



**Cloud Computing
(SOFE 4630U)**

**Project Milestone 3: Data Processing: Dataflow/Apache
Beam**

Name	Student ID
Rohan Radadiya	100704614

Due Date: February 24, 2025

GitHub Link:

<https://github.com/rohanradadiya/Cloud-Computing-Project-MS-3>

Google Drive for Videos:

https://drive.google.com/drive/folders/1HcPcpJlkzAwhymcY-gjWLjAgn-H_Qgwy?usp=sharing

Screenshots:

The screenshot displays the Google Cloud Platform (GCP) console interface for a project named "My First Project". The "Service accounts" page is active, showing a list of service accounts. Below the list, a terminal window shows the execution of a Python script named `wordcount.py` using the `fine-cycling-451417-q7` service account.

Service accounts for project "My First Project"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, App Engine apps, or systems running outside Google. [Learn more about service accounts.](#)

Organization policies can be used to secure service accounts and block risky service account features, such as automatic IAM Grants, key creation/upload, or the creation of service accounts entirely. [Learn more about service account organization policies.](#)

Filter	Email	Status	Name	Description	Key ID	Key creation date	OAuth 2 Client ID	Actions
	554797386760-compute@developer.gserviceaccount.com	Enabled	Compute Engine default service account		No keys		102237916812002112889	
	fine-cycling-451417-q7-dfsa@fine-cycling-451417-q7.iam.gserviceaccount.com	Enabled	fine-cycling-451417-q7-DFSAs		No keys		117513887737621949532	

```

WARNING: The script gcp_client is installed in '/home/rohan_radadiya/work/10041/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
...
Successfully installed apache-beam-2.63.0 cloudpickle-2.2.1 dill-0.3.1.1 docopt-0.6.2 fastavro-1.10.0 fasteners-0.19 google-apitools-0.5.31 google-cloud-bigquery-storage-2.28.0 google-cloud-bigtable-2.28.1 google-cloud-dlp-3.27.0 google-clou
d-pubsub-2.28.0 google-cloud-pubsublite-1.11.1 google-cloud-recommendations-ai-0.10.16 grpcio-1.65.5 hdfs-2.7.3 jsonpickle-3.4.2 objsize-0.7.1 opentelemetry-api-1.30.0 opentelemetry-sdk-1.30.0 opentelemetry-semantic-conventions-0.51b0 orjson
-3.10.15 overrides-7.7.0 pyarrow-16.1.0 pyarrow-hdfs-0.6 pydot-1.4.2 pytz-2025.1 redis-5.2.1 regex-2024.11.6 sortedcontainers-2.4.0 standard-0.23.0
rohan_radadiya@cloudshell:~ (fine-cycling-451417-q7) $
rohan_radadiya@cloudshell:~ (fine-cycling-451417-q7) $ python wordcount.py --output outputs
INFO:root:Missing pipeline option (runner). Executing pipeline using the default runner: DirectRunner.
INFO:apache_beam.runners.worker.statecache:Creating state cache with size 104857600
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.
rohan_radadiya@cloudshell:~ (fine-cycling-451417-q7) $ ls outputs*
outputs-00000-of-00001
rohan_radadiya@cloudshell:~ (fine-cycling-451417-q7) $ 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: 498
Gloucester: 26
EDMUND: 99
bastard: 7
CURAN: 6
a: 366
courtier: 1
Old: 13
Man: 11
tenant: 3
--More--

```

← Bucket details ↗

fine-cycling-451417-q7-bucket

Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Not public	Soft Delete

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser 🔗

Buckets > fine-cycling-451417-q7-bucket 📁

CREATE FOLDER UPLOAD ▾ TRANSFER DATA ▾ OTHER SERVICES ▾

Filter by name prefix only ▾ 🔍 Filter Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created ?	Storage class	Last modified	Public access ?	Version history ?	Encryption ?	0
No rows to display										

Bucket created:

```
rohan_radadiyawork@cloudshell:~ (fine-cycling-451417-q7) $ BUCKET=gs://$PROJECT-bucket
echo $BUCKET
gs://fine-cycling-451417-q7-bucket
rohan_radadiyawork@cloudshell:~ (fine-cycling-451417-q7) $
```

