Important Commands

To set up GOOGLE_APPLICATION_CREDENTIALS

Linux - export GOOGLE_APPLICATION_CREDENTIALS="/home/dezyre/svc/fabric-1333-29edd96ad1f1.json"

Windows - set GOOGLE_APPLICATION_CREDENTIALS=C:\Documents\Learning\SVC\fabric-1333-29edd96ad1f1.json

To Configure service account with google cloud SDK

gcloud auth activate-service-account svc-rw@fabric-1333.iam.gserviceaccount.com --key-file=C:\Documents\Learning\SVC\fabric-1333-29edd96ad1f1.json --project=fabric-1333

To Copy data from Local to Cloud Bucket using SDK

gsutil cp yelp_academic_dataset_business.json gs://gcs-dezyre-bucket/data/

*gsutil -m cp *.json gs://gcs-dezyre-bucket/data/*

To make bucket using Cloud SDK

gsutil mb gs://gcs-dezyre-bucket2

To publish messages to pubsub topic

python3.7 scripts/publish_messages.py --config_path=config/publish_config.ini

To run dataflow stream job from console

python3 scripts/dataflow_stream.py --input_subscription=projects/fabric-1333/subscriptions/gcp-yelp-topic-sub --output_table=test.yelp_covid --output_error_table=test.error --runner DataflowRunner --project fabric-1333 --region us-west1 --service_account_email svc-rw@fabric-1333.iam.gserviceaccount.com --staging_location gs://gcp-file-source/dataflow/staging --temp_location gs://gcp-file-source/dataflow/temp --job_name test-stream-bq --num_workers 1 --max_num_workers 2

To Run Batch Dataflow Job from Console

```
python3 scripts/dataflow_batch_test.py --input_path=gs://gcp-file-source/data1/* --
table=test.business --error_table=test.error --runner DataflowRunner --project fabric-
1333 --region us-west1 --service_account_email svc-rw@fabric-
1333.iam.gserviceaccount.com --staging_location gs://gcp-file-source/dataflow/staging
--temp_location gs://gcp-file-source/dataflow/temp --job_name test-batch-bq --
num_workers 1 --max_num_workers 4
```

To Create Batch Dataflow Job Template from Console

```
python3 scripts/dataflow_batch_templated.py --runner DataflowRunner --project fabric-
1333 --region us-west1 --staging_location gs://gcp-file-source/dataflow/staging --
temp_location gs://gcp-file-source/dataflow/temp --num_workers 1 --max_num_workers 6 -
-template_location gs://gcp-file-source/batchtemplate --save_main_session True
```

To Create Stream Dataflow Job Template from Console

```
Python3 dataflow_stream_templated.py --runner DataflowRunner --project fabric-1333 --
region us-west1 --staging_location gs://gcp-file-source/dataflow/staging --
temp_location gs://gcp-file-source/dataflow/temp --num_workers 1 --max_num_workers 4 -
-template_location gs://gcp-file-source/streamtemplate --save_main_session True
```