

# Cloud Computing (SOFE 4630U)

# Project Milestone 1: Data Ingestion Software (Cloud Pub/Sub)

Name	Student ID
Rohan Radadiya	100704614

Due Date: January 27, 2025

Show inherited roles in table
Display roles inherited from the parent resources in the table below

Filter Enter property name or value

Owner (1)

#### GitHub Link:

https://github.com/rohanradadiya/Cloud-Computing-Project-MS-1/tree/main

## **Video Link 1:**

https://youtu.be/HpQnY2xQqJs

## Video Link 2:

https://youtu.be/dngTpA5jiko

EXPORT DATA

SNAPSHOTS

METRICS

messages to BigQuery or Cloud Storage by creating a subscription from a Cloud Dataflow job. Learn more 2

DETAILS

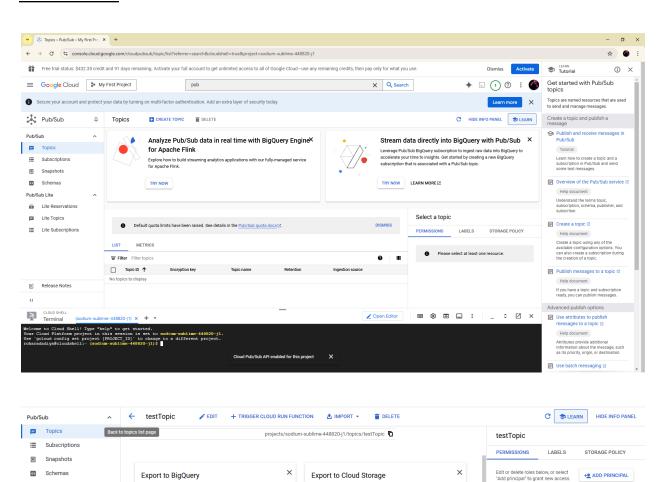
projects/sodium-sublime-448820-j1/subscriptions/testTopic-sub sodium-sublime-448820-j1

Only subscriptions attached to this topic are displayed. A subscription captures the stream of messages published to a given topic. You can also stream

Release Notes

Ž

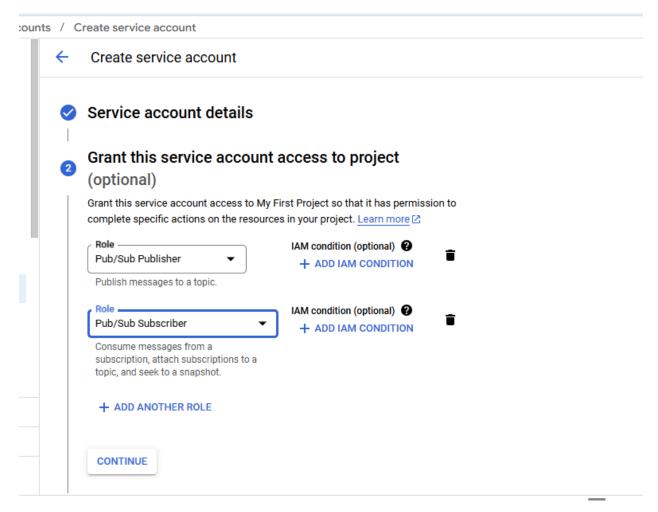
### **Screenshots:**



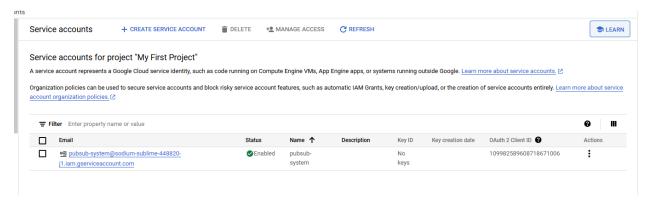
EXPORT DATA

MESSAGES

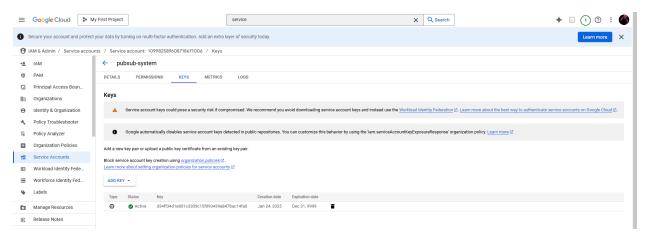
# testTopic was created:



