Important Commands

**To set up GOOGLE\_APPLICATION\_CREDENTIALS**

*Linux - export GOOGLE\_APPLICATION\_CREDENTIALS=/home/shrutijoshi/helpful-cat-324414-bbf5a431efb0.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*

gcloud auth activate-service-account svc-dezyre-editor@helpful-cat-324414.iam.gserviceaccount.com --key-file=helpful-cat-324414-bbf5a431efb0.json --project=helpful-cat-324414

**Install Python modules for apache beam**

Pip install apache-beam[gcp]

**Run Demo and branched pipeline**

Python3 pipeline\_demo.py

Python3 branched\_pipeline.py

**To publish messages to pubsub topic**

python3 pulish\_csv\_2\_pubsub.py --config\_path=config/publish\_config.ini

**To run stream job in dataflow**

*python3*  flights\_stream\_dataflow*.py --input\_subscription=projects/*helpful-cat-324414*/subscriptions/flights-sub --output\_table=dataset1.flights --output\_error\_table=dataset1.error --runner DataflowRunner --project* helpful-cat-324414 *--region us-west1 --service\_account\_email* svc-dezyre-editor@helpful-cat-324414.iam.gserviceaccount.com *--staging\_location gs://dezyre-test-bucket1/dataflow/staging --temp\_location gs://dezyre-test-bucket1/dataflow/temp --job\_name flight-test-stream-bq --num\_workers 1 --max\_num\_workers 2*

**Dataflow parameters**

<https://cloud.google.com/dataflow/docs/reference/pipeline-options>

**To Run Batch Dataflow Job from Console using Dataflow Runner**

*python3 flights\_batch\_templated.py --input\_path=gs://dezyre-test-bucket1/data/flights.csv --side\_input\_path=gs://dezyre-test-bucket1/data/airlines.csv --table=dataset1.flights\_side --error\_table=dataset1.error --runner DataflowRunner --project helpful-cat-324414 --region us-west1 --service\_account\_email svc-dezyre-editor@helpful-cat-324414.iam.gserviceaccount.com --staging\_location gs://dezyre-test-bucket1/dataflow/staging --temp\_location gs://dezyre-test-bucket1/dataflow/temp --job\_name flight-test-batch-bq --num\_workers 1 --max\_num\_workers 2*

**To Run Batch Dataflow Job from Console using Direct Runner**

*python3 flights\_batch\_templated.py --input\_path=gs://dezyre-test-bucket1/data/flights\_sample.csv -–side\_input\_path=gs://dezyre-test-bucket1/data/airlines.csv --table= helpful-cat-324414:dataset1.flights\_side --error\_table= helpful-cat-324414:dataset1.error --staging\_location gs://dezyre-test-bucket1/dataflow/staging --temp\_location gs://dezyre-test-bucket1/dataflow/temp --project helpful-cat-324414 --job\_name flight-test-batch-bq*

**To create template of dataflow job**

*python3 flights\_batch\_templated.py --runner DataflowRunner --project helpful-cat-324414 --region us-west1 --staging\_location gs://dezyre-test-bucket1/dataflow/staging --temp\_location gs://dezyre-test-bucket1/dataflow/temp --num\_workers 1 --max\_num\_workers 2 --template\_location gs:// dezyre-test-bucket1/batchtemplate --save\_main\_session True*