



Faculty of Engineering & Applied Science

SOFE4630U – Cloud Computing

Project Milestone 2 – Data Processing: Dataflow – Apache Beam

GitHub: https://github.com/abdulbhutta/Cloud-Computing/tree/main/Final%20Project/Milestone2_DataFlow

Due Date: 02/16/2023

First Name	Last Name	Student ID
Abdul	Bhutta	100785884
Andrew	Mikaeel	100525236
Anson	Tu	100655482
Owen	Musselman	100657709

D1: Wordcount Example

Run the wordcount.py script

```
abdulbhutta7@cloudshell:~ (projectmilestone2)$ python wordcount.py --output outputs
INFO:root:Missing pipeline option (runner). Executing pipeline using the default runner: DirectRunner.
INFO:apache_beam.internal.gcp.auth:Setting socket default timeout to 60 seconds.
INFO:root:Default Python SDK image for environment is apache/beam_python3.9_sdk:2.44.0
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function annotate_downstream_side_inputs at 0x7f2fd580f1f0> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function fix_side_input_pcoll_coders at 0x7f2fd580f310> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function pack_combiners at 0x7f2fd580f820> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function lift_combiners at 0x7f2fd580f8b0> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function expand_sdf at 0x7f2fd580fa60> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function expand_gbk at 0x7f2fd580faf0> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function sink_flatten at 0x7f2fd580fc10> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function greedily_fuse at 0x7f2fd580fc40> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function read_to_Impulse at 0x7f2fd580fd30> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function impulse_to_input at 0x7f2fd580fdc0> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function sort_stages at 0x7f2fd5810040> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function add_Impulse_to_dangling_transforms at 0x7f2fd5810160> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function setup_timer_mapping at 0x7f2fd580ff70> =====
INFO:apache_beam.runners.portability.fn_api_runner.translations:<function populate_data_channel coders at 0x7f2fd58100d0> =====
INFO:apache_beam.runners.worker.statecache:Creating state cache with size 104857600
INFO:apache_beam.runners.portability.fn_api_runner.worker_handlers:Created Worker handler <apache_beam.runners.portability.fn_api_runner.worker_handlers.EmbeddedWorker@ref Environment default_environment 1 (beam:env:embedded_python:v1, b'')
INFO:apache_beam.io.filebasedsink:Starting finalize write threads with num_shards: 1 (skipped: 0), batches: 1, num_threads: 1
INFO:apache_beam.io.filebasedsink:Renamed 1 shards in 0.01 seconds.
```

View the output

```
abdulbhutta7@cloudshell:~ (projectmilestone2)$ ls outputs*
outputs-00000-of-00001
abdulbhutta7@cloudshell:~ (projectmilestone2)$ ls
outputs-00000-of-00001  README-cloudshell.txt  smartMeter  'wordcount.py'  wordcount.py
abdulbhutta7@cloudshell:~ (projectmilestone2)$ cat outputs* | more
KING: 243
LEAR: 236
DRAMATIS: 1
PERSONAE: 1
king: 65
of: 447
Britain: 2
OF: 15
FRANCE: 10
DUKE: 3
BURGUNDY: 8
CORNWALL: 63
ALBANY: 67
EARL: 2
KENT: 156
GLOUCESTER: 141
EDGAR: 126
son: 29
to: 438
Gloucester: 26
EDMUND: 99
bastard: 7
CURAN: 6
a: 366
courtier: 1
Old: 13
Man: 11
tenant: 3
Doctor: 12
Fool: 73
OSWALD: 53
steward: 2
Goneril: 12
A: 51
Captain: 12
--More--
```

Get the current project ID

```
abdulbhutta7@cloudshell:~ (projectmilestone2)$ PROJECT=$(gcloud config list project --format "value(core.project)")
abdulbhutta7@cloudshell:~ (projectmilestone2)$ echo $PROJECT
projectmilestone2
abdulbhutta7@cloudshell:~ (projectmilestone2)$
```

Create a bucket

[Create a bucket](#)

Name your bucket
Pick a globally unique, permanent name. [Naming guidelines](#)

projectmilestone2-bucket
Tip: Don't include any sensitive information

LABELS (OPTIONAL)

[CONTINUE](#)

Choose where to store your data
Location: us (multiple regions in United States)
Location type: Multi-region

Choose a storage class for your data
Default storage class: Standard

Choose how to control access to objects
Public access prevention: On
Access control: Uniform

Choose how to protect object data
Protection tools: None
Data encryption: Google-managed key

[CREATE](#) [CANCEL](#)

Public access will be prevented

This bucket is set to prevent exposure of its data on the public internet.

Keep this setting enabled unless you have a use case that requires public access (such as static website hosting). You can change it now or later. [Learn more](#)

Enforce public access prevention on this bucket
 Don't show this message again

[CANCEL](#) [CONFIRM](#)

Save bucket ID

```
abdulbhutta7@cloudshell:~ (projectmilestone2)$ BUCKET=gs://$PROJECT-bucket
abdulbhutta7@cloudshell:~ (projectmilestone2)$ echo $BUCKET
gs://projectmilestone2-bucket
abdulbhutta7@cloudshell:~ (projectmilestone2)$ 
```

Run pipeline using Dataflow

```
abdulbhutta7@cloudshell:~ (projectmilestone2)$ python wordcount.py --region northamerica-northeast2 --runner DataflowRunner --project $PROJECT --temp_location $BUCKET/tmp/ --input gs://dataflow-samples/shakespeare/winterstale.txt --output $BUCKET/result/output --experiment use_unsupported_python_version
INFO:apache_beam.internal.gcp.auth:Setting socket default timeout to 60 seconds.
INFO:apache_beam.internal.gcp.auth:socket default timeout is 60 seconds.
INFO:apache_beam.runners.portability.stager:Uploading source distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpdhx3umxc', 'apache-beam==2.44.0', '--no-deps', '--no-binary', ':all:']
INFO:apache_beam.runners.portability.stager:Staging SDK sources from PyPi: dataflow_python_sdk.tar
INFO:apache_beam.runners.portability.stager:Downloading binary distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpdhx3umxc', 'apache-beam==2.44.0', '--no-deps', '--only-binary', ':all:', '--python-version', '39', '--implementation', 'cp', '--abi', 'cp39', '--platform', 'manylinux2014_x86_64']
INFO:apache_beam.runners.portability.stager:Staging binary distribution of the SDK from PyPi: apache_beam-2.44.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
INFO:root:Python SDK image for environment is apache_beam-python3.9.sdk[2.44.0]
INFO:root:Default Python SDK image for environment is apache_beam-python3.9.sdk[2.44.0]
INFO:root:Python SDK container image set to gcr.io/cloud-dataflow/vtbase/python39:2.44.0
INFO:root:Python SDK container image set to gcr.io/cloud-dataflow/vtbase/python39:2.44.0" for Docker environment
INFO:apache_beam.runners.portability.rn_apache_runner.translations:-----> <function pack_combiners at 0x7f72d6d835310>
INFO:apache_beam.runners.portability.rn_apache_runner.translations:-----> <function sort_stages at 0x7f72d6d835a0>
INFO:apache_beam.runners.portability.rn_apache_runner.translations:-----> <function get_gcs_location at 0x7f72d6d835c0>
INFO:apache_beam.runners.portability.rn_apache_runner.translations:-----> <function copy_to_temp_location at 0x7f72d6d835e0>
INFO:apache_beam.runners.portability.rn_apache_runner.translations:-----> <function copy_to_temp_location at 0x7f72d6d835f0>
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/pickled_main_session...
INFO:apache_beam.runners.dataflow.internal.apiclient:Completed GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/pickled_main_session in 0 seconds.
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/dataflow_python_sdk.tar...
INFO:apache_beam.runners.dataflow.internal.apiclient:Completed GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/dataflow_python_sdk.tar in 0 seconds.
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/apache_beam-2.44.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl...
INFO:apache_beam.runners.dataflow.internal.apiclient:Completed GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/apache_beam-2.44.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl in 0 seconds.
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/pipeline.pb...
INFO:apache_beam.runners.dataflow.internal.apiclient:Completed GCS upload to gs://projectmilestone2-bucket/tmp/beamapp-abdulbhutta7-0210210951-549111-09fe1zab.1676063391.549463/pipeline.pb in 0 seconds.
INFO:apache_beam.runners.dataflow.internal.apiclient>Create job: <Job
```