Environment variable for bucket created:

```
n) *Write/Write/WriteImpl/WriteBundles*Write/Write/WriteImpl/Pair*Write/Write/WriteImpl/GroupByKey/Write
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:48.030Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/Close
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:48.063Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/Close
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:48.089Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/Read*Write/Write/WriteImpl/Extract
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.181Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/GroupByKey/Read*Write/Write/WriteImpl/Extract
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.254Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/GroupByKey/Read*View-python_side_input1
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.285Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/PreFinalize/View-python_side_input1
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.309Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/PreFinalize/View-python_side_input1
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.347Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/PreFinalize/View-python_side_input1
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:50.408Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:53.406Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/PreFinalize
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:53.462Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:53.512Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:53.570Z: JOB MESSAGE BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:55.535Z: JOB MESSAGE BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:25:55.910Z: JOB MESSAGE BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:26:41.632Z: JOB MESSAGE BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 18 22 19-5872698157292134026 is in state JOB_STATE_DONE
rohan_radadiyawork@cloudshell:~ (fine-cycling-451417-q7) $
```

Dataflow job details

beamapp-roh... ■ STOP + IMPORT AS PIPELINE ↗ SHARE 📁 ARCHIVE + TRIGGER ▾ SEND FEEDBACK

beamapp-rohanradadiyawork-0221022217-855671-sm3l8etp JOB METRICS COST RECOMMENDATIONS

Job steps view Graph view

CLEAR SELECTION

```

graph TD
    Read[Read Succeeded  
1 sec  
2 of 2 stages succeeded] --> Split[Split Succeeded  
1 of 1 stage succeeded]
    Split --> PairWithOne[PairWithOne Succeeded  
0 sec  
1 of 1 stage succeeded]
    PairWithOne --> GroupAndSum[GroupAndSum Succeeded  
0 sec  
2 of 2 stages succeeded]
  
```

Logs SHOW

Job info

Job name beamapp-rohanradadiyawork-0221022217-855671-sm3l8etp

Job ID 2025-02-20-18_22_19-5872698157292134026

Job type Batch

Job status ✓ Succeeded

SDK version Apache Beam Python 3.12 SDK 2.63.0

Job region ? northamerica-northeast2

Current workers ? 0

Latest worker status Worker pool stopped.

Start time February 20, 2025 at 9:22:20 PM GMT-5

Elapsed time 4 min 29 sec

Encryption type Google-managed

Dataflow Prime ? Disabled

Runner v2 ? Enabled

Dataflow Shuffle ? Enabled

Resource metrics

Current vCPUs ? 1

Total vCPU time ? 0.039 vCPU hr

Current memory ? 3.75 GB

Total memory time ? 0.149 GB hr

Current HDD PD ? 25 GB

Total HDD PD time ? 0.99 GB hr

Current SSD PD ? 0 B

Total SSD PD time ? 0 B hr

Open Editor

dataflow job details

beamapp-roh... STOP + IMPORT AS PIPELINE SHARE ARCHIVE + TRIGGER SEND FEEDBACK

beamapp-rohanradadiywork-0221022217-855671-sm3l8etp JOB METRICS COST RECOMMENDATIONS

Job steps view Graph view CLEAR SELECTION

Logs HIDE MAX TIME

JOB LOGS **WORKER LOGS** **DIAGNOSTICS** **DATA SAMPLING**

Severity Filter Search all fields and values

SEVERITY INFO WARNING ERROR FATAL DEBUG TRACE SUMMARY

No older entries found matching current filter.

2025-02-20 21:22:23.932 EST Worker configuration: nl-standard-1 in northamerica-northeast-2-a.

2025-02-20 21:22:25.452 EST Executing operation Read/Read/Impulse+Read/Read/EmitsSource+Read/Read/SDFBoundedSourceReader/ParDo(SDFBoundedSourceDoFn)/Pa...

2025-02-20 21:22:25.462 EST Executing operation Write/Write/WriteImpl/DoOnce/Impulse+Write/Write/WriteImpl/DoOnce/FlatMap(lambda at core.py:397b)+Wri...

2025-02-20 21:22:25.486 EST Starting 1 workers in northamerica-northeast-2...

2025-02-20 21:25:42.918 EST All workers have finished the startup processes and began to receive work requests.

