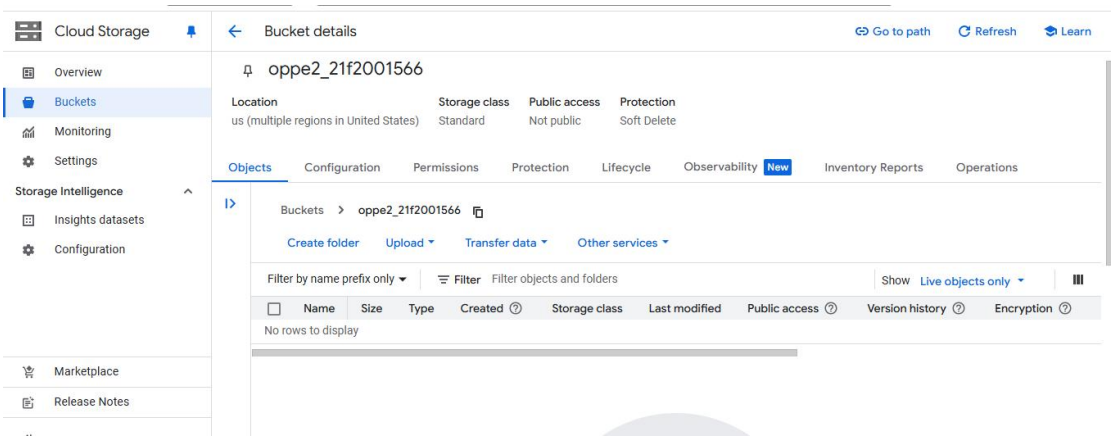
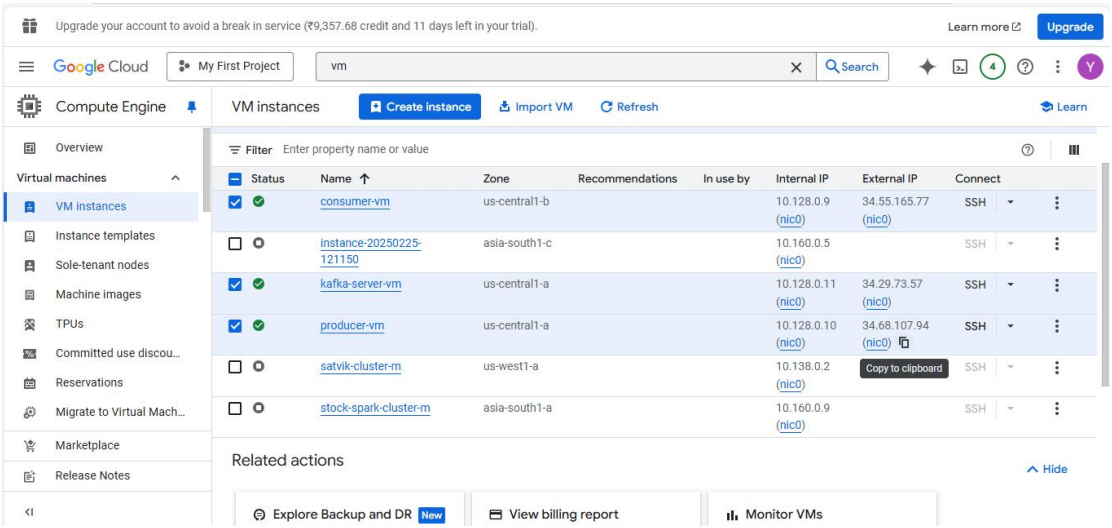
Oppe 2:

Name: Yashawini
Roll No:21f2001566

STEP 1: Upload Your ZIP File to GCS



STEP 2: SSH into Producer VM



```
ssh.cloud.google.com/v2/ssh/projects/data-mind-448014-b3/zones/us-central1-a/instances/producer-vm?authuser=0&hl=en_US&projec...
https://ssh.cloud.google.com/v2/ssh/projects/data-mind-448014-b3/zones/us-central1-a/instances/producer-vm?authuser=0...
SSH-in-browser                                   UPLOAD FILE   DOWNLOAD FILE

Linux producer-vm 5.10.0-34-cloud-amd64 #1 SMP Debian 5.10.234-1 (2025-02-24) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Apr  5 11:30:18 2025 from 35.235.245.128
yashawini0704@producer-vm:~$
```