GROUPT2_PROJECTMILESTONE2

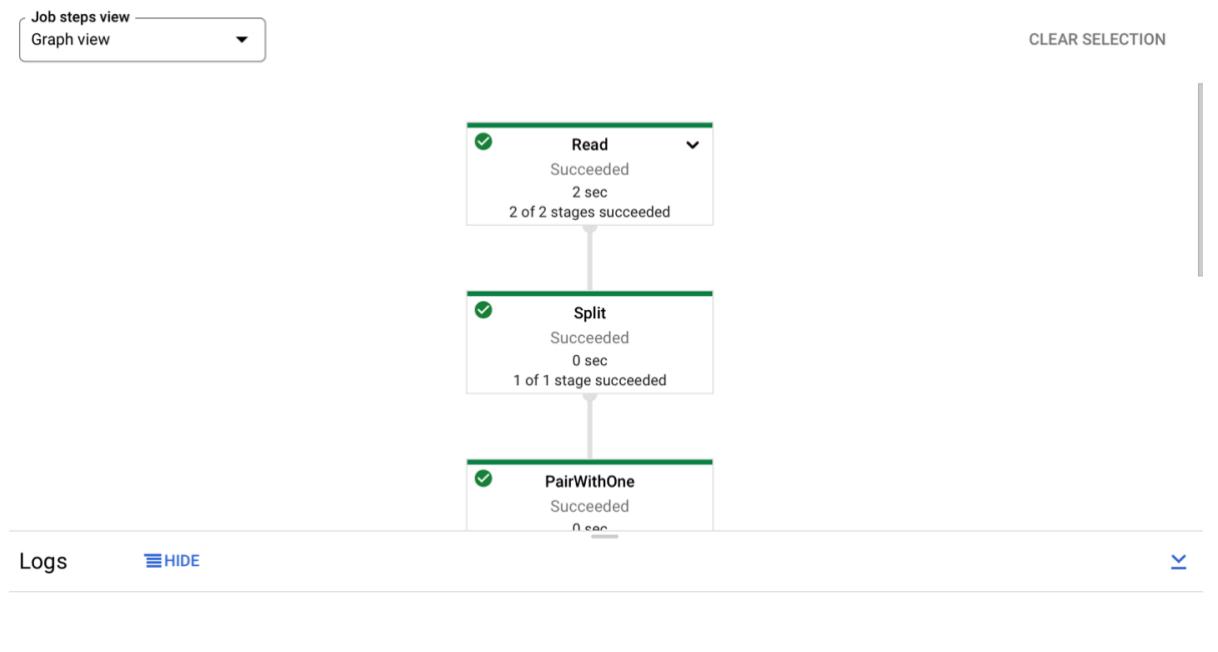
2

```
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:23.269Z: JOB_MESSAGE_BASIC: Executing operation GroupAndSum/GroupByKey/Close
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:24.111Z: JOB_MESSAGE_BASIC: Finished operation GroupAndSum/GroupByKey/Close
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:24.151Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/C
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:24.241Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/C
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:24.277Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/GroupByKey/Session" mater
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:24.319Z: JOB_MESSAGE_BASIC: Executing operation GroupAndSum/GroupByKey/Read/GroupA
pl/WindowInto(WindowIntoFn)+Write/Write/WriteImpl/WriteBundles+Write/Write/WriteImpl/Pair+Write/Write/WriteImpl/GroupByKey/Write
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:36.114Z: JOB_MESSAGE_BASIC: Finished operation GroupAndSum/GroupByKey/Read/GroupAr
l/WindowInto(WindowIntoFn)+Write/Write/WriteImpl/WriteBundles+Write/Write/WriteImpl/GroupByKey/Write
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:36.156Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/C
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:37.682Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/C
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:37.700Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/R
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.068Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/R
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.086Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/GroupByKey/Re
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.105Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrit
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.113Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/PrefFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.161Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrit
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.174Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/PrefFinalize/W
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.180Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/FinalizeWrite/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.192Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/PrefFinalize/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:40.210Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/PrefFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.414Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/PrefFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.432Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/PrefFinalize.None" materia
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.449Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/PrefFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.506Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrit
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.524Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/PrefFinalizeWrite/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:43.543Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/PrefFinalizeWrit
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:46.122Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:46.141Z: JOB_MESSAGE_DEBUG: Executing success step success32
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:46.163Z: JOB_MESSAGE_DETAILED: Cleaning up.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:46.482Z: JOB_MESSAGE_DEBUG: Starting worker pool teardown.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:13:46.490Z: JOB_MESSAGE_BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:14:29.566Z: JOB_MESSAGE_DETAILED: Autoscaling: Resized worker pool from 1 to 0.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:14:29.585Z: JOB_MESSAGE_BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:14:29.595Z: JOB_MESSAGE_DEBUG: Tearing down pending resources...
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2023-02-10_13_09_53-11709624175540378258 is in state JOB_STATE_DONE
abduhnuta@cloudshell-(projectmilestone2)$
```

View status of the job

Jobs	CREATE JOB FROM TEMPLATE		ENABLE SORTING		REFRESH	LEARN			
Running	Filter		?	!					
Name	Type	End time	Elapsed time	Start time	Status	SDK version	ID	Region	Insights
beamapp-abdulbutta7-0210210951-549111-09fe7ab	Batch	Feb 10, 2023, 4:14:34 PM	4 min 39 sec	Feb 10, 2023, 4:09:55 PM	Succeeded	2.44.0	2023-02-10_13_09_53-11709624175540378258	northamerica-northeast	1 INSIGHT

View the dataflow graph of each process



View the job logs

The screenshot shows the Google Cloud Platform Job Logs interface. At the top, there are tabs for 'JOB LOGS', 'WORKER LOGS', and 'DIAGNOSTICS'. A dropdown menu under 'Severity' is set to 'Info'. Below the tabs is a search bar with the placeholder 'Search all fields and values'. To the right of the search bar are several filter and sorting icons, including 'MAX TIME' and a question mark icon.

SEVERITY	TIMESTAMP	SUMMARY
> i	2023-02-10 16:13:43.415 EST	Finished operation Write/Write/WriteImpl/PreFinalize
> i	2023-02-10 16:13:43.451 EST	Executing operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2-Write/Write/WriteImpl/...
> i	2023-02-10 16:13:43.508 EST	Finished operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2-Write/Write/WriteImpl/F...
> i	2023-02-10 16:13:43.544 EST	Executing operation Write/Write/WriteImpl/FinalizeWrite
> i	2023-02-10 16:13:46.123 EST	Finished operation Write/Write/WriteImpl/FinalizeWrite
> i	2023-02-10 16:13:46.492 EST	Stopping worker pool...
> i	2023-02-10 16:14:29.586 EST	Worker pool stopped.

Download the output file and view the file

```
abdulbhutta7@cloudshell:~/result (projectmilestone2)$ gcloud storage cp $BUCKET/result/outputs/* ~/result
Copying gs://projectmilestone2-bucket/result/outputs-00000-of-00001 to file:///home/abdulbhutta7/result/outputs-00000-of-00001
Completed files 1/1 | 44.1kiB/44.1kiB
abdulbhutta7@cloudshell:~/result (projectmilestone2)$
```

```
abdulbhutta7@cloudshell:~/result (projectmilestone2)$ cat outputs-00000-of-00001 | more
stabbing: 1
deserves: 3
she: 79
sudden: 1
silly: 1
More: 12
'twill: 1
out: 42
believe: 7
breast: 1
sighted: 1
ones: 5
sword: 2
fabric: 1
posterity: 1
pity: 8
law: 8
thorns: 2
Alack: 2
horrible: 1
All: 8
Added: 1
distaffs: 1
any: 34
rebuke: 1
trunk: 2
```

Run the wordcount2 script

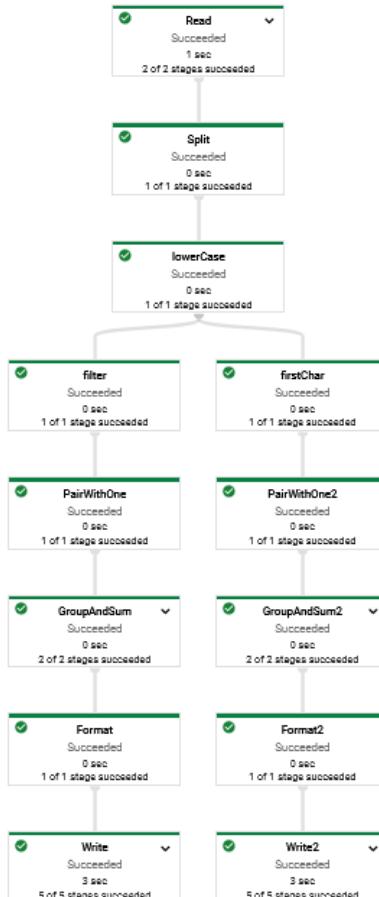
```
abdulbhutta7@cloudshell:~/SOFE4630U-MS2/wordcount (projectmilestone2)$ python wordcount2.py \
--region northamerica-northeast2 \
--runner DataflowRunner \
--project $PROJECT \
--temp_location $BUCKET/tmp/ \
--input gs://dataflow-samples/shakespeare/winterstale.txt \
--output $BUCKET/result/outputs \
--output2 $BUCKET/result/outputs2 \
--experiment use_unsupported_python_version
INFO:apache_beam.internal.gcp.auth:Setting socket default timeout to 60 seconds.
INFO:apache_beam.internal.gcp.auth:socket default timeout is 60.0 seconds.
INFO:apache_beam.runners.portability.stager:Downloading source distribution of the SDK from PyPI
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip',
e-beam==2.44.0', '--no-deps', '--no-binary', ':all:']
INFO:apache_beam.runners.portability.stager:Staging SDK sources from PyPI: dataflow_python_sdk.t
INFO:apache_beam.runners.portability.stager:Downloading binary distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip',
e-beam==2.44.0', '--no-deps', '--only-binary', ':all:', '--python-version', '39', '--implementat
linux2014_x86_64']
INFO:apache_beam.runners.portability.stager:Staging binary distribution of the SDK from PyPI: ap
.manylinux2014_x86_64.whl
INFO:apache_beam.runners.portability.stager:Uploading binary distribution of the SDK to GCP: ap...
```

```

INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:48.764Z: JOB_MESSAGE_DEBUG: Value "Write2/Write/WriteImpl/PreFinalize/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:48.782Z: JOB_MESSAGE_BASIC: Executing operation Write2/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.644Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/None
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.662Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/Extract.None" materialized
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.680Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.731Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.734Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.750Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/FinalizeWrite/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.758Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/PreFinalize/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:49.776Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:52.144Z: JOB_MESSAGE_BASIC: Finished operation Write2/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:52.162Z: JOB_MESSAGE_DEBUG: Value "Write2/Write/WriteImpl/PreFinalize.None" materialized
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:52.180Z: JOB_MESSAGE_BASIC: Executing operation Write2/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:52.234Z: JOB_MESSAGE_BASIC: Finished operation Write2/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:52.252Z: JOB_MESSAGE_DEBUG: Value "Write2/Write/WriteImpl/FinalizeWrite/View-python"
.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.270Z: JOB_MESSAGE_BASIC: Executing operation Write2/Write/WriteImpl/FinalizeWrite/View-python
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.562Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.581Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/PreFinalize.None" materialized
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.598Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.660Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.678Z: JOB_MESSAGE_DEBUG: Value "Write/Write/WriteImpl/FinalizeWrite/View-python"
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.696Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:53.842Z: JOB_MESSAGE_BASIC: Finished operation Write2/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:54.626Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:56.842Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:56.860Z: JOB_MESSAGE_DEBUG: Executing success step success56
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:57.208Z: JOB_MESSAGE_DETAILED: Cleaning up.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:57.216Z: JOB_MESSAGE_BASIC: Starting worker pool teardown.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:49:57.216Z: JOB_MESSAGE_BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:50:39.205Z: JOB_MESSAGE_DETAILED: Autoscaling: Resized worker pool from 1 to 0.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:50:39.224Z: JOB_MESSAGE_BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T21:50:39.233Z: JOB_MESSAGE_DEBUG: Tearing down pending resources...
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2023-02-10_13_46_28-663861519406990545 is in state JOB_STATE_DONE
abdulbhutta7@cloudshell:~/SOFE4630U-MS2/wordcount (projectmilestone2)$

```

DataFlow Graphs



Output of the wordcount 2 which displayes all the occurrence of the letters within the winterstale.txt book.

```
outputs-00000-of-00001 outputs2-00000-of-00001 ×
result > outputs2-00000-of-00001
  1 l: 905
  2 s: 2051
  3 j: 62
  4 b: 1450
  5 o: 1241
  6 f: 1017
  7 u: 255
  8 q: 82
  9 c: 1004
 10 g: 577
 11 r: 305
 12 n: 881
 13 p: 951
 14 t: 3664
 15 y: 869
 16 i: 1813
 17 ': 156
 18 z: 1
 19 e: 424
 20 m: 1585
 21 v: 112
 22 k: 234
 23 w: 1633
 24 a: 2423
 25 h: 1762
 26 d: 724
 27
```

D1: Data flow Processing the MNIST database from BigQuery

Create a BigQuery Table and upload the MNIST csv data

Create table

Source

Create table from —
Upload
Select file * mnist.csv
File format CSV

Destination

Project * projectmilestone2
Dataset * MNIST
Table * Images
Table type Native table

Schema

Auto detect
Schema will be automatically generated.