2025-02-20 21:25:44.812 EST Finished operation Read/Read/Impulse+Read/Read/EmitsSource+Read/Read/SDFBoundedSourceReader/ParDo(SDFBoundedSourceDoFn)/Pa...

2025-02-20 21:25:44.876 EST Executing operation GroupAndSum/GroupByKey/Create

2025-02-20 21:25:45.249 EST Finished operation GroupAndSum/GroupByKey/Create

2025-02-20 21:25:45.381 EST Executing operation Read/Read/SDFBoundedSourceReader/ParDo(SDFBoundedSourceDoFn)/ProcessElementAndRestrictionWithSizing+Spl...

Job info

Job name beamapp-rohanradadiywork-0221022217-855671-sm3l8etp

Job ID 2025-02-20_18_22_19-5872698157292134026

Job type Batch

Job status Succeeded

SDK version Apache Beam Python 3.12 SDK 2.63.0

Job region northamerica-northeast2

Current workers 0

Latest worker status Worker pool stopped.

Start time February 20, 2025 at 9:22:20 PM GMT-5

Elapsed time 4 min 29 sec

Encryption type Google-managed

Dataflow Prime Disabled

Runner v2 Enabled

Dataflow Shuffle Enabled

Resource metrics

Current vCPUs 1

Total vCPU time 0.039 vCPU hr

[Open Editor](#)

2. Running the wordcount2 Example

```
type: typeValueValuesFrom(JOB_TYPE_BATCH, 1))
INFO:apache_beam.runners.dataflow.internal.apiclient:Created job with id: [2025-02-20 18 39 14-11053338657560234488]
INFO:apache_beam.runners.dataflow.internal.apiclient:Submitted job: 2025-02-20 18 39 14-11053338657560234488
INFO:apache_beam.runners.dataflow.internal.apiclient:To access the Dataflow monitoring console, please navigate to https://console.cloud.google.com/dataflow/jobs/northamerica-northeast2/2025-02-20_18_39_14-11053338657560234488?project=fine-c
ycling-451417-q7
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 18 39 14-11053338657560234488 is in state JOB_STATE_PENDING
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:13:16.163Z: JOB_MESSAGE_BASIC: Worker configuration: nl-standard-1 in northamerica-northeast-2-b.
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 18 39 14-11053338657560234488 is in state JOB_STATE_RUNNING
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:13:19.172Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/DoOnce/Impulse+Write/Write/WriteImpl/DoOnce/FlatMap(lambda at core.py:397b)+Write/Write/WriteImpl/D
oOnce/Map(decode)+Write/Write/WriteImpl/InitialiseWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:13:19.893Z: JOB_MESSAGE_BASIC: Executing operation Read/Read/Impulse+Read/Read/EmitsSource+Read/Read/SDFBoundedSourceReader/ParDo(SDFBoundedSourceDoFn)/SplitWithSizing
eAd/SDFBoundedSourceReader/ParDo(SDFBoundedSourceDoFn)/SplitWithSizing
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:13:19.904Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/DoOnce/Impulse+Write/Write/WriteImpl/DoOnce/FlatMap(lambda at core.py:397b)+Write/Write/WriteImpl/DoO
n/Map(decode)+Write/Write/WriteImpl/InitialiseWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:13:19.908Z: JOB_MESSAGE_BASIC: Starting 1 workers in northamerica-northeast-2...
```

```
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:16.642Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:16.697Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:16.742Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:17.782Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/PreFinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:17.827Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:17.874Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite/View-python_side_input2
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:17.921Z: JOB_MESSAGE_BASIC: Executing operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:18.489Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:19.777Z: JOB_MESSAGE_BASIC: Finished operation Write/Write/WriteImpl/FinalizeWrite
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:42:20.136Z: JOB_MESSAGE_BASIC: Stopping worker pool...
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T02:43:04.717Z: JOB_MESSAGE_BASIC: Worker pool stopped.
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 18 39 14-11053338657560234488 is in state JOB_STATE_DONE
rohan_radadiywork@cloudshell:~/SOPF4630U-HS3/wordcount (fine-cycling-451417-q7)$
```