# Creating service accounts with the two roles:

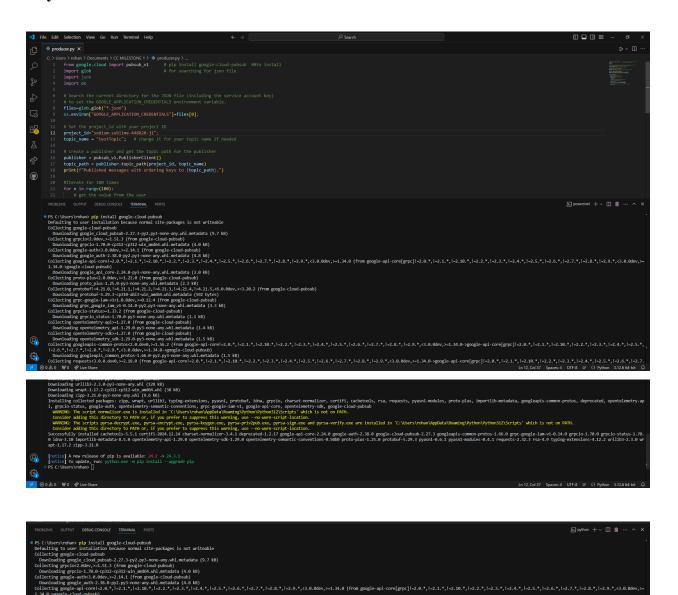


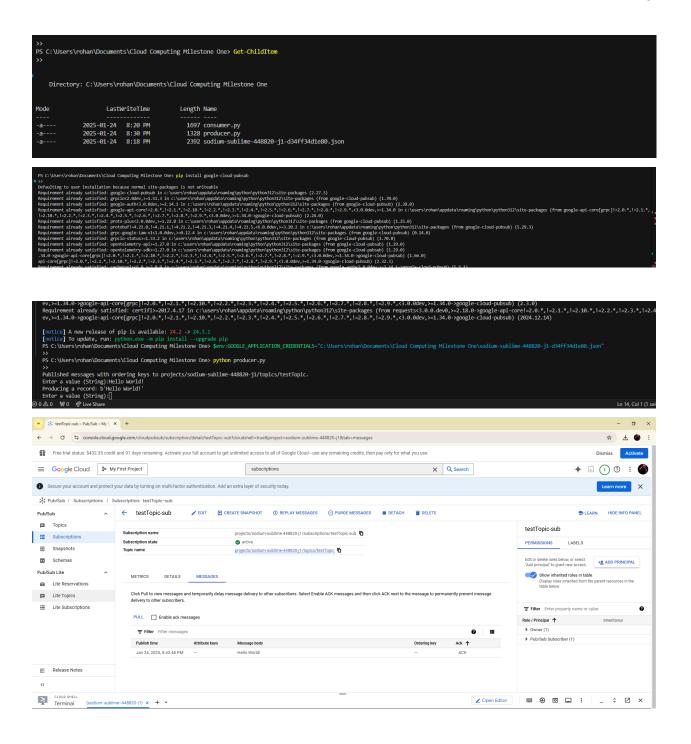
**Service account was created:** 



## **Key was created:**

re-2.24.0-py3-none-any.whl.metadata (3.0 kB) lev,>=1.22.0 (from google-cloud-pubsub) 5.0-py3-none-any.whl metadata (2.2 kB)





# Message was pulled ("Hello World!"):

```
| Section | Sect
```

## Producer/Consumer interaction.

V2 files and edits.

```
File "C:Users'rohan/Agoptat/Noaming/Python/Pythoni2\site-packages\google\arth\default.py", line 719, in default
raise exceptions.DefaultCredentialsFrom? (CLOD_SOK_MISSING_CREDRITALS)
google.auth.exceptions.DefaultCredentialsFrom? CutOD_SOK_MISSING_CREDRITALS)
SC:USers'rohan/Documents\CC Milestonety2\sedim=-befault Credentials, see https://cloud.google.com/docs/authentication/external/set-up-adc for more information.
PS:C:USers'rohan/Documents\CC Milestonety2\sedim=-befaultCredentialsFrom? (CutO_SOK_MISSING_CREDRITALS-"C:USers\rohan/Documents\CC Milestonety2\sedim=-sublime-448820-j1-d34ff34dle80.json"
>>
SC:USers\rohan/Documents\CC Milestonety2\sedim=-befaultCredentialsFrom(CutO_SOK_MISSING_CREDRITALS-"C:USers\rohan/Documents\CC Milestonety2\sedim=-sublime-448820-j1-d34ff34dle80.json"
>>
SC:USers\rohan/Documents\CC Milestonety2\sedim=-sublime-448820-j1-d34ff34dle80.json"
>>
SC:USers\rohan/Documents\CC Milestonety2\sedim=-sublime-448820-j1-d34ff34dle80.json"
>>
SC:USers\rohan/Documents\CC Milestonety2\sedim=-sublime-448820-j1-d34ff34ff34dle80.json"
>>
SC:USers\rohan/Documents\CC Milestonety2\sedi
```

V2 files, and the smartMeter.py and consumer.py files communicating as required.

