

Tweet Listener to Kafka topic

KAFKA TOPIC

Kafka Topic is Streamed with latest offset

LDA DATA LABELING

MULTICLASS CLASSIFICATION

KERAS COMPLAIN CLASSIFICATION

KERAS MODEL SENTIMENT ANALYSIS

SPARK LAYER - STREAMING JOB LIFECYCLE

SPARK STREAMING

ENTITY EXTRACTION

Enriched data sink to parquet file

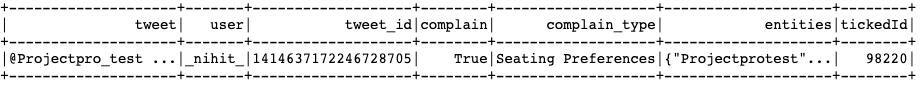
(S3, Hive, Hadoop, kafka)

API CALL TO

REPLY TWEET AND GENERATE TICKED ID

PARQUET

FILE



# AWS

**Instance 1**

1. Kafka Instance AWS
2. Hosting API Call
3. Tweet Listener

**Instance 2**

1. Spark Submit Job

**Steps:**

sudo apt update

sudo apt install default-jdk

sudo wget https://packages.confluent.io/archive/5.5/confluent-5.5.0-2.12.tar.gz

sudo tar xvzf confluent-5.5.0-2.12.tar.gz

sudo vi /etc/profile

export PATH=/home/ubuntu/confluent-5.5.0/bin:$PATH

source /etc/profile

echo $PATH

/home/ubuntu/confluent-5.5.0/bin/confluent-hub install --no-prompt confluentinc/kafka-connect-datagen:0.1.0

/home/ubuntu/kafka/kafka/confluent-5.5.0/bin/confluent start

confluent local start

Filezilla to copy code

Requirement.txt

**Spark Job Submit:**

sudo vi /etc/profile

export PATH=/home/ubuntu/.local/lib/python3.8/site-packages/pyspark/bin:$PATH