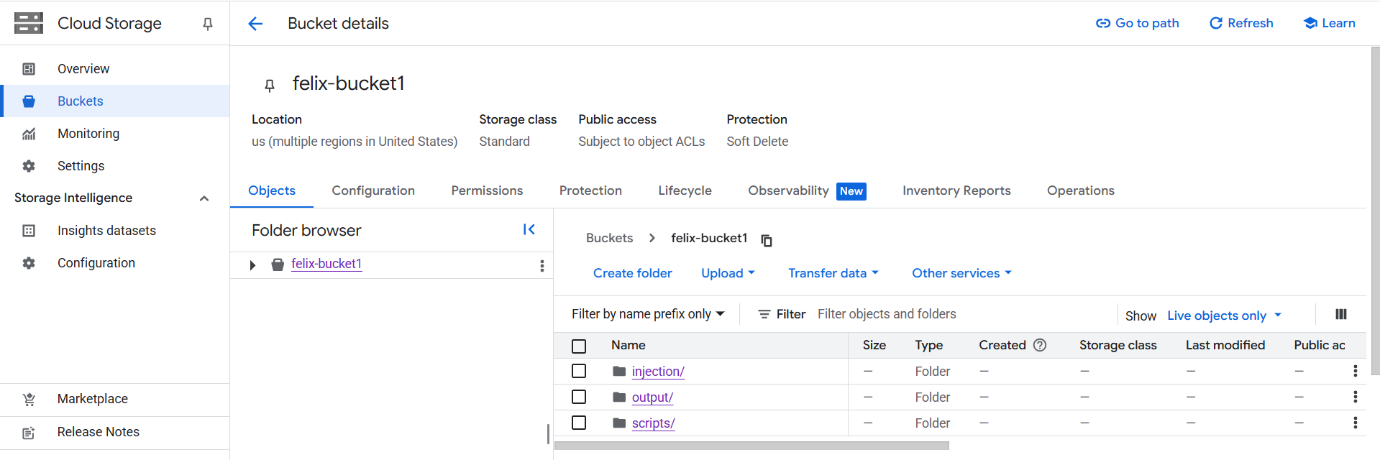
Follow the below steps to understand about GCP Dataproc service.



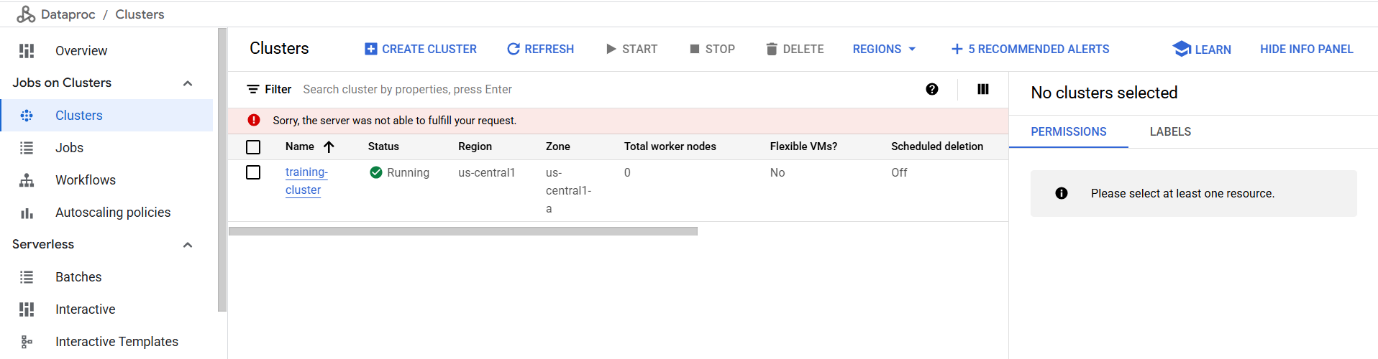
Step1: Download and Install Google Cloud SDK. For that, navigate to google.com and search for Google Cloud SDK Download. After that, click the URL which is named as [Install the gcloud CLI | Google Cloud CLI Documentation.](https://cloud.google.com/sdk/docs/install)

Step2: After configuring the Google Cloud SDK, check its version by **gcloud –version**. After that, to get started with it, enter the command named **gcloud init**.

Step3: Bucket Creation. Go to GCP cloud and search for Buckets. To create the bucket, give this command **gsutil mb gs://felix-bucket/**. The bucket name should be globally unique.



Step4: After creating the bucket in GCP bucket service, now give this command in the Google Cloud SDK or shell, **gsutil cp “C:/Users/Sharley Angel/Desktop/input/0\_input.csv” gs://felix-bucket/injection/** and **gsutil cp “C:/Users/Sharley Angel/Desktop/input/0\_input.csv” gs://felix-bucket/scripts/**. This command is used to upload a file in the Bucket storage in separate folders.



Step5: After uploading the files in the bucket storage, now we need to create the cluster by 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.**

****

Step6: After creating the cluster, execute this command to process the output. **gsutil cp -r gs://felix-bucket/output/processed/part-00000-46c304ee-a443-438b-a1b5-c43daac31d93-c000.csv c:/**