SHINDE SHUBHAM SUNIL
smail: me18b183@smail.iitm.ac.in
Course: CS4830 - Big Data Laboratory
Instructor: Prof. Balaraman Ravindran

Lab-6 Assignment
Roll No.: ME18B183
Semester: JAN-MAY 2022
Due Date: 27th March, 2022

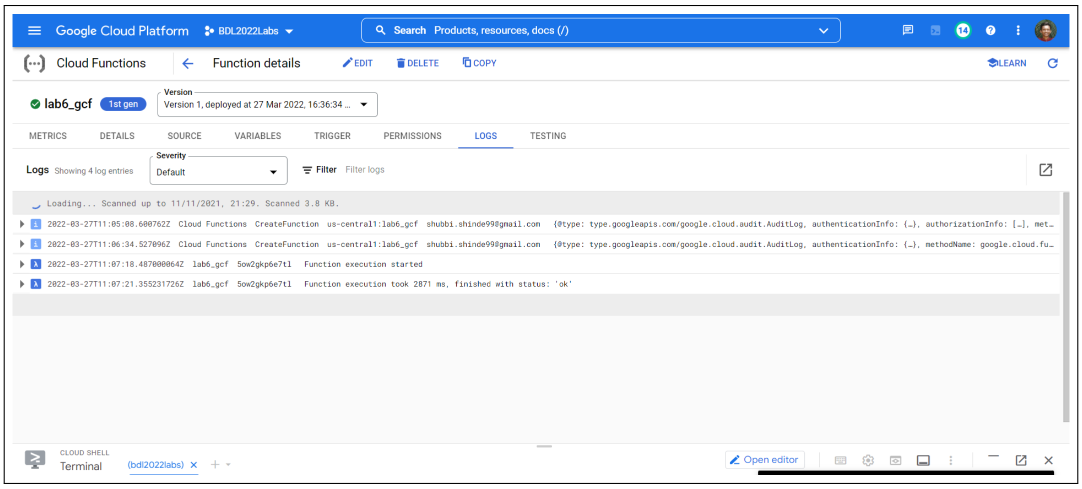
Problem 1

Write a Google cloud Function which gets triggered whenever a file is added to a bucket and publishes the file name to a topic in Pub/Sub.

Solution:

Python code (*main.py*):

```python
def lab6_gcf(data, context):
    from google.cloud import pubsub_v1

    publisher = pubsub_v1.PublisherClient()
    topic_name = "projects/bdl2022labs/topics/lab6_pubsub"
    topic_id = "lab6_pubsub"

    future = publisher.publish(topic_name, bytes(data['name'] ,'utf-8'))
    future.result()
```



```
Loading... Scanned up to 11/11/2021, 21:29. Scanned 3.8 KB.
2022-03-27T11:05:08.600762Z  Cloud Functions  CreateFunction  us-central1:lab6_gcf  shubbi.shinde99@gmail.com   {@type: type.googleapis.com/google.cloud.audit.AuditLog, authenticationInfo: {…}, authorizationInfo: […], met…
2022-03-27T11:06:34.527096Z  Cloud Functions  CreateFunction  us-central1:lab6_gcf  shubbi.shinde99@gmail.com   {@type: type.googleapis.com/google.cloud.audit.AuditLog, authenticationInfo: {…}, methodName: google.cloud.fu…
2022-03-27T11:07:18.487000064Z  lab6_gcf  5ow2gkp6e7tl   Function execution started
2022-03-27T11:07:21.355231726Z  lab6_gcf  5ow2gkp6e7tl   Function execution took 2871 ms, finished with status: 'ok'
```

Figure 1: Successful deployment of cloud function- *lab6_gcf*

Problem 2

Write a python file, which acts as a subscriber to this topic and prints out the number of lines in the file in real-time.