Partition and cluster settings

Partitioning No partitioning
Clustering order

Clustering order determines the sort order of the data. Clustering can be used on both partitioned and non-partitioned tables.

Buttons

CREATE TABLE CANCEL

Upload the model folder

projectmilestone2-bucket

Location Storage class Public access Protection
us (multiple regions in United States) Standard Not public None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY NEW

Buckets > projectmilestone2-bucket

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS DOWNLOAD DELETE

Filter by name prefix only ▾ Filter objects and folders Show deleted data

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption
model/	—	Folder	—	—	—	—	—	—
result/	—	Folder	—	—	—	—	—	—
tmp/	—	Folder	—	—	—	—	—	—

one2) + * 300-M82/wordcount [projectmilestone2]\$ gcloud storage cp \$BUCKET/result/outputs2* ~/result
bucket/result/outputs2-00000-of-00001 to file:///home/abdulbhutta7/result/outputs2-00000-of-00001
189.0B 300-M82/wordcount [projectmilestone2]: 3 files successfully uploaded

Uploads and ProjectMilestone2 operations

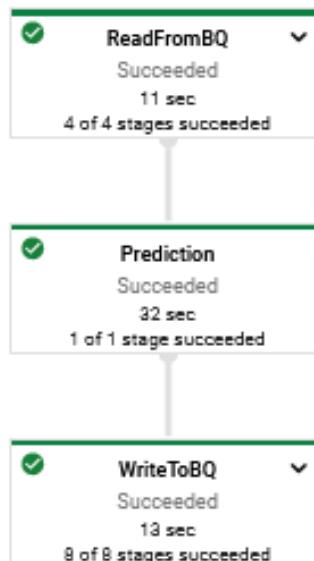
Uploading 3 files

- export_index Complete
- export_meta Complete
- export_data-00000-of-00001 Complete

Run the DataFlow job

```
abdulbhutta7@cloudshell:~/SOFE4630U-MS2/wordcount [projectmilestone2]$ cd ~/SOFE4630U-MS2/mnist
python mnistBQ.py \
    --runner DataflowRunner \
    --project $PROJECT \
    --staging_location $BUCKET/staging \
    --temp_location $BUCKET/temp \
    --setup_file ./setup.py \
    --input $PROJECT:MNIST.Images \
    --region northamerica-northeast2 \
    --experiment use_unsupported_python_version
2023-02-10 22:05:30.603102: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlsym: libcudart.so.11.0: cannot open shared object file: No such file or directory
2023-02-10 22:05:30.603150: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlsym error if you do not have a GPU set up on your machine.
/home/abdulbhutta7/.local/lib/python3.9/site-packages/apache_beam/io/gcp/bigquery.py:2485: BeamDeprecationWarning: options is deprecated since First stable release. References to <pipeline>.options will not be supported
    temp_location = pcoll_pipeline.options.view_as(
/home/abdulbhutta7/.local/lib/python3.9/site-packages/apache_beam/io/gcp/bigquery_file_loads.py:1142: BeamDeprecationWarning: options is deprecated since First stable release. References to <pipeline>.options will not be supported
        self.project = self.project or p.options.view_as(GoogleCloudOptions).project
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', 'setup.py', 'sdist', '--dist-dir', '/tmp/tmpf186ks51']
WARNING: sdist: Standard file not found: should have one of README, README.rst, README.txt, README.md
INFO:apache_beam.runners.portability.stager:Downloading source distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpf186ks51', 'apache-beam==2.44.0', '--no-deps', '--no-binary', ':all:']
INFO:apache_beam.runners.portability.stager:Staging SDK sources from PyPi: dataflow_python_sdk.tar
INFO:apache_beam.runners.portability.stager:Downloading binary distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpf186ks51', 'apache-beam==2.44.0', '--no-deps', '--no-binary', ':all:', '--python-version', '39', '--implementation', 'cp', '--abi', 'cp39', '--platform', 'manylinux2014_x86_64']
INFO:apache_beam.runners.portability.stager:Staging binary distribution of the SDK from PyPi: apache_beam-2.44.0-cp39-manylinux2014_x86_64.whl
INFO:root:Default Python SDK image for environment is apache/beam-dataflow:9_sdk:2.44.0
INFO:root:Using provided Python SDK container image: gcr.io/cloud-dataflow/v1beta3/python39:2.44.0
INFO:root:Python SDK container image set to "gcr.io/cloud-dataflow/v1beta3/python39:2.44.0" for Docker environment
INFO:apache_beam.runners.portability.fn_api_runner.translations:===== <function pack_combiners at 0x7fe8843d8670> =====
INFO:apache_beam.runners.portability.stager:Reading from BigQueryBatchFileLoads/RemoveTempTables/GetTableNames/Keys+WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/Delete
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.536Z: JOB_MESSAGE_BASIC: Finished operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Read+WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/GetTableNames/Keys+WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/Delete
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.536Z: JOB_MESSAGE_BASIC: Executing success step success42
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.577Z: JOB_MESSAGE_DETAILED: Cleaning up.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.729Z: JOB_MESSAGE_DEBUG: Starting worker pool teardown.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.737Z: JOB_MESSAGE_BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.910Z: JOB_MESSAGE_DETAILED: Autoscaling: Resized worker pool from 1 to 0.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.926Z: JOB_MESSAGE_BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.938Z: JOB_MESSAGE_DEBUG: Tearing down pending resources...
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2023-02-10_14_05_43-17038938066430139014 is in state JOB_STATE_DONE
abdulbhutta7@cloudshell:~/SOFE4630U-MS2/mnist [projectmilestone2]$
```

DataFlow Graph



Run a query

The screenshot shows a data exploration interface with a sidebar on the left containing an 'Explorer' section. The 'Predict' item under the 'MNIST' category is selected and highlighted with a blue background. The main area is a query editor titled 'Run a query'. The toolbar includes buttons for RUN, SAVE, SHARE, SCHEDULE, and MORE. A search bar at the top says 'Type to search'. Below it, a message says 'Viewing all resources. Show starred resources only.' A code editor window displays the following SQL query:

```
1 SELECT * FROM `projectmilestone2.MNIST.Predict` LIMIT 1000
```

View the query output

The screenshot shows the same data exploration interface after running the query. The status bar at the top right indicates 'Query completed'. The 'Query results' section displays the output of the query. The table has columns: Row, ID, P0, P1, P2, P3, P4, P5, and F. The data is as follows:

Row	ID	P0	P1	P2	P3	P4	P5	F
1	8448	1.0	6.36827560...	8.86756154...	1.02936153...	2.83226472...	8.07854651...	
2	8458	1.0	3.10568838...	2.91970267...	2.80161180...	3.93055845...	1.99952165...	
3	8481	1.0	1.54292892...	3.89899739...	9.56256016...	3.76161605...	9.0071666...	
4	8513	1.0	1.22901261...	1.08167724...	8.47215613...	8.38592389...	6.43156949...	
5	8535	1.0	2.60132600...	2.09912220...	5.36175010...	1.95211673...	6.28202177...	
6	8567	1.0	3.29208692...	2.94471069...	1.42370195...	1.36069771...	7.09583149...	
7	8595	1.0	6.01713552...	2.34865377...	1.25974670...	5.49075459...	3.46307184...	
8	8619	1.0	1.26582181...	3.31328936...	1.96425627...	1.84887287...	1.42652884...	
9	8657	1.0	3.80602735...	1.51620338...	3.00870602...	1.15719375...	4.24890201...	
10	8671	1.0	6.07441207...	1.03449799...	3.92728221...	1.18613041...	2.18257893...	
11	8730	1.0	1.91185920...	1.07626951...	1.35363810...	1.60220957...	7.04176571...	
12	8731	1.0	2.05738768...	2.21387885...	2.28982523...	4.91702858...	2.14701867...	
13	8763	1.0	1.56876888...	1.06087210...	1.03443161...	2.07287447...	6.93942299...	
14	8798	1.0	1.81124273...	1.01470183...	1.76658262...	1.66800864...	6.49116448...	
15	8808	1.0	7.37771280...	2.29562431...	2.37855169...	1.31236287...	8.25805691...	
16	8845	1.0	5.74639294...	1.13828200...	6.81174409...	3.10233197...	1.82059705...	