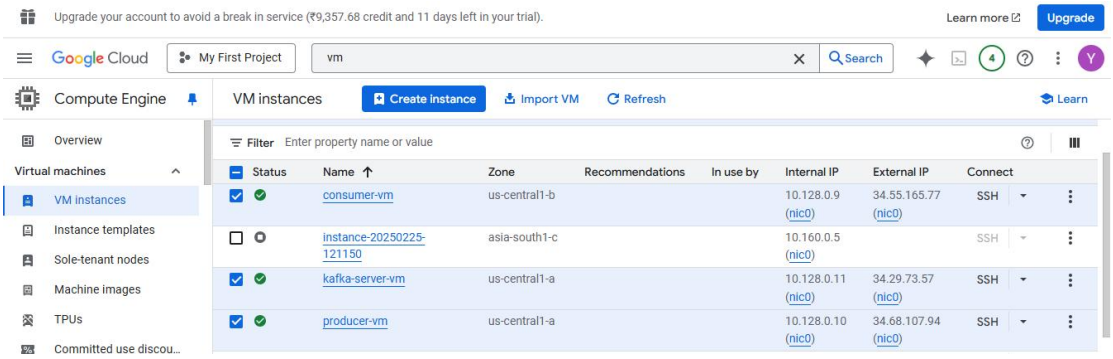
STEP 3: Copy ZIP from GCS to Producer VM

```
yashawini0704@producer-vm:~$ gsutil cp gs://oppe2_21f2001566/Train_details_22122017_Set_A.csv.zip ~/
Copying gs://oppe2_21f2001566/Train_details_22122017_Set_A.csv.zip...
/ [1 files][  2.4 MiB/  2.4 MiB]
Operation completed over 1 objects/2.4 MiB.
yashawini0704@producer-vm:~$
```

STEP 4: Unzip the File

```
NSE_Stocks_Data                                  batchstream.py          spark
NSE_Stocks_Data-20250405T085349Z-001.zip         demo.py                 spark-3.5.5-bin-hadoop3.tgz
NSE_Stocks_Data-20250405T085349Z-001.zip.csupload install_dependencies.sh
Train_details_22122017_Set_A.csv.zip             producer.py
yashawini0704@producer-vm:~$ unzip Train_details_22122017_Set_A.csv.zip
Archive:  Train_details_22122017_Set_A.csv.zip
  inflating: Train_details_22122017.csv
  inflating: __MACOSX/._Train_details_22122017.csv
yashawini0704@producer-vm:~$
```

Step 5 :
Created Vm for Consumer,Kafka and Producer separately



Step 6:

SSH into kafka vm

```
CLOUD SHELL
Terminal   (data-mind-448014-b3)                                   Open Editor

Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to data-mind-448014-b3.
Use `gcloud config set project [PROJECT_ID]` to change to a different project.
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ ./ssh_kafka_vm.sh
Enter passphrase for key '/home/yashawini0704/.ssh/google_compute_engine':
Linux kafka-server-vm 5.10.0-34-cloud-amd64 #1 SMP Debian 5.10.234-1 (2025-02-24) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Apr  5 11:29:12 2025 from 35.198.201.30
yashawini0704@kafka-server-vm:~$
```

Step 7:

Now start zookeeper

```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Apr  5 11:29:12 2025 from 35.198.201.30
yashawini0704@kafka-server-vm:~$ ls
create_kafka_topic.sh  install_dependencies.sh  kafka  kafka_2.13-3.7.2.tgz  start_kafka.sh  start_zookeeper.sh
yashawini0704@kafka-server-vm:~$ ./start_zookeeper.sh
[2025-04-06 09:10:46,837] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,839] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,843] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,843] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,843] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,843] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig
)
```