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser IK

Buckets > fine-cycling-451417-q7-bucket > result IK

CREATE FOLDER UPLOAD TRANSFER DATA OTHER SERVICES

Filter by name prefix only Filter Filter objects and folders Show Live objects only IK

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	<input type="checkbox"/>
<input type="checkbox"/>	outputs-0000-of-00001	13.9 KB	text/plain	Feb 20, 2025, 9:42:19 PM	Standard	Feb 20, 2025, 9:42:19 PM	Not public	—	<input type="checkbox"/>
<input type="checkbox"/>	outputs2-00000-of-00001	189 B	text/plain	Feb 20, 2025, 9:42:18 PM	Standard	Feb 20, 2025, 9:42:18 PM	Not public	—	<input type="checkbox"/>

3. Batch Processing of the MNIST Dataset

The screenshot shows the Google Cloud Storage interface for a bucket named 'fine-cycling-451417-q7-bucket'. The bucket is located in 'us (multiple regions in United States)', has a 'Standard' storage class, and is 'Not public'. The 'Folder browser' tab is active, showing a folder structure with 'model/' and 'result/' folders. The 'Filter by name prefix only' is set to 'model/'.

Model folder uploaded:

The screenshot shows a Cloud Shell terminal window with a terminal session for the 'fine-cycling-451417-q7' job. The terminal output displays various log messages from the Dataflow job, including 'Executing operation WriteToBQ/BigQueryBatchFileLoads/Flatmap', 'Finished operation WriteToBQ/BigQueryBatchFileLoads/GroupByKey/Close', and 'Finished operation WriteToBQ/BigQueryBatchFileLoads/Flatmap'. The job appears to be running successfully.

Running Dataflow after uploading the /mnist/model folder:

The screenshot shows the Google Cloud Dataflow console with a query result displayed. The query is 'SELECT * FROM 'fine-cycling-451417-q7.MNIST.Predict' LIMIT 1000'. The results are shown in a table with columns for 'JOB INFORMATION', 'RESULTS', 'CHART', 'JSON', 'EXECUTION DETAILS', and 'EXECUTION GRAPH'. The table contains 10 rows of data, including job IDs, timestamps, and various metrics.

4. Stream Processing the MNIST Dataset

LIST						Select a topic	
Filter						PERMISSIONS	
Topic ID ↑	Encryption key	Topic name	Retention	Ingestion source			
<input type="checkbox"/> mnist_image	Google-managed	projects/fine-cycling-451417-q7/topics/mnist_image	—	—			
<input type="checkbox"/> mnist_predict	Google-managed	projects/fine-cycling-451417-q7/topics/mnist_predict	—	—			

“mnist_image” and “mnist_predict” topics created:

```
location: 'northamerica-northeast2'
name: 'beampp-rohanreddiyawork-0221032142-299956-in7lakj0'
projectId: 'fine-cycling-451417-q7'
stageState: {}
startTime: '2025-02-21T03:21:44.100748Z'
steps: []
tempFiles: []
type: TypeValueValuesEnum(JOB_TYPE_STREAMING, 2)>
INFO:apache_beam.runners.dataflow.internal.apiclient:Created job with id: (2025-02-20 19 21 43-13316342389863091699)
INFO:apache_beam.runners.dataflow.internal.apiclient:Submitted job: 2025-02-20 19 21 43-13316342389863091699
INFO:apache_beam.runners.dataflow.internal.apiclient:To access the Dataflow monitoring console, please navigate to https://console.cloud.google.com/dataflow/jobs/northamerica-northeast2/2025-02-20_19_21_43-13316342389863091699?project=fine-cycling-451417-q7
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 19 21 43-13316342389863091699 is in state JOB_STATE_PENDING
WARNING:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T03:21:45.1805Z: JOB MESSAGE WARNING: Autoscaling is enabled for Dataflow Streaming Engine. Workers will scale between 1 and 100 unless maxNumWorkers is specified.
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T03:21:49.413Z: JOB MESSAGE BASIC: Worker configuration: nl-standard-2 in northamerica-northeast2-b.
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T03:21:51.036Z: JOB MESSAGE BASIC: Running job using Streaming Engine
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T03:21:53.016Z: JOB MESSAGE BASIC: Starting 1 workers in northamerica-northeast2...
INFO:apache_beam.runners.dataflow.dataflow_runner:Job 2025-02-20 19 21 43-13316342389863091699 is in state JOB_STATE_RUNNING
INFO:apache_beam.runners.dataflow.dataflow_runner:2025-02-21T03:25:27.092Z: JOB MESSAGE BASIC: All workers have finished the startup processes and began to receive work requests.
[]
```