D1: Data flow Processing the MNIST database from Pub/Sub

Create two topics

Topics	CREATE TOPIC	DELETE	
LIST	METRICS		
<input type="checkbox"/> Filter Filter topics			
<input type="checkbox"/> Topic ID ↑	Encryption key	Topic name	Retention
<input type="checkbox"/> mnist_image	Google-managed	projects/projectmilestone2/topics/mnist_image	[]
<input type="checkbox"/> mnist_predic	Google-managed	projects/projectmilestone2/topics/mnist_predic	[]

Run the DataFlow jobs

```
abdulhutta@cloudshell:~/SOFE4630U-MS2/wordcount [projectmilestone2]$ cd ~/SOFE4630U-MS2/mnist
python mnistBQ.py \
--runner DataflowRunner \
--project $PROJECT \
--staging_location $BUCKET/staging \
--temp_location $BUCKET/temp \
--model $BUCKET/model \
--setup_file ./setup.py \
--image gcr.io/ucb-mlops/MNIST_Predict \
--output $PROJECT$BUCKET/Predict \
--region northamerica-northeast2 \
--experiment use unsupported python version
2023-02-10 22:05:30.603102: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot open shared object file: No such file or directory
2023-02-10 22:05:30.603150: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU set up on your machine.
/home/abdulhutta/.local/lib/python3.9/site-packages/apache_beam/io/gcp/bigquery.py:248: BeamDeprecationWarning: options is deprecated since First stable release. References to <pipeline>.options will not be supported
temp_location = pccol.pipeline.options.view_as(
/home/abdulhutta/.local/lib/python3.9/site-packages/apache_beam/io/gcp/bigquery_file_loads.py:1142: BeamDeprecationWarning: options is deprecated since First stable release. References to <pipeline>.options will not be supported
self.project = self.project or options.view_as(GoogleCloudOptions).project
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', 'setup.py', '--dist-dir', '/tmp/tmpfi86ks51']
WARNING: distdir not found: should have one of README, README.rst, README.txt, README.md

INFO:apache_beam.runners.portability.stager:Downloading source distribution of the SDK from PyPI
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpfi86ks51', 'apache-beam==2.44.0', '--no-deps', '--no-binary', ':all:']
INFO:apache_beam.runners.portability.stager:Staging SDK sources from PyPI: dataflow_python_sdk.tar
INFO:apache_beam.runners.portability.stager:Downloading binary distribution of the SDK from PyPI
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download', '--dest', '/tmp/tmpfi86ks51', 'apache-beam==2.44.0', '--no-deps', '--no-binary', ':all:', '--python-version', '39', '--implementation', 'cp', '--abi', 'cp39', '--platform', 'manylinux2014_x86_64']
INFO:apache_beam.runners.portability.stager:Staging binary distribution of the SDK from PyPI: apache_beam-2.44.0-cp39-cp39-manylinux2_17_x86_64.manylinux2014_x86_64.whl
INFO:root:Default Python SDK image for environment is apache_beam_python3.9_sdk:2.44.0

INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.613Z: JOB_MESSAGE_DEBUG: Value "WriteToBQ/BigQueryBatchFileLoads/ParDo(TriggerCopyJobs)/ParDo(TriggerCopyJobs)_TriggerDeleteTempTables" materialized.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.652Z: JOB_MESSAGE_BASIC: Executing operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue/View+python_side_input0=WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.704Z: JOB_MESSAGE_BASIC: Finished operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.742Z: JOB_MESSAGE_BASIC: Value "WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue/View+python_side_input0=WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue.out" materialized.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.785Z: JOB_MESSAGE_BASIC: Executing operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Create
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.962Z: JOB_MESSAGE_BASIC: Finished operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Create
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.980Z: JOB_MESSAGE_DEBUG: Value "WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Session" materialized.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:38.982Z: JOB_MESSAGE_BASIC: Executing operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue/write+WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Write
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.376Z: JOB_MESSAGE_BASIC: Finished operation WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/AddUselessValue+WriteToBQ/BigQueryBatchFileLoads/RemoveTempTables/DeduplicateTables/Write
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.553Z: JOB_MESSAGE_DEBUG: Executing success step success42
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.577Z: JOB_MESSAGE_DETAILED: Cleaning up.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.729Z: JOB_MESSAGE_DEBUG: Starting worker pool teardown.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:11:39.737Z: JOB_MESSAGE_BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.910Z: JOB_MESSAGE_DETAILED: Autoscaling: Resized worker pool from 1 to 0.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.926Z: JOB_MESSAGE_BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:2023-02-10T22:12:21.938Z: JOB_MESSAGE_DEBUG: Tearing down pending resources...
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2023-02-10 14:05:43-17038938066430139014 is in state JOB_STATE_DONE
abdulhutta@cloudshell:~/SOFE4630U-MS2/mnist [projectmilestone2]$ 
```

Edit the Consumer python script

```
consumerMnistPubSup.py > ...  
import os  
from google.cloud import pubsub_v1      #pip install google-cloud-pubsub  
os.environ["GOOGLE_APPLICATION_CREDENTIALS"]="/Users/abdulbhutta/Downloads/credentials.json"  
project_id = "projectmilestone2"  
subscription_id = "projects/projectmilestone2/subscriptions/mnist_predic-sub"  
  
subscriber = pubsub_v1.SubscriberClient()  
subscription_path = subscriber.subscription_path(project_id, subscription_id)  
  
import json  
def callback(message: pubsub_v1.subscriber.message.Message) -> None:  
    print(f"Received {json.loads(message)}."  
    print(f"Received {json.loads(message.data)}."  
    message.ack()  
  
streaming_pull_future = subscriber.subscribe(subscription_path, callback=callback)  
print(f"Listening for messages on {subscription_path}..\n")  
  
with subscriber:  
    streaming_pull_future.result()
```

Edit the producer python script

```
producerMnistPubSup.py > ...  
# environment variable setup for private key file  
import os  
import pandas as pd  #pip install pandas  ##to install  
  
from google.cloud import pubsub_v1    #pip install google-cloud-pubsub  ##to install  
import time  
import json;  
import io;  
  
# TODO : fill the credential json and project id  
os.environ["GOOGLE_APPLICATION_CREDENTIALS"]="/Users/abdulbhutta/Downloads/credentials.json"  
project_id = "projectmilestone2"  
topic_id = "projects/projectmilestone2/subscriptions/mnist_image-sub"  
  
publisher = pubsub_v1.PublisherClient()  
# The `topic_path` method creates a fully qualified identifier  
# in the form 'projects/{project_id}/topics/{topic_id}'  
topic_path = publisher.topic_path(project_id, topic_id)  
  
df=pd.read_csv('mnist.csv')  
  
for row in df.iterrows():  
    value=row[1].to_dict()  
    future = publisher.publish(topic_path, json.dumps(value).encode('utf-8'));  
    print("Image with key "+str(value["ID"])+ " is sent")  
    time.sleep(0.1);
```

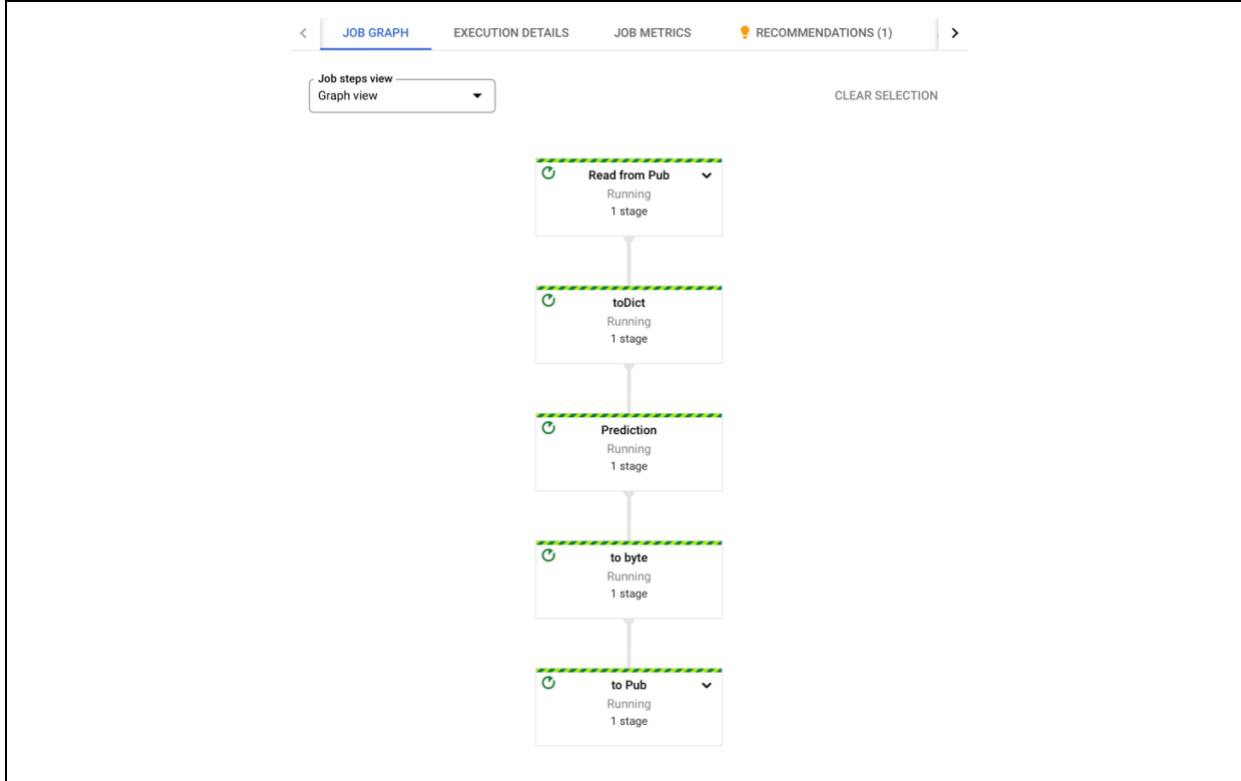
Execute producer.py script which will produce image data to the mnist_image topic

```
data -- python producerMnistPubSup.py -- 122x70
Image with key 1783 is sent
Image with key 1783 is sent
Image with key 1784 is sent
Image with key 1785 is sent
Image with key 1786 is sent
Image with key 1787 is sent
Image with key 1788 is sent
Image with key 1789 is sent
Image with key 1790 is sent
Image with key 1791 is sent
Image with key 1792 is sent
Image with key 1793 is sent
Image with key 1794 is sent
Image with key 1795 is sent
Image with key 1796 is sent
Image with key 1797 is sent
Image with key 1798 is sent
Image with key 1799 is sent
Image with key 1800 is sent
Image with key 1801 is sent
Image with key 1802 is sent
Image with key 1803 is sent
Image with key 1804 is sent
Image with key 1805 is sent
Image with key 1806 is sent
Image with key 1807 is sent
Image with key 1808 is sent
Image with key 1809 is sent
Image with key 1810 is sent
Image with key 1811 is sent
```