Zookeeper Running:

```
CLOUD SHELL
Terminal (data-mind-448014-b3) ×                                                    Open Editor
[2025-04-06 09:10:46,851] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,851] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2025-04-06 09:10:46,851] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig
)
[2025-04-06 09:10:46,851] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2025-04-06 09:10:46,866] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@1460a8c0 (org.apache.zookeeper.server.Ser
verMetrics)
[2025-04-06 09:10:46,870] INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2025-04-06 09:10:46,870] INFO zookeeper.DigestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2025-04-06 09:10:46,875] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2025-04-06 09:10:46,892] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO                                                      (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO  |  ___ /                |_ |                      (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO    / / ___   ___   | | __  ___  ___  _ __   ___  _ __  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO   / /  / _ \ / _ \ | |/ /  / _ \ / _ \ | '_ \ / _ \ | '__| (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO  / /__ | (_) | (_) | |  <  |  __/ |  __/ | |_) | |_) | |  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO  /____| \___/  \___/ |_|\_\ \___| | ___/  \___| |_| (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO                                  | |                  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO                                  |_|                  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,892] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,896] INFO Server environment:zookeeper.version=3.8.4-9316c2a7a97e1666d8f4593f34dd6fc36ecc436c, built on 2024-02-12 22:16 UTC (org.apache.zookeeper.s
erver.ZooKeeperServer)
[2025-04-06 09:10:46,896] INFO Server environment:host.name=kafka-server-vm.us-central1-a.c.data-mind-448014-b3.internal (org.apache.zookeeper.server.ZooKeeperServer)
[2025-04-06 09:10:46,896] INFO Server environment:java.version=11.0.26 (org.apache.zookeeper.server.ZooKeeperServer)
```

Now open another terminal and ssh into kafka_vm

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to data-mind-448014-b3.
Use `gcloud config set project [PROJECT_ID]` to change to a different project.
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ ./ssh_kafka_vm.sh
Enter passphrase for key '/home/yashawini0704/.ssh/google_compute_engine':
Linux kafka-server-vm 5.10.0-34-cloud-amd64 #1 SMP Debian 5.10.234-1 (2025-02-24) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sun Apr  6 09:09:38 2025 from 35.197.131.209
yashawini0704@kafka-server-vm:~$ ls
create_kafka_topic.sh  install_dependencies.sh  kafka  kafka_2.13-3.7.2.tgz  start_kafka.sh  start_zookeeper.sh
yashawini0704@kafka-server-vm:~$
```

Now start_kafka.sh server