#### **Discussion:**

#### 1. What is EDA? What are its advantages and disadvantages?

EDA stands for "Exploratory Data Analysis". It represents an approach in data analysis that has to do with summarizing and visualizing data to identify patterns and relationships. Statistics and graphs are also utilized to better understand the data's structure and further analyze it. Some advantages of EDA are that it greatly helps with better understanding the data and identifying outliers or missing data as well. EDA also helps to make certain predictions of the data and the visualization tools make it easier to understand the data better and this also helps to communicate the data better to stakeholders. One disadvantage of EDA is that it can be very time-consuming. It can also be very difficult when dealing with large datasets, and there is usually a need for a deep understanding of the technology and tools to do this.

2. Cloud Pub/Sub has two types of subscriptions: push and pull. Describe them, showing the strengths and weaknesses of each based on potential applications.

A pull subscription has to do with subscribers requesting messages from the Pub/Sub server, and the subscriber is also able to control when and how often to pull the messages. Some scenarios when a pull subscription might be useful is when the messages need to be processed in batches rather than in real-time, and when subscribers need to pull the messages when they require, depending on the workloads. Some weaknesses may be that there may be high latency and more delays, and also an increased level of complexity.

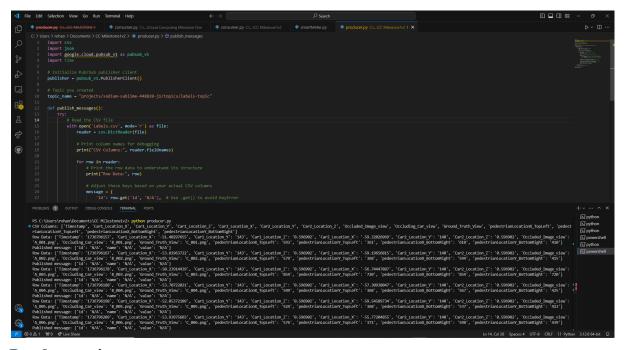
On the other hand, a pull subscription has to do with the process that Pub/Sub pushes the messages automatically when the messages arrive, to an HTTP/HTTPS endpoint. Some strengths are that there is low latency, and the implementation is simpler, but some disadvantages are that there has to be an available endpoint to be able to receive the messages. The complexity can also be increased, and there may be scalability issues due to being able to handle an unexpected amount of messages at certain times. Some applications of this can be for real-time processing scenarios and also when there is communication between microservices in distributed systems.

3. When publishing a message into a topic, an ordering key can be specified. Using examples, describe the role and benefits of ordering keys.

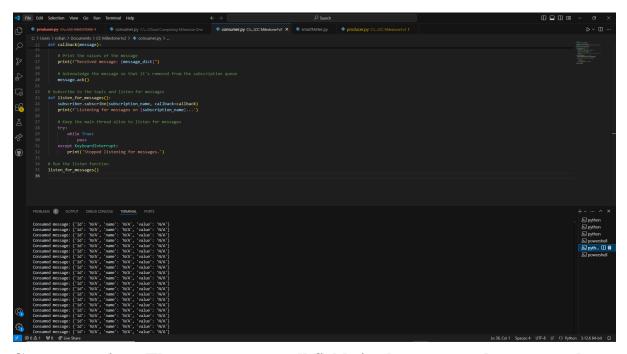
To start, ordering keys in Pub/Sub is useful when applications require the messages to be processed in a specific order. With the use of ordering keys, companies, businesses, and other organizations are able to ensure consistency, simplify logic, etc. The way it works is that the publisher specifies an ordering key when sending a message to the topic. With the ordering key, the Pub/Sub makes sure that the messages with the same ordering key are delivered in the correct order that they were sent. If there is a scenario where there is a delivery fail, the Pub/Sub holds

certain messages with the same key until the previous messages are acknowledged/confirmed. An example of when ordering keys may be used is during financial transactions, delivery companies, and even IoT-related systems.

# **Design:**



# **Producer view:**



Consumer view: There are many null fields in the consumed messages, but this may be due to the connectivity and functionality between the .csv file and the other corresponding project files.