Running the Data flow job that reads JSON objects from the mnist_image topic, applying it to the ML model, and sending the prediction results via the mnist_predict topic:

```
25
26 df=pd.read_csv('mnist.csv')
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
[notice] To update, run: python.exe -m pip install --upgrade pip
PS C:\Users\rohan\Downloads\mnist\data> python producerMnistPubSup.py
o Image with key 0 is sent
Image with key 1 is sent
Image with key 2 is sent
Image with key 3 is sent
Image with key 4 is sent
Image with key 5 is sent
Image with key 6 is sent
Image with key 7 is sent
Image with key 8 is sent
Image with key 9 is sent
Image with key 10 is sent
Image with key 11 is sent
Image with key 12 is sent
Image with key 13 is sent
Image with key 14 is sent
Image with key 15 is sent
Image with key 16 is sent
Image with key 17 is sent
Image with key 18 is sent
Image with key 19 is sent
Image with key 20 is sent
Image with key 21 is sent
Image with key 22 is sent
Image with key 23 is sent
Image with key 24 is sent
Image with key 25 is sent
Image with key 26 is sent
Image with key 27 is sent
Image with key 28 is sent
Image with key 29 is sent
```

After updating producerMnistPubSup.py script and running it:

```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

.186789212639434e-14, 'P7': 4.818560143277009e-06, 'P8': 4.4119585148813806e-07, 'P9': 0.999948395355225},
Received ('ID': 377, 'P0': 1.0162728444618099e-08, 'P1': 0.9999817609782897, 'P2': 5.886346911130644e-07, 'P3': 2.3853276626998086e-08, 'P4': 7.858274329919368e-06, 'P5': 5.380304646251943e-09, 'P6': 4.076
175397216275e-09, 'P7': 9.47867474044161e-06, 'P8': 2.391214763974858e-07, 'P9': 5.2222283386527124e-08},
Received ('ID': 217, 'P0': 5.86363739787851562e-07, 'P1': 1.1061812923984782e-11, 'P2': 2.572931641793057e-09, 'P3': 6.928018714269513e-10, 'P4': 2.571886315028138e-10, 'P5': 0.036524783819913864, 'P6': 0.96
33223414421882, 'P7': 2.3631399308562037e-14, 'P8': 0.00015224385424517895, 'P9': 2.1833851393537888e-10},
Received ('ID': 389, 'P0': 4.9180559798245316e-12, 'P1': 1.8488867319408273e-08, 'P2': 3.7519681805875393e-10, 'P3': 4.593324547386146e-07, 'P4': 0.000663252035143602, 'P5': 6.769045057808398e-08, 'P6': 2.
2605325187538333e-15, 'P7': 0.0014610294019803405, 'P8': 2.5184385776810814e-06, 'P9': 0.9978277102279663},
Received ('ID': 176, 'P0': 1.9463459466351196e-06, 'P1': 0.9996246099472045, 'P2': 4.131226887693629e-06, 'P3': 2.8298742904553364e-07, 'P4': 0.0003343856606491804, 'P5': 1.5387116718557081e-06, 'P6': 5.8
69821052328916e-06, 'P7': 1.4656588064099190e-05, 'P8': 1.2138538109950807e-05, 'P9': 4.4258973730393336e-07},
Received ('ID': 398, 'P0': 5.130879228432139e-12, 'P1': 4.0085704200443217e-10, 'P2': 8.158788089573932e-10, 'P3': 1.6775551217249052e-13, 'P4': 0.9999986886978149, 'P5': 5.680823589315764e-10, 'P6': 4.045