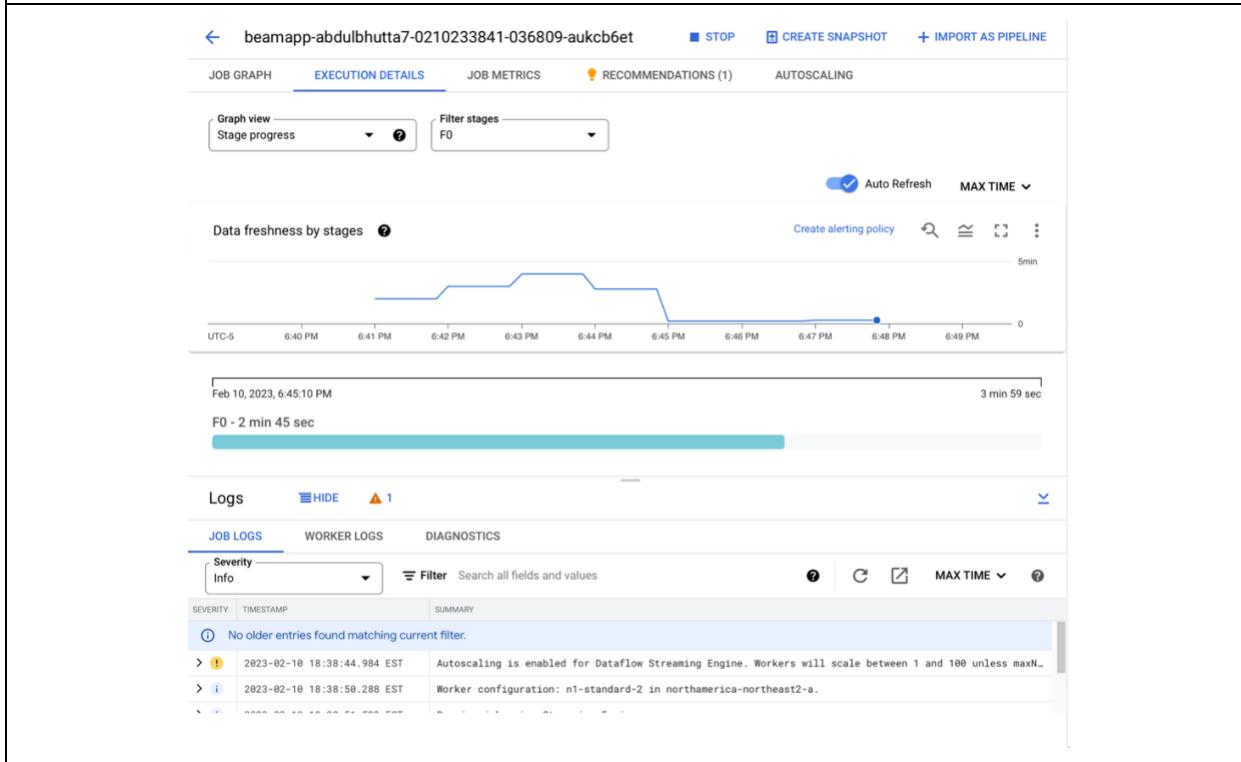
Execute the consumer.py script and view the results from mnist_predict topic

```
data -- python consumerMnistPubSup.py -- 121x70
Received ('ID': 1925, 'P1': 4.473367341698356e-14, 'P2': 3.109128954292828e-08, 'P3': 1.264699750251111e-06, 'P4': 6.012123131399129394e-13, 'P5': 4.0261835217845e-13, 'P6': 1.758225951926123e-18, 'P7': 0.9999986886978149, 'P8': 1.1785444587051996e-12, 'P9': 2.6753932402812097e-09, 'P10': 1.2595519285692582e-11, 'P11': 2.5873106324079446e-06, 'P12': 1.417233394627252e-12, 'P13': 1.0121997641896558e-13, 'P14': 8.474864998154693e-06, 'P15': 0.9999872146530151, 'P16': 7.641338925168384e-06, 'P17': 9.531394385745898e-07, 'P18': 2.1876841791832324e-13, 'P19': 6.81688403780446e-08, 'P20': 7.736618945719488e-17, 'P21': 5.4787298911189852e-14, 'P22': 4.754682758623119e-12, 'P23': 0.99999487409095493, 'P24': 6.138996814643421e-01, 'P25': 6.075506289325471e-10, 'P26': 2.777016503571537e-16, 'P27': 1.4537872017328136e-09, 'P28': 8.138996814643421e-01, 'P29': 6.075506289325471e-10, 'P30': 1.251000026511854824e-10, 'P31': 0.9999749668491943, 'P32': 2.788990149931783e-09, 'P33': 0.10738932811317058e-09, 'P34': 9.546823775963276e-07, 'P35': 1.742562008580286e-08, 'P36': 2.4098137525993e-08, 'P37': 0.99998740445234995e-13, 'P38': 1.94089361831569e-12, 'P39': 5.87951023782722e-09, 'P40': 0.999997615814269, 'P41': 8.803308876898804e-11, 'P42': 2.395808775745863e-06, 'P43': 1.63056400838595352e-14, 'P44': 8.929051942274668e-10, 'P45': 3.173438229708174e-08, 'P46': 2.50904391890348643e-09, 'P47': 0.999987615814269, 'P48': 6.138996814643421e-01, 'P49': 5.150683536158467e-07, 'P50': 0.999910831451416, 'P51': 3.406688077932364e-07, 'P52': 3.1955858098564605e-11, 'P53': 1.07596130423837276e-14, 'P54': 3.459594264711152e-13, 'P55': 8.82655440364031e-09, 'P56': 0.999987489117968491943, 'P57': 2.85256846426225984e-08, 'P58': 1.3078392811317058e-09, 'P59': 9.546823775963276e-07, 'P60': 1.742562008580286e-08, 'P61': 2.4098137525993e-08, 'P62': 0.99998740445234995e-13, 'P63': 1.94089361831569e-12, 'P64': 5.87951023782722e-09, 'P65': 0.999997615814269, 'P66': 8.803308876898804e-11, 'P67': 2.395808775745863e-06, 'P68': 1.63056400838595352e-14, 'P69': 8.929051942274668e-10, 'P70': 3.173438229708174e-08, 'P71': 2.50904391890348643e-09, 'P72': 0.999987615814269, 'P73': 6.138996814643421e-01, 'P74': 5.150683536158467e-07, 'P75': 0.999910831451416, 'P76': 3.406688077932364e-07, 'P77': 3.1955858098564605e-11, 'P78': 1.07596130423837276e-14, 'P79': 3.459594264711152e-13, 'P80': 8.82655440364031e-09, 'P81': 0.999987489117968491943, 'P82': 2.85256846426225984e-08, 'P83': 1.3078392811317058e-09, 'P84': 9.546823775963276e-07, 'P85': 1.742562008580286e-08, 'P86': 2.4098137525993e-08, 'P87': 0.99998740445234995e-13, 'P88': 1.94089361831569e-12, 'P89': 5.87951023782722e-09, 'P90': 0.999997615814269, 'P91': 8.803308876898804e-11, 'P92': 2.395808775745863e-06, 'P93': 1.63056400838595352e-14, 'P94': 8.929051942274668e-10, 'P95': 3.173438229708174e-08, 'P96': 2.50904391890348643e-09, 'P97': 0.999987615814269, 'P98': 6.138996814643421e-01, 'P99': 5.150683536158467e-07, 'P100': 0.999910831451416, 'P101': 3.406688077932364e-07, 'P102': 3.1955858098564605e-11, 'P103': 1.07596130423837276e-14, 'P104': 3.459594264711152e-13, 'P105': 8.82655440364031e-09, 'P106': 0.999987489117968491943, 'P107': 2.85256846426225984e-08, 'P108': 1.3078392811317058e-09, 'P109': 9.546823775963276e-07, 'P110': 1.742562008580286e-08, 'P111': 2.4098137525993e-08, 'P112': 0.99998740445234995e-13, 'P113': 1.94089361831569e-12, 'P114': 5.87951023782722e-09, 'P115': 0.999997615814269, 'P116': 8.803308876898804e-11, 'P117': 2.395808775745863e-06, 'P118': 1.63056400838595352e-14, 'P119': 8.929051942274668e-10, 'P120': 3.173438229708174e-08, 'P121': 2.50904391890348643e-09, 'P122': 0.999987615814269, 'P123': 6.138996814643421e-01, 'P124': 5.150683536158467e-07, 'P125': 0.999910831451416, 'P126': 3.406688077932364e-07, 'P127': 3.1955858098564605e-11, 'P128': 1.07596130423837276e-14, 'P129': 3.459594264711152e-13, 'P130': 8.82655440364031e-09, 'P131': 0.999987489117968491943, 'P132': 2.85256846426225984e-08, 'P133': 1.3078392811317058e-09, 'P134': 9.546823775963276e-07, 'P135': 1.742562008580286e-08, 'P136': 2.4098137525993e-08, 'P137': 0.99998740445234995e-13, 'P138': 1.94089361831569e-12, 'P139': 5.87951023782722e-09, 'P140': 0.999997615814269, 'P141': 8.803308876898804e-11, 'P142': 2.395808775745863e-06, 'P143': 1.63056400838595352e-14, 'P144': 8.929051942274668e-10, 'P145': 3.173438229708174e-08, 'P146': 2.50904391890348643e-09, 'P147': 0.999987615814269, 'P148': 6.138996814643421e-01, 'P149': 5.150683536158467e-07, 'P150': 0.999910831451416, 'P151': 3.406688077932364e-07, 'P152': 3.1955858098564605e-11, 'P153': 1.07596130423837276e-14, 'P154': 3.459594264711152e-13, 'P155': 8.82655440364031e-09, 'P156': 0.999987489117968491943, 'P157': 2.85256846426225984e-08, 'P158': 1.3078392811317058e-09, 'P159': 9.546823775963276e-07, 'P160': 1.742562008580286e-08, 'P161': 2.4098137525993e-08, 'P162': 0.99998740445234995e-13, 'P163': 1.94089361831569e-12, 'P164': 5.87951023782722e-09, 'P165': 0.999997615814269, 'P166': 8.803308876898804e-11, 'P167': 2.395808775745863e-06, 'P168': 1.63056400838595352e-14, 'P169': 8.929051942274668e-10, 'P170': 3.173438229708174e-08, 'P171': 2.50904391890348643e-09, 'P172': 0.999987615814269, 'P173': 6.138996814643421e-01, 'P174': 5.150683536158467e-07, 'P175': 0.999910831451416, 'P176': 3.406688077932364e-07, 'P177': 3.1955858098564605e-11, 'P178': 1.07596130423837276e-14, 'P179': 3.459594264711152e-13, 'P180': 8.82655440364031e-09, 'P181': 0.999987489117968491943, 'P182': 2.85256846426225984e-08, 'P183': 1.3078392811317058e-09, 'P184': 9.546823775963276e-07, 'P185': 1.742562008580286e-08, 'P186': 2.4098137525993e-08, 'P187': 0.99998740445234995e-13, 'P188': 1.94089361831569e-12, 'P189': 5.87951023782722e-09, 'P190': 0.999997615814269, 'P191': 8.803308876898804e-11, 'P192': 2.395808775745863e-06, 'P193': 1.63056400838595352e-14, 'P194': 8.929051942274668e-10, 'P195': 3.173438229708174e-08, 'P196': 2.50904391890348643e-09, 'P197': 0.999987615814269, 'P198': 6.138996814643421e-01, 'P199': 5.150683536158467e-07, 'P200': 0.999910831451416, 'P201': 3.406688077932364e-07, 'P202': 3.1955858098564605e-11, 'P203': 1.07596130423837276e-14, 'P204': 3.459594264711152e-13, 'P205': 8.82655440364031e-09, 'P206': 0.999987489117968491943, 'P207': 2.85256846426225984e-08, 'P208': 1.3078392811317058e-09, 'P209': 9.546823775963276e-07, 'P210': 1.742562008580286e-08, 'P211': 2.4098137525993e-08, 'P212': 0.99998740445234995e-13, 'P213': 1.94089361831569e-12, 'P214': 5.87951023782722e-09, 'P215': 0.999997615814269, 'P216': 8.803308876898804e-11, 'P217': 2.395808775745863e-06, 'P218': 1.63056400838595352e-14, 'P219': 8.929051942274668e-10, 'P220': 3.173438229708174e-08, 'P221': 2.50904391890348643e-09, 'P222': 0.999987615814269, 'P223': 6.138996814643421e-01, 'P224': 5.150683536158467e-07, 'P225': 0.999910831451416, 'P226': 3.406688077932364e-07, 'P227': 3.1955858098564605e-11, 'P228': 1.07596130423837276e-14, 'P229': 3.459594264711152e-13, 'P230': 8.82655440364031e-09, 'P231': 0.999987489117968491943, 'P232': 2.85256846426225984e-08, 'P233': 1.3078392811317058e-09, 'P234': 9.546823775963276e-07, 'P235': 1.742562008580286e-08, 'P236': 2.4098137525993e-08, 'P237': 0.99998740445234995e-13, 'P238': 1.94089361831569e-12, 'P239': 5.87951023782722e-09, 'P240': 0.999997615814269, 'P241': 8.803308876898804e-11, 'P242': 2.395808775745863e-06, 'P243': 1.63056400838595352e-14, 'P244': 8.929051942274668e-10, 'P245': 3.173438229708174e-08, 'P246': 2.50904391890348643e-09, 'P247': 0.999987615814269, 'P248': 6.138996814643421e-01, 'P249': 5.150683536158467e-07, 'P250': 0.999910831451416, 'P251': 3.406688077932364e-07, 'P252': 3.1955858098564605e-11, 'P253': 1.07596130423837276e-14, 'P254': 3.459594264711152e-13, 'P255': 8.82655440364031e-09, 'P256': 0.999987489117968491943, 'P257': 2.85256846426225984e-08, 'P258': 1.3078392811317058e-09, 'P259': 9.546823775963276e-07, 'P260': 1.742562008580286e-08, 'P261': 2.4098137525993e-08, 'P262': 0.99998740445234995e-13, 'P263': 1.94089361831569e-12, 'P264': 5.87951023782722e-09, 'P265': 0.999997615814269, 'P266': 8.803308876898804e-11, 'P267': 2.395808775745863e-06, 'P268': 1.63056400838595352e-14, 'P269': 8.929051942274668e-10, 'P270': 3.173438229708174e-08, 'P271': 2.50904391890348643e-09, 'P272': 0.999987615814269, 'P273': 6.138996814643421e-01, 'P274': 5.150683536158467e-07, 'P275': 0.999910831451416, 'P276': 3.406688077932364e-07, 'P277': 3.1955858098564605e-11, 'P278': 1.07596130423837276e-14, 'P279': 3.459594264711152e-13, 'P280': 8.82655440364031e-09, 'P281': 0.999987489117968491943, 'P282': 2.85256846426225984e-08, 'P283': 1.3078392811317058e-09, 'P284': 9.546823775963276e-07, 'P285': 1.742562008580286e-08, 'P286': 2.4098137525993e-08, 'P287': 0.99998740445234995e-13, 'P288': 1.94089361831569e-12, 'P289': 5.87951023782722e-09, 'P290': 0.999997615814269, 'P291': 8.803308876898804e-11, 'P292': 2.395808775745863e-06, 'P293': 1.63056400838595352e-14, 'P294': 8.929051942274668e-10, 'P295': 3.173438229708174e-08, 'P296': 2.50904391890348643e-09, 'P297': 0.999987615814269, 'P298': 6.138996814643421e-01, 'P299': 5.150683536158467e-07, 'P300': 0.999910831451416, 'P301': 3.406688077932364e-07, 'P302': 3.1955858098564605e-11, 'P303': 1.07596130423837276e-14, 'P304': 3.459594264711152e-13, 'P305': 8.82655440364031e-09, 'P306': 0.999987489117968491943, 'P307': 2.85256846426225984e-08, 'P308': 1.3078392811317058e-09, 'P309': 9.546823775963276e-07, 'P310': 1.742562008580286e-08, 'P311': 2.4098137525993e-08, 'P312': 0.99998740445234995e-13, 'P313': 1.94089361831569e-12, 'P314': 5.87951023782722e-09, 'P315': 0.999997615814269, 'P316': 8.803308876898804e-11, 'P317': 2.395808775745863e-06, 'P318': 1.63056400838595352e-14, 'P319': 8.929051942274668e-10, 'P320': 3.173438229708174e-08, 'P321': 2.50904391890348643e-09, 'P322': 0.999987615814269, 'P323': 6.138996814643421e-01, 'P324': 5.150683536158467e-07, 'P325': 0.999910831451416, 'P326': 3.406688077932364e-07, 'P327': 3.1955858098564605e-11, 'P328': 1.07596130423837276e-14, 'P329': 3.459594264711152e-13, 'P330': 8.82655440364031e-09, 'P331': 0.999987489117968491943, 'P332': 2.85256846426225984e-08, 'P333': 1.3078392811317058e-09, 'P334': 9.546823775963276e-07, 'P335': 1.742562008580286e-08, 'P336': 2.4098137525993e-08, 'P337': 0.99998740445234995e-13, 'P338': 1.94089361831569e-12, 'P339': 5.87951023782722e-09, 'P340': 0.999997615814269, 'P341': 8.803308876898804e-11, 'P342': 2.395808775745863e-06, 'P343': 1.63056400838595352e-14, 'P344': 8.929051942274668e-10, 'P345': 3.173438229708174e-08, 'P346': 2.50904391890348643e-09, 'P347': 0.999987615814269, 'P348': 6.138996814643421e-01, 'P349': 5.150683536158467e-07, 'P350': 0.999910831451416, 'P351': 3.406688077932364e-07, 'P352': 3.1955858098564605e-11, 'P353': 1.07596130423837276e-14, 'P354': 3.459594264711152e-13, 'P355': 8.82655440364031e-09, 'P356': 0.999987489117968491943, 'P357': 2.85256846426225984e-08, 'P358': 1.3078392811317058e-09, 'P359': 9.546823775963276e-07, 'P360': 1.742562008580286e-08, 'P361': 2.4098137525993e-08, 'P362': 0.99998740445234995e-13, 'P363': 1.94089361831569e-12, 'P364': 5.87951023782722e-09, 'P365': 0.999997615814269, 'P366': 8.803308876898804e-11, 'P367': 2.395808775745863e-06, 'P368': 1.63056400838595352e-14, 'P369': 8.929051942274668e-10, 'P370': 3.173438229708174e-08, 'P371': 2.50904391890348643e-09, 'P372': 0.999987615814269, 'P373': 6.138996814643421e-01, 'P374': 5.150683536158467e-07, 'P
```