```
CLOUD SHELL
Terminal (data-mind-448014-b3) ×  (data-mind-448014-b3) ×                           Open Editor
create_kafka_topic.sh  install_dependencies.sh  kafka  kafka_2.13-3.7.2.tgz  start_kafka.sh  start_zookeeper.sh
yashawini0704@kafka-server-vm:~$ ./start_kafka.sh
[2025-04-06 09:14:28,760] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2025-04-06 09:14:29,153] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509
Util)
[2025-04-06 09:14:29,259] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2025-04-06 09:14:29,261] INFO starting (kafka.server.KafkaServer)
[2025-04-06 09:14:29,261] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2025-04-06 09:14:29,278] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2025-04-06 09:14:29,281] INFO Client environment:zookeeper.version=3.8.4-9316c2a7a97e1666d8f4593f34dd6fc36ecc436c, built on 2024-02-12 22:16 UTC (org.apache.zookeeper.Z
ooKeeper)
[2025-04-06 09:14:29,281] INFO Client environment:host.name=kafka-server-vm.us-central1-a.c.data-mind-448014-b3.internal (org.apache.zookeeper.ZooKeeper)
[2025-04-06 09:14:29,281] INFO Client environment:java.version=11.0.26 (org.apache.zookeeper.ZooKeeper)
[2025-04-06 09:14:29,281] INFO Client environment:java.vendor=Debian (org.apache.zookeeper.ZooKeeper)
[2025-04-06 09:14:29,281] INFO Client environment:java.home=/usr/lib/jvm/java-11-openjdk-amd64 (org.apache.zookeeper.ZooKeeper)
[2025-04-06 09:14:29,281] INFO Client environment:java.class.path=/home/yashawini0704/kafka/bin/../libs/activation-1.1.1.jar:/home/yashawini0704/kafka/bin/../libs/aopall
iance-repackaged-2.6.1.jar:/home/yashawini0704/kafka/bin/../libs/argparse4j-0.7.0.jar:/home/yashawini0704/kafka/bin/../libs/audience-annotations-0.12.0.jar:/home/yashawi
ni0704/kafka/bin/../libs/caffeine-2.9.3.jar:/home/yashawini0704/kafka/bin/../libs/checker-qual-3.19.0.jar:/home/yashawini0704/kafka/bin/../libs/commons-beanutils-1.9.4.j
ar:/home/yashawini0704/kafka/bin/../libs/commons-cli-1.4.jar:/home/yashawini0704/kafka/bin/../libs/commons-collections-3.2.2.jar:/home/yashawini0704/kafka/bin/../libs/co
mmons-digester-2.1.jar:/home/yashawini0704/kafka/bin/../libs/commons-io-2.14.0.jar:/home/yashawini0704/kafka/bin/../libs/commons-lang3-3.8.1.jar:/home/yashawini0704/kafk
a/bin/../libs/commons-logging-1.2.jar:/home/yashawini0704/kafka/bin/../libs/commons-validator-1.7.jar:/home/yashawini0704/kafka/bin/../libs/connect-api-3.7.2.jar:/home/y
ashawini0704/kafka/bin/../libs/connect-basic-auth-extension-3.7.2.jar:/home/yashawini0704/kafka/bin/../libs/connect-json-3.7.2.jar:/home/yashawini0704/kafka/bin/../libs/
connect-mirror-3.7.2.jar:/home/yashawini0704/kafka/bin/../libs/connect-mirror-client-3.7.2.jar:/home/yashawini0704/kafka/bin/../libs/connect-runtime-3.7.2.jar:/home/yash
awini0704/kafka/bin/../libs/connect-transforms-3.7.2.jar:/home/yashawini0704/kafka/bin/../libs/error_prone_annotations-2.10.0.jar:/home/yashawini0704/kafka/bin/../libs/h
```

Now open another terminal and create a Kafka topic

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to data-mind-448014-b3.
Use `gcloud config set project [PROJECT_ID]` to change to a different project.
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ ./ssh_kafka_vm.sh
Enter passphrase for key '/home/yashawini0704/.ssh/google_compute_engine':
Linux kafka-server-vm 5.10.0-34-cloud-amd64 #1 SMP Debian 5.10.234-1 (2025-02-24) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sun Apr  6 09:12:56 2025 from 35.197.131.209
yashawini0704@kafka-server-vm:~$ ls
create_kafka_topic.sh  install_dependencies.sh  kafka  kafka_2.13-3.7.2.tgz  start_kafka.sh  start_zookeeper.sh
yashawini0704@kafka-server-vm:~$ ./create_kafka_topic.sh
Usage: ./create_kafka_topic.sh <topic-name> [partitions] [replication-factor]
yashawini0704@kafka-server-vm:~$  ./create_kafka_topic.sh input-1 1
```

```
create_kafka_topic.sh  install_dependencies.sh  kafka  kafka_2.13-3.7.2.tgz  start_kafka.sh  start_zookeeper.sh
yashawini0704@kafka-server-vm:~$ ./create_kafka_topic.sh
Usage: ./create_kafka_topic.sh <topic-name> [partitions] [replication-factor]
yashawini0704@kafka-server-vm:~$  ./create_kafka_topic.sh input-1 1
Checking Kafka broker at 34.29.73.57:9092...
Connection to 34.29.73.57 9092 port [tcp/*] succeeded!
Creating Kafka topic: input-1 with 1 partitions and 1 replication factor...
Created topic input-1.
Verifying topic creation...
input-1
Kafka topic 'input-1' created successfully!
yashawini0704@kafka-server-vm:~$
```

Now ssh into producer and create a python file for producer code:

nano train_producer.py

```
yashawini0704@producer-vm:~$ nano train_ producer.py
```

Add code to the file:

ssh.cloud.google.com/v2/ssh/projects/data-mind-448014-b3/zones/us-central1-a/instances/producer-vm?authuser=0&hl=en_US&projec...

https://ssh.cloud.google.com/v2/ssh/projects/data-mind-448014-b3/zones/us-central1-a/instances/producer-vm?authuser=0...