8764810624714e-11, 'P7': 4.371257773527759e-07, 'P8': 3.279616823199594e-10, 'P9': 7.907381132699454e-07},
Received ('ID': 358, 'P0': 5.152361610277767e-12, 'P1': 0.80375492397463e-06, 'P2': 3.039711372276077e-05, 'P3': 1.9293131572339917e-06, 'P4': 0.001221937476657331, 'P5': 2.5581887719283714e-09, 'P6': 1.73
04772998122213e-14, 'P7': 0.9987226128578186, 'P8': 7.226081066046746e-08, 'P9': 4.1476963815512136e-05},
Received ('ID': 177, 'P0': 0.000415507838582471, 'P1': 2.6670802562453595e-07, 'P2': 6.279485387494788e-05, 'P3': 3.641077910288004e-06, 'P4': 1.3551773918152321e-05, 'P5': 5.3969324653735384e-06, 'P6': 0.
00024069570645224303, 'P7': 7.003588393672544e-07, 'P8': 0.9992209672927856, 'P9': 3.629064667620696e-05},
Received ('ID': 365, 'P0': 0.9999676942825317, 'P1': 1.962642721953234e-11, 'P2': 3.551254792454383e-07, 'P3': 3.71943448262968e-12, 'P4': 5.898443625890204e-11, 'P5': 8.393375772186573e-09, 'P6': 3.15103
8981741294e-05, 'P7': 8.812532925617234e-10, 'P8': 4.5681321125812246e-07, 'P9': 7.479902031093388e-09},
Received ('ID': 205, 'P0': 8.517662486787203e-13, 'P1': 3.774443535498051e-10, 'P2': 2.15292583760629e-08, 'P3': 0.999995231628418, 'P4': 4.333894354502377e-11, 'P5': 3.669954651286389e-07, 'P6': 1.461302
731381106e-13, 'P7': 3.573343221319192e-08, 'P8': 7.276377722575944e-08, 'P9': 6.69073891122537e-08},
Received ('ID': 200, 'P0': 3.461779751745553e-09, 'P1': 2.9239441090567198e-08, 'P2': 1.442951997887576e-06, 'P3': 0.9999432563781738, 'P4': 1.8105529228651385e-11, 'P5': 5.217174020799575e-06, 'P6': 6.917
972444897558e-11, 'P7': 3.0479255230808633e-10, 'P8': 5.008789194093534e-05, 'P9': 2.647787044454194e-08},
Received ('ID': 148, 'P0': 0.999998310699463, 'P1': 3.137540146008627e-14, 'P2': 1.5069611112027902e-08, 'P3': 3.068473751997869e-14, 'P4': 6.967312875727147e-14, 'P5': 6.894917969901826e-11, 'P6': 1.6021
904230001383e-06, 'P7': 4.662669898091226e-13, 'P8': 7.830424086918697e-10, 'P9': 5.4798453968049654e-11},
Received ('ID': 155, 'P0': 8.55915993530303e-16, 'P1': 1.9878319630215353e-13, 'P2': 5.819129900164199e-13, 'P3': 6.112525596790874e-08, 'P4': 2.039736601844311e-09, 'P5': 0.999994039535522, 'P6': 7.65188
2599768456e-11, 'P7': 4.252979533218154e-15, 'P8': 1.0787026383241027e-08, 'P9': 5.0794687922246189e-07},

```

After updating consumerMnistPubSup.py script and running it:

The screenshot displays the Google Cloud Dataflow console interface. The main panel shows the job graph for 'beampapp-roh...' with two steps: 'Read from Pub Running' and 'toDict Running'. The job status is 'Cancelled'. The job info panel on the right provides details such as Job ID, Job type (Streaming), and Service zones. The terminal at the bottom shows logs indicating the job was cancelled and the console URL.

Dataflow job is stopped manually:

Design:

```
Google Cloud SDK Shell
---
C:\Users\rohan\AppData\Local\Google\Cloud SDK>gsutil version
gsutil version: 5.33

C:\Users\rohan\AppData\Local\Google\Cloud SDK>cd C:\Users\rohan\Documents\pedestrian_detection_pipeline

C:\Users\rohan\Documents\pedestrian_detection_pipeline>gsutil -m cp -r . gs://fine-cycling-451417-q7-bucket/
Copying file:///.\models\depth_model.pth [Content-Type=application/octet-stream]...
Copying file:///.\models\yolov8n.pt [Content-Type=application/octet-stream]...
Copying file:///.\src\object_detection.py [Content-Type=text/x-python]...
Copying file:///.\src\pipeline.py [Content-Type=text/x-python]...
Copying file:///.\Dockerfile [Content-Type=application/octet-stream]...
Copying file:///.\pubsub_publisher.py [Content-Type=text/x-python]...
Copying file:///.\data\Labels.csv [Content-Type=application/vnd.ms-excel]...
Copying file:///.\pedestrian_detection_pipeline_requirements.txt [Content-Type=text/plain]...
Copying file:///.\src\depth_estimation.py [Content-Type=text/x-python]...
Copying file:///.\data\Dataset_Occluded_Pedestrian\A_085.png [Content-Type=image/png]...
Copying file:///.\data\Dataset_Occluded_Pedestrian\A_020.png [Content-Type=image/png]...
Copying file:///.\data\images\depth.jpg [Content-Type=image/jpeg]...
Copying file:///.\pubsub_subscriber.py [Content-Type=text/x-python]...
Copying file:///.\data\Dataset_Occluded_Pedestrian\C_020.png [Content-Type=image/png]...
Copying file:///.\data\images\problem.jpg [Content-Type=image/jpeg]...
Copying file:///.\data\images\ped_depth.jpg [Content-Type=image/jpeg]...
Copying file:///.\data\images\yolo.jpg [Content-Type=image/jpeg]...
/ [17/17 files][ 3.2 MiB/ 3.2 MiB] 100% Done
Operation completed over 17 objects/3.2 MiB.

C:\Users\rohan\Documents\pedestrian_detection_pipeline>
C:\Users\rohan\Documents\pedestrian_detection_pipeline>
```

```
[notice] A new release of pip is available: 25.0 -> 25.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip

C:\Users\rohan\Documents\pedestrian_detection_pipeline>
C:\Users\rohan\Documents\pedestrian_detection_pipeline>set PYTHONPATH=C:\Users\rohan\Documents\pedestrian_detection_pipeline\src

C:\Users\rohan\Documents\pedestrian_detection_pipeline>
C:\Users\rohan\Documents\pedestrian_detection_pipeline>python src/pipeline.py --runner DataflowRunner --project fine-cycling-451417-q7 --staging_location gs://fine-cycling-451417-q7-bucket/staging --temp_location gs://fine-cycling-451417-q7-bucket/temp --region northamerica-northeast2 --template_location gs://fine-cycling-451417-q7-bucket/templates/pedestrian_detection_template
Traceback (most recent call last):
  File "C:\Users\rohan\Documents\pedestrian_detection_pipeline\src\pipeline.py", line 5, in <module>
    from src.object_detection import detect_pedestrians
ModuleNotFoundError: No module named 'src'

C:\Users\rohan\Documents\pedestrian_detection_pipeline>
C:\Users\rohan\Documents\pedestrian_detection_pipeline>set PYTHONPATH=C:\Users\rohan\Documents\pedestrian_detection_pipeline

C:\Users\rohan\Documents\pedestrian_detection_pipeline>python src/pipeline.py --runner DataflowRunner --project fine-cycling-451417-q7 --staging_location gs://fine-cycling-451417-q7-bucket/staging --temp_location gs://fine-cycling-451417-q7-bucket/temp --region northamerica-northeast2 --template_location gs://fine-cycling-451417-q7-bucket/templates/pedestrian_detection_template
```

```
C:\Users\rohan\Documents\pedestrian_detection_pipeline>
C:\Users\rohan\Documents\pedestrian_detection_pipeline>curl -L -o models\yolov8n.pt https://github.com/ultralytics/yolo
v8/releases/latest/download/yolov8n.pt
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
100    9 100    9  0     0    28      0  0 --:--:-- --:--:-- --:--:--    28

C:\Users\rohan\Documents\pedestrian_detection_pipeline>python src/pipeline.py --runner DataflowRunner --project fine-cycling-451417-q7 --staging_location gs://fine-cycling-451417-q7-bucket/staging --temp_location gs://fine-cycling-451417-q7-bucket/temp --region northamerica-northeast2 --template_location gs://fine-cycling-451417-q7-bucket/templates/pedestrian_detection_template
```


