Perform Foundational Data ML - DevOps - Lab Internal - 2

Task 1: Run a simple Dataflow job

1. Navigation menu → Cloud storage → Create bucket [Leave others as default]
2. Name the bucket “ `YOUR_BUCKET_NAME` “
3. Click on create

1. Go to Big Query → Select Project ID Three dots → Create Dataset → Name it as what is given in the challenge “ `LAB` “ [Leave others as default]
2. Now we copy the CSV File given in the task by writing the following commands in shell

Task 1: Run a simple Dataflow job

You have used Dataflow in the quest to load data into BigQuery from Pub/Sub, now use the Dataflow batch template **Text Files on Cloud Storage to BigQuery** under "Process Data in Bulk (batch)" to transfer data from a Cloud Storage bucket (`gs://cloud-training/gsp323/lab.csv`). The following table has the values you need to correctly configure the Dataflow job.

```
gsutil cp gs://cloud-training/gsp323/lab.csv .

cat lab.csv

gsutil cp gs://cloud-training/gsp323/lab.schema .

cat lab.schema
```

3. Copy the data types from the terminal that is obtained from the cat command […] part

Now we will be copying it to the big query's datasets table

1. Navigation menu → Big Query → Dataset name [the lab thing] → Three dots → Open
2. Create table
3. Source = Google cloud storage → File form = " The link of csv given gs://cloud-training/gsp323/lab.csv " → Name it as what is shown in the table " `BigQuery output table` " → Click on edit as text → Past the schema that you copied earlier → Create table

Now creating a pipeline " Test file from cloud storage to Big Query "

1. Navigation menu → Dataflow → Create Job From template
2. Name of Job = anything → Region = one given in challenge → Data Flow Template = " The one given " **Text Files on Cloud Storage to BigQuery** under "Process Data in Bulk (batch)" "
3. Fill the details from the table given in the challenge → Make sure to change the `YOUR_PROJECT` to the project ID

Task 2: Run a simple Dataproc job

We will first create a cluster and then a Job

1. Navigation menu → Dataproc → Create cluster → Change the region [I think] → Keep everything as it is → Create cluster
2. Let the cluster finish
3. Click the cluster → The SSH in master node → Paste the command given in the task " hdfs dfs -cp gs://cloud-training/gsp323/data.txt /data.txt "

Now will create a Dataproc Job

1. Navigating menu → Job → Submit Job → Name = default whatever it is → Fill the details form the table given → Submit

Task 3: Run a simple Dataprep job

1. Navigation menu → Dataprep → Agree to everything
2. Import data → GCS → Past the link given in the task “ gs://cloud-training/gsp323/runs.csv “ → Go
3. Do all the task given in the table [Like Column → Filter rows → On Column value → Contains → Paste the pattern given in the task → Delete matching rows → Add
4. Rename all the columns as given
5. Run

Task 4: AI

All the commands will be there in the “ Cloud Natural Language API Quick Lab “

- Dont forget to change the content of the results page
- Then paste this command `gsutil cp result1.json gs://YOUR_PROJECT-marking/task4-gvi.result` [Don't forget to change the place holder name]

```
Task 4 :::=>

gcloud iam service-accounts create my-natlang-sa \
  --display-name "my natural language service account"

-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------


gcloud iam service-accounts keys create ~/key.json \
  --iam-account my-natlang-sa@$DEVSHELL_PROJECT_ID.iam.gserviceaccount.com


-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------

 wget https://raw.githubusercontent.com/guys-in-the-cloud/cloud-skill-boosts/main/Chal
lenge-labs/Perform%20Foundational%20Data%2C%20ML%2C%20and%20AI%20Tasks%20in%20Google%2
0Cloud%3A%20Challenge%20Lab/speech-request.json
```

```
--------------------------------------------------------------------------------
--------------------------------------------------------------------

curl -s -X POST -H "Content-Type: application/json" --data-binary @speech-request.json
\
"https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}" > speech.json

--------------------------------------------------------------------------------
--------------------------------------------------------------------

gsutil cp speech.json gs://$DEVSHELL_PROJECT_ID-marking/<changefilename>
--------------------------------------------------------------------------------
--------------------------------------------------------------------

gcloud ml language analyze-entities --content="Old Norse texts portray Odin as one-eye
d and long-bearded, frequently wielding a spear named Gungnir and wearing a cloak and
 a broad hat." > language.json
--------------------------------------------------------------------------------
--------------------------------------------------------------------

gsutil cp language.json gs://$DEVSHELL_PROJECT_ID-marking/<changefilename>

--------------------------------------------------------------------------------
--------------------------------------------------------------------

wget https://github.com/guys-in-the-cloud/cloud-skill-boosts/blob/main/Challenge-labs/
Perform%20Foundational%20Data%2C%20ML%2C%20and%20AI%20Tasks%20in%20Google%20Cloud:%20C
hallenge%20Lab/video-intelligence-request.json
--------------------------------------------------------------------------------
--------------------------------------------------------------------


curl -s -H 'Content-Type: application/json' \
    -H 'Authorization: Bearer '$(gcloud auth print-access-token)'' \
    'https://videointelligence.googleapis.com/v1/videos:annotate' \
    -d @video-intelligence-request.json  > video.json

--------------------------------------------------------------------------------
--------------------------------------------------------------------


 gsutil cp video.json gs://$DEVSHELL_PROJECT_ID-marking/<changefilename>
```