SSH-in-browser UPLOAD FILE DOWNLOAD FILE

```
  [1/2]                                          train_ *
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
import time
import os
from kafka import KafkaProducer
import json

# 🔧CONFIGURATION
KAFKA_BROKER = "34.29.73.57:9092"  # Replace with your Kafka VM internal IP
TOPIC = "input-1"
CSV_FOLDER = "./"  # Current folder contains the CSV

# 🚀Spark Session
spark = SparkSession.builder \
    .appName("TrainCSVToKafkaStreamer") \
    .getOrCreate()

# 🛠️Kafka Producer
producer = KafkaProducer(
    bootstrap_servers=KAFKA_BROKER,
    value_serializer=lambda v: json.dumps(v).encode('utf-8')
)

# 📂Look for CSV in folder
for file in os.listdir(CSV_FOLDER):
    if file.endswith(".csv"):
        full_path = os.path.join(CSV_FOLDER, file)
        print(f"\n📂Processing: {full_path}")

        # 📩Load CSV using Spark
        df = spark.read.option("header", True).csv(full_path)
        print(f"🔍Raw rows in {file}: {df.count()}")

        # 🧹Clean: remove bad rows, parse fields
        df = df.select(
```

```
yashawini0704@producer-vm:~$ python3 train_producer.py
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/04/06 09:38:10 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
-java classes where applicable
✅Cleaned rows: 186124
✅Sent: {'station_code': 'SWV', 'train_no': '107', 'train_name': 'SWV-MAO-VLNK', 'arrival': '00:00:00', 'departu
re': '10:25:00'}
✅Sent: {'station_code': 'THVM', 'train_no': '107', 'train_name': 'SWV-MAO-VLNK', 'arrival': '11:06:00', 'depart
ure': '11:08:00'}
✅Sent: {'station_code': 'KRMI', 'train_no': '107', 'train_name': 'SWV-MAO-VLNK', 'arrival': '11:28:00', 'depart
ure': '11:30:00'}
✅Sent: {'station_code': 'MAO', 'train_no': '107', 'train_name': 'SWV-MAO-VLNK', 'arrival': '12:10:00', 'departu
re': '00:00:00'}
✅Sent: {'station_code': 'MAO', 'train_no': '108', 'train_name': 'VLNK-MAO-SWV', 'arrival': '00:00:00', 'departu
re': '20:30:00'}
```

Now ssh into consumer vm

```
CLOUD SHELL
Terminal    a-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3   Open Editor

Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to data-mind-448014-b3.
Use `gcloud config set project [PROJECT_ID]` to change to a different project.
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ ./ssh_consumer_vm.sh
Enter passphrase for key '/home/yashawini0704/.ssh/google_compute_engine':
Linux consumer-vm 5.10.0-34-cloud-amd64 #1 SMP Debian 5.10.234-1 (2025-02-24) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Apr  5 11:35:12 2025 from 35.198.201.30
yashawini0704@consumer-vm:~$
```

Create nano train_consumer.py

```
CLOUD SHELL
Terminal    a-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3   Open Editor

  GNU nano 5.4                                   train_consumer.py *
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.types import StructType, StringType

# 🔧 Kafka Config
KAFKA_BROKER = "34.29.73.57:9092"  # Your Kafka VM internal IP
TOPIC = "input-1"

# 🚀 Spark Session
spark = SparkSession.builder \
    .appName("TrainStationCongestionConsumer") \
    .getOrCreate()

spark.sparkContext.setLogLevel("WARN")

# 📦 Define schema for JSON values from Kafka
schema = StructType() \
    .add("station_code", StringType()) \
    .add("train_no", StringType()) \
    .add("train_name", StringType()) \
    .add("arrival", StringType()) \
    .add("departure", StringType())

^G Help      ^O Write Out   ^W Where Is    ^K Cut       ^T Execute   ^C Location   M-U Undo     M-A Set Mark   M-] To Bracket  M-Q Previous
^X Exit      ^R Read File   ^\ Replace     ^U Paste     ^J Justify   ^_ Go To Line  M-E Redo     M-6 Copy       ^Q Where Was    M-W Next
```

