

Please install all Dependencies

1. `sudo apt update`
2. `sudo apt install -y tmux vim`
3. `sudo apt-get install openjdk-11-jre openjdk-11-jdk`
4. `wget curl git python3-pip # python-is-python3 # (Optional)`
5. `pip install kafka-python pyspark pyspark[sql]`
6. `wget https://dlcdn.apache.org/kafka/3.4.0/kafka_2.13-3.4.0.tgz # Get Kafka Sources`

Once Done above then only follow below setups.

- A. Installing Kafka(Open New Terminal Session)
`tar -xzf kafka_2.13-3.4.0.tgz`
- B. Open New Terminal Session and Start zookeeper (Keep on running do not close)
`cd kafka_2.13-3.4.0`
`bin/zookeeper-server-start.sh config/zookeeper.properties`

[illegible]

- C. Open New Terminal Session and Start Server (Keep on running do not close)
- ```
cd kafka_2.13-3.4.0
bin/kafka-server-start.sh config/server.properties
```

[illegible]

- D. Open New Terminal Session (Create Events –Topic)
- ```
cd kafka_2.13-3.4.0
```
1. `bin/kafka-topics.sh --create --topic test-topic --bootstrap-server localhost:9092`
 2. **To check if my Topic is created or not**
`bin/kafka-topics.sh --describe --topic test-topic --bootstrap-server localhost:9092`

E. Open New Terminal Session (Execute Producer.sh)

It will read data from GCP and generate the message for 10 record with 10 sec sleep.

```
g2if1000350@instance-1:~/kafka_2.13-3.4.0/demo/kafka_producer_consumer$ python3 kafka_producer_gcp.py
Kafka Producer Application Started ...
Preparing message: 0
Message: {'id': 1, 'msg': ['Line1\r', 'Line2\r', 'Line3\r', 'Line4\r', 'Line5\r', 'Line6\r', 'Line7\r', 'Line8\r', 'Line9\r', 'Line10\r'], 'no_of_line': 10}
Preparing message: 10
Message: {'id': 1, 'msg': ['Line11\r', 'Line12\r', 'Line13\r', 'Line14\r', 'Line15\r', 'Line16\r', 'Line17\r', 'Line18\r', 'Line19\r', 'Line20\r'], 'no_of_line': 10}
Preparing message: 20
Message: {'id': 1, 'msg': ['Line21\r', 'Line22\r', 'Line23\r', 'Line24\r', 'Line25\r', 'Line26\r', 'Line27\r', 'Line28\r', 'Line29\r', 'Line30\r'], 'no_of_line': 10}
Preparing message: 30
Message: {'id': 1, 'msg': ['Line31\r', 'Line32\r', 'Line33\r', 'Line34\r', 'Line35\r', 'Line36\r', 'Line37\r', 'Line38\r', 'Line39\r', 'Line40\r'], 'no_of_line': 10}
Preparing message: 40
Message: {'id': 1, 'msg': ['Line41\r', 'Line42\r', 'Line43\r', 'Line44\r', 'Line45\r', 'Line46\r', 'Line47\r', 'Line48\r', 'Line49\r', 'Line50\r'], 'no_of_line': 10}
Preparing message: 50
Message: {'id': 1, 'msg': ['Line51\r', 'Line52\r', 'Line53\r', 'Line54\r', 'Line55\r', 'Line56\r', 'Line57\r', 'Line58\r', 'Line59\r', 'Line60\r'], 'no_of_line': 10}
Preparing message: 60
Message: {'id': 1, 'msg': ['Line61\r', 'Line62\r', 'Line63\r', 'Line64\r', 'Line65\r', 'Line66\r', 'Line67\r', 'Line68\r', 'Line69\r', 'Line70\r'], 'no_of_line': 10}
Preparing message: 70
Message: {'id': 1, 'msg': ['Line71\r', 'Line72\r', 'Line73\r', 'Line74\r', 'Line75\r', 'Line76\r', 'Line77\r', 'Line78\r', 'Line79\r', 'Line80\r'], 'no_of_line': 10}
Preparing message: 80
Message: {'id': 1, 'msg': ['Line81\r', 'Line82\r', 'Line83\r', 'Line84\r', 'Line85\r', 'Line86\r', 'Line87\r', 'Line88\r', 'Line89\r', 'Line90\r'], 'no_of_line': 10}
Preparing message: 90
Message: {'id': 1, 'msg': ['Line91\r', 'Line92\r', 'Line93\r', 'Line94\r', 'Line95\r', 'Line96\r', 'Line97\r', 'Line98\r', 'Line99\r', 'Line100\r'], 'no_of_line': 10}
Kafka Producer Application Completed.
g2if1000350@instance-1:~/kafka_2.13-3.4.0/demo/kafka_producer_consumer$
```

F. Open New Terminal Session (Execute Produce.sh)

It will read data from Producer and count no of lines every 5 sec.

Execute

spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.3.2

kafka_streaming_json_demo_gcp.py

```
+---+-----+
|id|total_line_count|
+---+-----+
+---+-----+

Batch: 1
-----
+---+-----+
|id|total_line_count|
+---+-----+
|1|10.0|
+---+-----+

Batch: 2
-----
+---+-----+
|id|total_line_count|
+---+-----+
|1|20.0|
+---+-----+

Batch: 3
-----
+---+-----+
|id|total_line_count|
+---+-----+
|1|40.0|
+---+-----+
```