View DataFlow Graph



View the job status



Design: SmartMeter

Create producer.py script to send the measurements data to a topic

```
consumer.py  producer.py 1
producer.py > ...
1 import os
2 import json
3 import time
4 import random
5 import numpy as np
6 from google.cloud import pubsub_v1
7
8 #Set the environment variable to the credentials file
9 credentials_path = os.path.join('/Users/abdulbhutta/Desktop/Cloud Computing/Project_Milestone2/Design/credentials.json')
10 os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = credentials_path
11
12 #Create a publisher client
13 publisher = pubsub_v1.PublisherClient()
14 topic_path = 'projects/projectmilestone2/topics/smartmeter_readings'
15
16 #device normal distributions profile used to generate random data
17 DEVICE_PROFILES = {
18     "boston": {'temp': (51.3, 17.7), 'humd': (77.4, 18.7), 'pres': (1.019, 0.091) },
19     "denver": {'temp': (49.5, 19.3), 'humd': (33.0, 13.9), 'pres': (1.512, 0.341) },
20     "losang": {'temp': (63.9, 11.7), 'humd': (62.8, 21.8), 'pres': (1.215, 0.201) },
21 }
22 profileNames=["boston","denver","losang"];
23
24 #Key and value for the message
25 record_key=0
26 while(True):
27     try:
28         profile_name = profileNames[random.randint(0, 2)];
29         profile = DEVICE_PROFILES[profile_name]
30         # get random values within a normal distribution of the value
31         temp = max(0, np.random.normal(profile['temp'][0], profile['temp'][1]))
32         humd = max(0, min(np.random.normal(profile['humd'][0], profile['humd'][1]), 100))
33         pres = max(0, np.random.normal(profile['pres'][0], profile['pres'][1]))
34
35         # create dictionary
36         msg={"time": time.time(), "profile_name": profile_name, "temperature": temp,"humidity": humd, "pressure":pres}
37
38         #randomly eliminate some measurements
39         for i in range(3):
40             if(random.randrange(0,10)<1):
41                 choice=random.randrange(0,3)
42                 if(choice==0):
43                     msg['temperature']=None
44                 elif (choice==1):
45                     msg['humidity']=None
46                 else:
47                     msg['pressure']=None
48
49         record_key=record_key+1
50         record_value=json.dumps(msg)
51         publisher.publish(topic_path, key=str(record_key),data=record_value.encode('utf-8'))
52         time.sleep(.5)
53
54         #if keyboard interrupt, break
55     except KeyboardInterrupt:
56         break
57
58     #Print produced record
59     print("Produced record to topic: {} with key {} values: {}".format(topic_path, record_key, msg))
```