Submit spark job

```
CLOUD SHELL
Terminal    a-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3   Open Editor

  GNU nano 5.4                                   run_consumer.sh

export PYSPARK_SUBMIT_ARGS="--packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.5.5 pyspark-shell"
python3 train_consumer.py
```

Output:

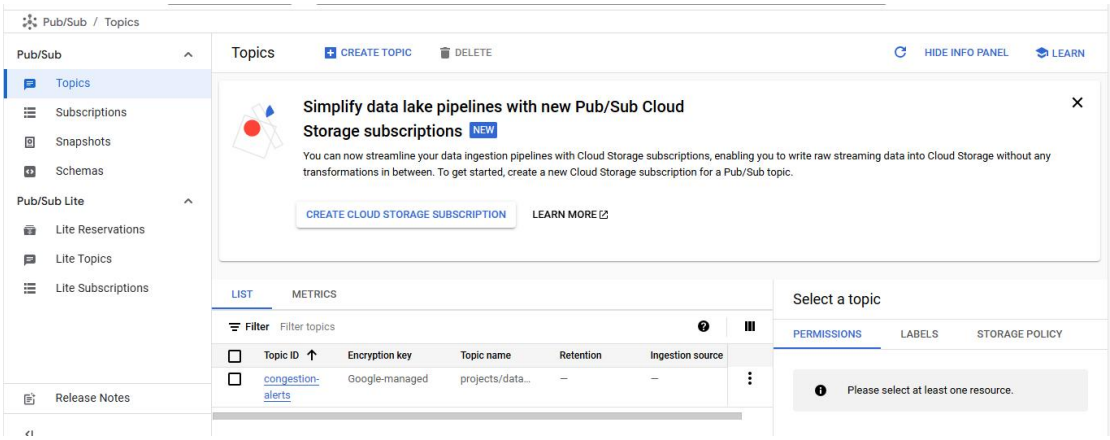
```
-------------------------------------------
Batch: 7
-------------------------------------------
+------------------------------------------+------------+-----------+-------------------------+
|window                                    |station_code|train_count|congestion_alert         |
+------------------------------------------+------------+-----------+-------------------------+
|{2025-04-06 09:45:00, 2025-04-06 10:05:00}|KGP         |8          |🚨Congested Train Station|
|{2025-04-06 09:30:00, 2025-04-06 09:50:00}|KGP         |8          |🚨Congested Train Station|
|{2025-04-06 09:35:00, 2025-04-06 09:55:00}|SRC         |8          |🚨Congested Train Station|
|{2025-04-06 09:40:00, 2025-04-06 10:00:00}|CTC         |2          |                         |
|{2025-04-06 09:30:00, 2025-04-06 09:50:00}|SRC         |8          |🚨Congested Train Station|
|{2025-04-06 09:45:00, 2025-04-06 10:05:00}|BBS         |2          |                         |
|{2025-04-06 09:45:00, 2025-04-06 10:05:00}|SRC         |8          |🚨Congested Train Station|
|{2025-04-06 09:45:00, 2025-04-06 10:05:00}|BHC         |2          |                         |
|{2025-04-06 09:35:00, 2025-04-06 09:55:00}|BBS         |2          |                         |
|{2025-04-06 09:35:00, 2025-04-06 09:55:00}|CTC         |2          |                         |
|{2025-04-06 09:30:00, 2025-04-06 09:50:00}|BHC         |2          |                         |
|{2025-04-06 09:40:00, 2025-04-06 10:00:00}|KGP         |8          |🚨Congested Train Station|
|{2025-04-06 09:40:00, 2025-04-06 10:00:00}|BHC         |2          |                         |
|{2025-04-06 09:45:00, 2025-04-06 10:05:00}|CTC         |2          |                         |
|{2025-04-06 09:30:00, 2025-04-06 09:50:00}|CTC         |2          |                         |
|{2025-04-06 09:35:00, 2025-04-06 09:55:00}|BHC         |2          |                         |
|{2025-04-06 09:30:00, 2025-04-06 09:50:00}|BBS         |2          |                         |
|{2025-04-06 09:40:00, 2025-04-06 10:00:00}|SRC         |8          |🚨Congested Train Station|
|{2025-04-06 09:40:00, 2025-04-06 10:00:00}|BBS         |2          |                         |
|{2025-04-06 09:35:00, 2025-04-06 09:55:00}|KGP         |8          |🚨Congested Train Station|
+------------------------------------------+------------+-----------+-------------------------+
```