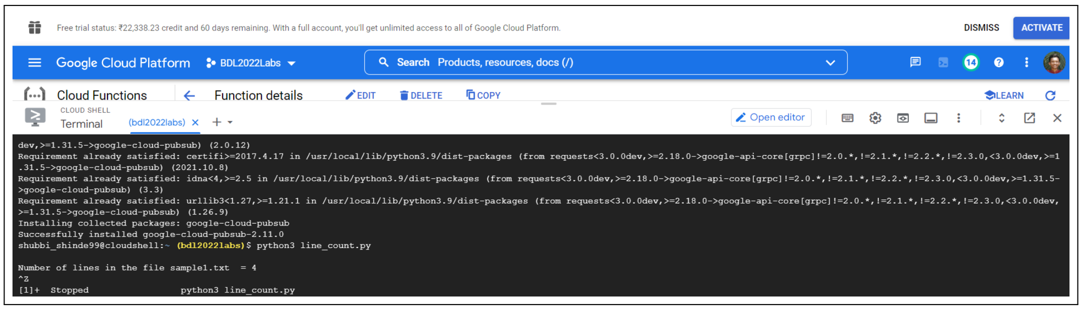
Solution:

Python code (*line_count.py*):

```python
import os

from google.cloud import storage
from google.cloud import pubsub_v1

client = storage.Client()
bucket = client.get_bucket('me18b183_bdl')

```

```python
subscriber = pubsub_v1.SubscriberClient()

topic_name = 'projects/bdl2022labs/topics/lab6_pubsub'
subscription_name = 'projects/bdl2022labs/subscriptions/lab6_pubsub-sub'

def line_print(message):
    blob = bucket.get_blob(message.data.decode('utf-8'))
    text = blob.download_as_string()
    text = text.decode('utf-8')
    print('\nNumber of lines in the file', message.data.decode('utf-8'),' =',
        len(text.split('\n')))
    message.ack()

future = subscriber.subscribe(subscription_name, line_print)

try:
    future.result()
except KeyboardInterrupt:
    future.cancel()
```

Number of lines in the *sample1.txt* file: 4.



Free trial status: ₹22,338.23 credit and 60 days remaining. With a full account, you'll get unlimited access to all of Google Cloud Platform.
DISMISS
ACTIVATE
Google Cloud Platform
BDL2022Labs
Search Products, resources, docs (/)
Cloud Functions
Function details
EDIT
DELETE
COPY
LEARN
CLOUD SHELL
Terminal
(bdl2022labs)
Open editor

```
dev,>=1.31.5->google-cloud-pubsub) (2.0.12)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0dev,>=2.18.0->google-api-core[grpc]!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0dev,>=1
.31.5->google-cloud-pubsub) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0dev,>=2.18.0->google-api-core[grpc]!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0dev,>=1.31.5-
>google-cloud-pubsub) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0dev,>=2.18.0->google-api-core[grpc]!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0dev,
>=1.31.5->google-cloud-pubsub) (1.26.9)
Installing collected packages: google-cloud-pubsub
Successfully installed google-cloud-pubsub-2.11.0
shubbi_shinde99@cloudshell:~ (bdl2022labs)$ python3 line_count.py

Number of lines in the file sample1.txt  = 4
^Z
[1]+  Stopped                 python3 line_count.py
```

Figure 2: Total number of lines in *sample1.txt* file.

Problem 3
There are two kinds of subscribers- pull and push subscribers. What are the differences between the two and when would you prefer one over the other?

Solution:

Pull subscription	**Push subscription**
Subscriber application initiates the requests to the Pub/Sub server to retrieve messages.	Pub/Sub initiates requests to your subscriber application to deliver the messages.
Achieves high throughput at low bandwidth by allowing batched delivery & acknowledgments as well as massively parallel consumption.	Delivers one message per request and limits maximum number of outstanding messages.
[**Preference**] Used when efficiency and throughput of message processing is critical.	[**Preference**] App Engine Standard and Cloud Functions subscribers.
[**Preference**] Large volume of messages (many more than 1/second).	[**Preference**] Multiple topics that must be processed by the same webhook.