Create consumer.py script to retrieve the data from a topic that has converted the measurements to PSI and Fahrenheit.

```
consumer.py > ...
1 import os
2 from google.cloud import pubsub_v1
3
4 #Set the environment variable to the credentials file
5 credentials_path = os.path.join('/Users/abdulbhutta/Desktop/Cloud Computing/Project_Milestone2/Design/credentials.json')
6 os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = credentials_path
7
8 #Create a subscriber client
9 subscriber = pubsub_v1.SubscriberClient()
10 subscription_path = 'projects/projectmilestone2/subscriptions/smartmeter_converted-sub'
11
12 #function to be called when a message is received
13 def callback(message):
14     print("Consumed record with key {} and value {}".format(message.attributes.get('key'), message.data))
15
16     #print(f'Received message with attributes: {message.attributes}')
17     message.ack() #Acknowledge the message and remove from the topic
18
19 #The subscriber listening for messages from the topic
20 streaming_pull_future = subscriber.subscribe(subscription_path, callback=callback)
21 print(f"Started listening for messages on {subscription_path}")
22
23 #Listen for messages
24 with subscriber:
25     try:
26         #Listen for messages indefinitely
27         streaming_pull_future.result()
28     #if error occurs, cancel
29     except TimeoutError:
30         streaming_pull_future.cancel()
31         streaming_pull_future.result()
32
```

Conversion class to convert temperature and pressure

```
class conversion(beam.DoFn):
    def process(self, element):
        conversionData=element
        celsiusTemp = float(conversionData['temperature'])
        pressureKPA = float(conversionData['pressure'])
        convertedTemp= celsiusTemp * 1.8 + 32
        convertedPSI = pressureKPA / 6.895
        conversionData['temperature'] = str(convertedTemp)
        conversionData['pressure'] = str(convertedPSI)
        return [conversionData]
```

Filter the data to remove missing measurements

```
filteredData = sensorData | "Filtered Data" >> beam.Filter(lambda x: x['temperature'] is not None and x['humidity'] is not None and x['pressure'] is not None)
```

Create a Dataflow pipeline

```
with beam.Pipeline(options=pipeline_options) as p:
    sensorData = (p | "Read from Pub/Sub" >> beam.io.ReadFromPubSub(topic=known_args.input)
    | "toDict" >> beam.Map(lambda x: json.loads(x)))

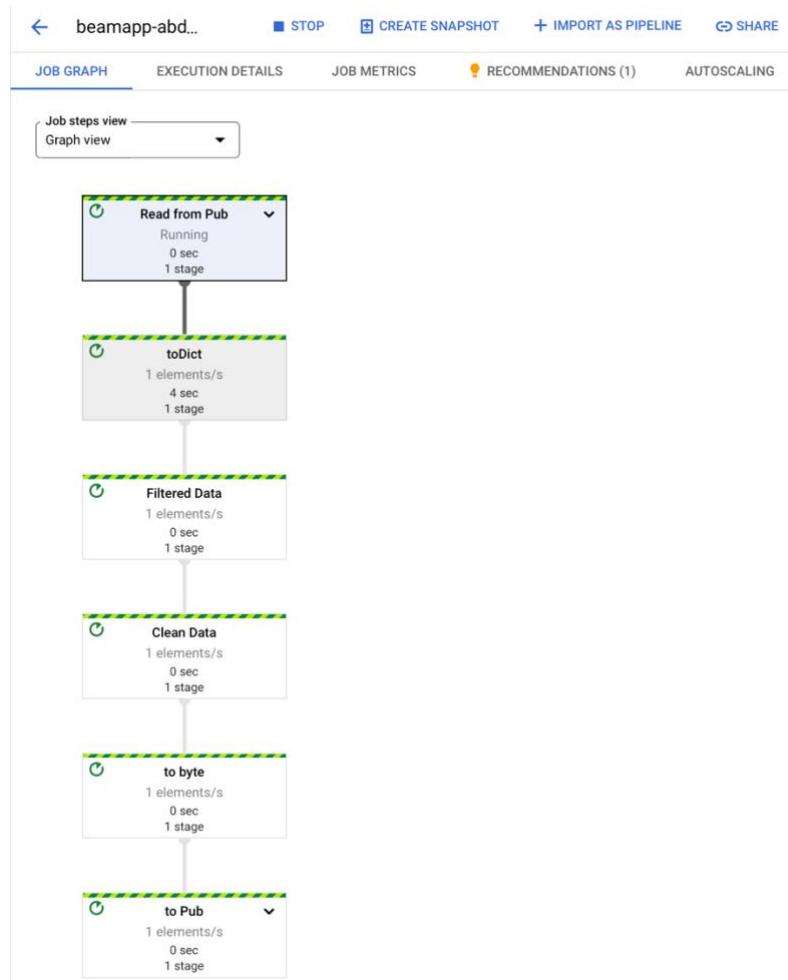
    filteredData = sensorData | "Filtered Data" >> beam.Filter(lambda x: x['temperature'] is not None and x['humidity'] is not None and x['pressure'] is not None)
    sensorCleanData = filteredData | 'Clean Data' >> beam.ParDo(conversion())

    (sensorCleanData | 'to byte' >> beam.Map(lambda x: json.dumps(x).encode('utf8'))
    | 'to Pub/Sub' >> beam.io.WriteToPubSub(topic=known_args.output));
```

Run the python script

```
abdulbhutta7@cloudshell:~/smartMeter (projectmilestone2)$ python stream_dataflow.py \
--runner DataflowRunner \
--project $PROJECT \
--staging_location $BUCKET/staging \
--temp_location $BUCKET/temp \
--model $BUCKET/model \
--input projects/$PROJECT/topics/smartmeter_readings \
--output projects/$PROJECT/topics/smartmeter_converted \
--region northamerica-northeast2 \
--experiment use_unsupported_python_version \
--streaming
2023-02-16 23:30:01.027281: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load
tory
2023-02-16 23:30:01.027322: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerr
INFO:apache_beam.runners.portability.stager:Downloading source distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download'
INFO:apache_beam.runners.portability.stager:Staging SDK sources from PyPi: dataflow_python_sdk.tar
INFO:apache_beam.runners.portability.stager:Downloading binary distribution of the SDK from PyPi
INFO:apache_beam.runners.portability.stager:Executing command: ['/usr/bin/python', '-m', 'pip', 'download'
'--implementation', 'cp', '--abi', 'cp39', '--platform', 'manylinux2014_x86_64']
INFO:apache_beam.runners.portability.stager:Staging binary distribution of the SDK from PyPi: apache_beam-
INFO:apache_beam.runners.dataflow.dataflow_runner:Pipeline has additional dependencies to be installed in
earn more on https://cloud.google.com/dataflow/docs/guides/using-custom-containers#prebuild
INFO:root:Default Python SDK image for environment is apache/beam_python3.9_sdk:2.44.0
INFO:root:Using provided Python SDK container image: gcr.io/cloud-dataflow/v1beta3/python39-fnapi:2.44.0
INFO:root:Python SDK container image set to "gcr.io/cloud-dataflow/v1beta3/python39-fnapi:2.44.0" for Dock
INFO:apache_beam.internal.gcp.auth:Setting socket default timeout to 60 seconds.
INFO:apache_beam.internal.gcp.auth:socket default timeout is 60.0 seconds.
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/
INFO:apache_beam.runners.dataflow.internal.apiclient:Starting GCS upload to gs://projectmilestone2-bucket/
```