Create Pub/Sub topic

```
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ gcloud pubsub topics create congestion-alerts
Created topic [projects/data-mind-448014-b3/topics/congestion-alerts].
```

```
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ gcloud compute instances describe consumer-vm \
  --zone=us-central1-b \
  --format='value(serviceAccounts.email)'
702363484992-compute@developer.gserviceaccount.com
```

give this account Pub/Sub publishing permission so your Spark job can send “Congested Train Station” alerts.




```
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ gcloud compute instances describe consumer-vm \
  --zone=us-central1-b \
  --format='value(serviceAccounts.email)'
702363484992-compute@developer.gserviceaccount.com
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ gcloud projects add-iam-policy-binding data-mind-448014-b3 \
  --member="serviceAccount:702363484992-compute@developer.gserviceaccount.com" \
  --role="roles/pubsub.publisher"
Updated IAM policy for project [data-mind-448014-b3].
bindings:
- members:
  - serviceAccount:service-702363484992@gcp-sa-artifactregistry.iam.gserviceaccount.com
  role: roles/artifactregistry.serviceAgent
- members:
  - serviceAccount:702363484992-compute@developer.gserviceaccount.com
  - serviceAccount:702363484992@cloudbuild.gserviceaccount.com
  role: roles/cloudbuild.builds.builder
- members:
  - serviceAccount:service-702363484992@gcp-sa-cloudbuild.iam.gserviceaccount.com
```

```
yashawini0704@cloudshell:~ (data-mind-448014-b3)$ nano pubsub_publisher.py
```

```
yashawini0704@consumer-vm:~$ ls
checkpoint                consumer.py   install_dependencies.sh  kafka_2.13-3.7.2.tgz  spark                          train_consumer.py
congestion_alerts_output  consumer2.py  kafka                    run_consumer.sh       spark-3.5.5-bin-hadoop3.tgz
yashawini0704@consumer-vm:~$ nano pubsub_publisher.py
```

Train CSV → Kafka (Producer)

↓

Structured Streaming (Consumer)

↓

- Console output
- JSON file output

pubsub_publisher.py → GCP Pub/Sub

```
CLOUD SHELL
Terminal   a-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3) ×   (data-mind-448014-b3   +   Open Editor
tation"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"CSMT","train_count":8,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"BDSW","train_count":12,"congestion_alert":"🚨Congested Train
 Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"BRC","train_count":12,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"ASR","train_count":6,"congestion_alert":"🚨Congested Train S
tation"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"RTGH","train_count":16,"congestion_alert":"🚨Congested Train
 Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"VSKP","train_count":6,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"KNW","train_count":12,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"PNVL","train_count":32,"congestion_alert":"🚨Congested Train
 Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"FGR","train_count":8,"congestion_alert":"🚨Congested Train S
tation"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"MTD","train_count":8,"congestion_alert":"🚨Congested Train S
tation"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"SWV","train_count":14,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"ANVT","train_count":8,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"JSG","train_count":16,"congestion_alert":"🚨Congested Train
Station"}
✅Published: {"window":{"start":"2025-04-06T10:15:00.000Z","end":"2025-04-06T10:35:00.000Z"},"station_code":"ET","train_count":14,"congestion_alert":"🚨Congested Train S
```