In [1]:
```
## Author: Yam Jason
```

In [1]:
```
from pyspark.sql import SparkSession

spark = SparkSession\
        .builder\
        .appName("Producer")\
        .getOrCreate()
```

```
24/09/07 15:09:02 WARN Utils: Your hostname, WeirdSmile. resolves to a loopback addr
ess: 127.0.1.1; using 10.255.255.254 instead (on interface lo)
24/09/07 15:09:02 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another addr
ess
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(ne
wLevel).
24/09/07 15:09:03 WARN NativeCodeLoader: Unable to load native-hadoop library for yo
ur platform... using builtin-java classes where applicable
```

Load predictions from JSON

In [2]:
```
sc = spark.sparkContext
sc.addFile("../../../de_classes/data_storage/hadoop_file_handler.py")

# Import the HadoopFileHandler class
from hadoop_file_handler import HadoopFileHandler

# Create an instance of HadoopFileHandler
handler = HadoopFileHandler()

# Read raw data from HDFS
df = handler.read_json('data/predictions/predictions3.json')
```

```
24/09/07 15:09:07 WARN SparkSession: Using an existing Spark session; only runtime S
QL configurations will take effect.
```

In [3]:
```
df.show(10, truncate=False)
```

```
+--------------------------------------------------------------------------------
--------------------------------------------------------------------+----------+
|Review
|prediction|
+--------------------------------------------------------------------------------
--------------------------------------------------------------------+----------+
|all item are good but get i order
|2.0       |
|good seller delivery man parcel good condition
|2.0       |
|comfortable mask
|2.0       |
|so good
|2.0       |
|best best best
|2.0       |
|v good tq
|2.0       |
|i am so happy i have received goods today i am very satisfied happy
|2.0       |
|fast delivery good quality nice packing will order again
|2.0       |
|received good condition
|2.0       |
|item not enough oder got je chat seller has asked send enough not enough items but
seller doesn t care other people say he says something else send s hard|0.0       |
+--------------------------------------------------------------------------------
--------------------------------------------------------------------+----------+
only showing top 10 rows
```

Filter only Positive Reviews

In [4]:
```
# Filter DataFrame to only contain rows where prediction = 1.0
filtered_df = df.filter(df.prediction == 2.0)
```

In [5]:
```
filtered_df.count()
```

Out[5]: 561

Produce Messages

In [ ]:
```
# kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1
```

In [6]:
```
sc = spark.sparkContext
sc.addFile("../../../de_classes/event_streaming/KafkaProducer.py")
```

In [7]:
```
from KafkaProducer import KafkaProducerClass

positiveProducer = KafkaProducerClass(bootstrap_servers='localhost:9092'
                                      , topic_name='positiveReviews')
```

In [8]:
```
# Assuming `df` is your DataFrame and it has a column named "Review"
for row in filtered_df.select("Review").collect():
    review = row["Review"]
    positiveProducer.produce_message({"Positive Review": review})
```

In [9]:
```
positiveProducer.close()
```

In [11]:
```
spark.stop()
```

In [ ]:
```

```