View the DataFlow graph



Run the producer.py script that produces random readings

```
(base) Abdul-MacBook-Pro:Design abdulbhutta$ python producer.py
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 1 values: {"time": "1676589831.863809", "profile_name": "losang", "temperature": 56.760551240212235, "humidity": 45.79085678052460, "pressure": 0.1484160313152480}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 2 values: {"time": "1676589834.354535", "profile_name": "denver", "temperature": 56.6986728526756, "humidity": 41.86472505789461, "pressure": 1.178774454896213}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 3 values: {"time": "1676589834.859843", "profile_name": "boston", "temperature": 26.51468184752412, "humidity": 31.3168618365607, "pressure": 1.19994178020813}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 4 values: {"time": "1676589835.369163", "profile_name": "boston", "temperature": 29.6579801449986, "humidity": 70.7067213174264, "pressure": 1.89164718825191}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 5 values: {"time": "1676589836.370495", "profile_name": "boston", "temperature": 29.6579801449986, "humidity": 70.7067213174264, "pressure": 1.89164718825191}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 6 values: {"time": "1676589836.379888", "profile_name": "denver", "temperature": 60.8433591448227, "humidity": 15.845418049488087, "pressure": 1.10182152131308}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 7 values: {"time": "1676589836.881872", "profile_name": "boston", "temperature": 62.09846970368807, "humidity": 22.96280815912135, "pressure": 1.3482368719451685}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 8 values: {"time": "1676589837.082253", "profile_name": "boston", "temperature": 53.78477175542414, "humidity": 48.8634731173973, "pressure": 1.22210217798931}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 9 values: {"time": "1676589837.887948", "profile_name": "losang", "temperature": 53.78477175542414, "humidity": 48.8634731173973, "pressure": 1.22210217798931}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 10 values: {"time": "1676589838.091355", "profile_name": "boston", "temperature": 59.33625790395465, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 11 values: {"time": "1676589838.295211", "profile_name": "boston", "temperature": 59.33625790395465, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 12 values: {"time": "1676589839.497639", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.4118730577935, "pressure": 1.2031926910404012}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 13 values: {"time": "1676589839.518418", "profile_name": "denver", "temperature": 65.19646067269913, "humidity": 100, "pressure": 1.081363764514078}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 14 values: {"time": "1676589840.023887", "profile_name": "boston", "temperature": 66.202570368807, "humidity": 30.2002036723008, "pressure": 0.9499377862634467}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 15 values: {"time": "1676589840.215796", "profile_name": "losang", "temperature": 66.4647224545789, "humidity": None, "pressure": 0.85963778626517}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 16 values: {"time": "1676589841.422405", "profile_name": "losang", "temperature": 45.29748314657933, "humidity": 61.30404216211327, "pressure": None}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 17 values: {"time": "1676589842.034503", "profile_name": "boston", "temperature": 60.83972675058033, "humidity": 38.480023659398135, "pressure": 1.6938404656536737}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 18 values: {"time": "1676589842.534503", "profile_name": "losang", "temperature": 58.4543951577389, "humidity": 39.641650923564282, "pressure": 1.02621225493756}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 19 values: {"time": "1676589842.541881", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 20 values: {"time": "1676589843.053483", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 21 values: {"time": "1676589843.553483", "profile_name": "denver", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 22 values: {"time": "1676589844.446881", "profile_name": "denver", "temperature": None, "humidity": 18.77721665118891, "pressure": 1.0718813591817885}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 23 values: {"time": "1676589844.546881", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 24 values: {"time": "1676589844.556289", "profile_name": "denver", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 25 values: {"time": "1676589844.556321", "profile_name": "losang", "temperature": 64.37728971768014, "humidity": 7.552273716456764, "pressure": 0.2.38445340562406, "pressure": 1.6514865515476611}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 26 values: {"time": "1676589844.556321", "profile_name": "boston", "temperature": 64.37728971768014, "humidity": 85.298751349980014, "pressure": 0.2.38445340562406, "pressure": 1.6514865515476611}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 27 values: {"time": "1676589844.556321", "profile_name": "denver", "temperature": 64.37728971768014, "humidity": 66.37728971768014, "pressure": 0.2.38445340562406, "pressure": 1.6514865515476611}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 28 values: {"time": "1676589844.890228", "profile_name": "boston", "temperature": 41.0587060788771, "humidity": None, "pressure": 0.888107624972226}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 29 values: {"time": "1676589844.890228", "profile_name": "losang", "temperature": 48.7495436055986, "humidity": 58.314544774109385, "pressure": 0.36745242462155}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 30 values: {"time": "1676589845.016028", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 31 values: {"time": "1676589845.016028", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 32 values: {"time": "1676589845.324080", "profile_name": "losang", "temperature": 58.42269959707771, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 33 values: {"time": "1676589845.324080", "profile_name": "boston", "temperature": 65.19646067269913, "humidity": 30.8266950168112, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 34 values: {"time": "1676589845.336209", "profile_name": "boston", "temperature": 78.422026930864, "humidity": 93.84686511811948, "pressure": 0.9065795428622883}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 35 values: {"time": "1676589845.442355", "profile_name": "boston", "temperature": 79.1865795428622883, "humidity": 88.30638231646449, "pressure": 0.81889767460616392}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 36 values: {"time": "1676589845.442355", "profile_name": "denver", "temperature": 79.1865795428622883, "humidity": 100, "pressure": 0.9286784234968986}
Produced record to topic: projects/projectmilestone2/topics/smarmeter.readings with key 37 values: {"time": "1676589852.056108", "profile_name": "denver", "temperature": 62.11758693137894, "humidity": None, "pressure": 1.459768725995051}
```

Run the consumer.py script and view the converted readings

```
(base) Abdul-MacBook-Pro:Design abdulbhutta$ python consumer.py
Started listening for messages on projects/projectmilestone2/subscriptions/smarmeter_converted-sub
Consumed record with key None and value b'{"time": "1676589834.359226", "profile_name": "losang", "temperature": "158.97751189267297", "humidity": "61.08472250578801", "pressure": "0.199068012081102"}'
Consumed record with key None and value b'{"time": "1676589834.859843", "profile_name": "denver", "temperature": "147.27740311930806", "humidity": "29.88320570800014", "pressure": "0.20725879135938837"}'
Consumed record with key None and value b'{"time": "1676589835.369163", "profile_name": "boston", "temperature": "130.98011405414195", "humidity": "91.36768247564575", "pressure": "0.14695044851703855"}'
Consumed record with key None and value b'{"time": "1676589843.556289", "profile_name": "boston", "temperature": "156.36204326965584", "humidity": "45.4543951577389", "pressure": "0.16839722173707547"}'
Consumed record with key None and value b'{"time": "1676589844.446881", "profile_name": "boston", "temperature": "158.649226398907", "humidity": "46.487185994897494", "pressure": "0.26565732270433307"}'
Consumed record with key None and value b'{"time": "1676589844.546881", "profile_name": "denver", "temperature": "74.49645211108361", "humidity": "100", "pressure": "0.24189235690864558"}'
Consumed record with key None and value b'{"time": "1676589844.556289", "profile_name": "denver", "temperature": "74.49645211108361", "humidity": "100", "pressure": "0.24189235690864558"}'
Consumed record with key None and value b'{"time": "1676589844.556321", "profile_name": "losang", "temperature": "137.27918485793", "humidity": "30.641650852564202", "pressure": "0.148831396869793"}'
Consumed record with key None and value b'{"time": "1676589844.556321", "profile_name": "boston", "temperature": "137.27918485793", "humidity": "30.641650852564202", "pressure": "0.148831396869793"}'
Consumed record with key None and value b'{"time": "1676589844.880228", "profile_name": "boston", "temperature": "101.3597848100631", "humidity": "56.314544774109385", "pressure": "0.1463401340458615"}'
Consumed record with key None and value b'{"time": "1676589844.880228", "profile_name": "boston", "temperature": "173.141965260863533", "humidity": "93.84686511811948", "pressure": "0.11514932543325134"}'
Consumed record with key None and value b'{"time": "1676589845.016028", "profile_name": "boston", "temperature": "139.50016561211737", "humidity": "53.5714373612282", "pressure": "0.162363554537787'}'
Consumed record with key None and value b'{"time": "1676589845.324080", "profile_name": "losang", "temperature": "134.50016561211737", "humidity": "53.5714373612282", "pressure": "0.162363554537787"}'
Consumed record with key None and value b'{"time": "1676589845.324080", "profile_name": "boston", "temperature": "185.6785213799541", "humidity": "93.32322292333015", "pressure": "0.15172948780816817"}'
Consumed record with key None and value b'{"time": "1676589845.336209", "profile_name": "boston", "temperature": "134.76589867.7658921", "humidity": "56.20148954949467", "pressure": "0.17937692914943127"}'
Consumed record with key None and value b'{"time": "1676589845.442355", "profile_name": "boston", "temperature": "172.8836484198844", "humidity": "96.28545182729883", "pressure": "0.128023829338867"}'
Consumed record with key None and value b'{"time": "1676589845.442355", "profile_name": "denver", "temperature": "33.31160618163507", "humidity": "44.514910375467", "pressure": "0.131616429694606464"}'
Consumed record with key None and value b'{"time": "1676589845.442355", "profile_name": "denver", "temperature": "62.3108923748137", "humidity": "44.514910375467", "pressure": "0.131616429694606464"}'
Consumed record with key None and value b'{"time": "1676589845.556289", "profile_name": "losang", "temperature": "181.40532422621894", "humidity": "71.6927258458891", "pressure": "0.191160599300686"}'
Consumed record with key None and value b'{"time": "1676589845.88701", "profile_name": "denver", "temperature": "100.91975868311959", "humidity": "62.13880434869554", "pressure": "0.14171999296321587"}'
Consumed record with key None and value b'{"time": "1676589868.71179", "profile_name": "losang", "temperature": "151.0041743639313", "humidity": "90.5934628960034", "pressure": "0.1679528729982810"}'
Consumed record with key None and value b'{"time": "1676589868.859843", "profile_name": "boston", "temperature": "134.18339274329806", "humidity": "45.7985547507569", "pressure": "0.1326202543169321"}'
Consumed record with key None and value b'{"time": "1676589868.916028", "profile_name": "boston", "temperature": "134.18339274329806", "humidity": "45.7985547507569", "pressure": "0.1326202543169321"}'
Consumed record with key None and value b'{"time": "1676589872.337023", "profile_name": "boston", "temperature": "172.8836484198844", "humidity": "96.28545182729883", "pressure": "0.128023829338867"}'
Consumed record with key None and value b'{"time": "1676589872.337023", "profile_name": "denver", "temperature": "33.31160618163507", "humidity": "44.514910375467", "pressure": "0.131616429694606464"}'
Consumed record with key None and value b'{"time": "1676589872.373023", "profile_name": "denver", "temperature": "62.3108923748137", "humidity": "44.514910375467", "pressure": "0.131616429694606464"}'
Consumed record with key None and value b'{"time": "1676589872.373023", "profile_name": "losang", "temperature": "164.55420938936125", "humidity": "100", "pressure": "0.16481710501081573"}'
Consumed record with key None and value b'{"time": "1676589872.776696", "profile_name": "losang", "temperature": "91.48497780780163", "humidity": "98.488163767583", "pressure": "0.19175338677642855"}'
Consumed record with key None and value b'{"time": "1676589872.776696", "profile_name": "denver", "temperature": "104.5701145335909", "humidity": "36.747936854353", "pressure": "0.2436479844284894"}'
Consumed record with key None and value b'{"time": "1676589876.889198", "profile_name": "losang", "temperature": "126.82696630880013", "humidity": "72.0943664954528", "pressure": "0.1760743677845766"}'
Consumed record with key None and value b'{"time": "1676589876.91304819", "profile_name": "boston", "temperature": "174.758249518978", "humidity": "100", "pressure": "0.1588338313290833"}'
```

Deliverables

GitHub Repo

https://github.com/abdulbhutta/Cloud-Computing/tree/main/Final%20Project/Milestone2_DataFlow

D2: Four Examples Execution Videos

https://drive.google.com/file/d/1K8h8UHjECHIF5CK5XW06O_PbC-YXv4pw/view

D3: Design Part Execution Videos

<https://drive.google.com/file/d/1lCVx0ma0UlkvpWs9-f6uuuh-f-QkUEGmG/view>