**MOONSHOT 2**

**Spark Enabled Real-time Streaming System**

**TEAM**: ZOOTOPIA

**TEAM MEMBERS**:

1. Sanjeedha Sanofer Raja (010698805)
2. Meera Mali
3. Ojas Milind Kale (010734672)
4. Diksha jain
5. Shagun Juneja (010742875)

Name of the App: **TweetOViz**

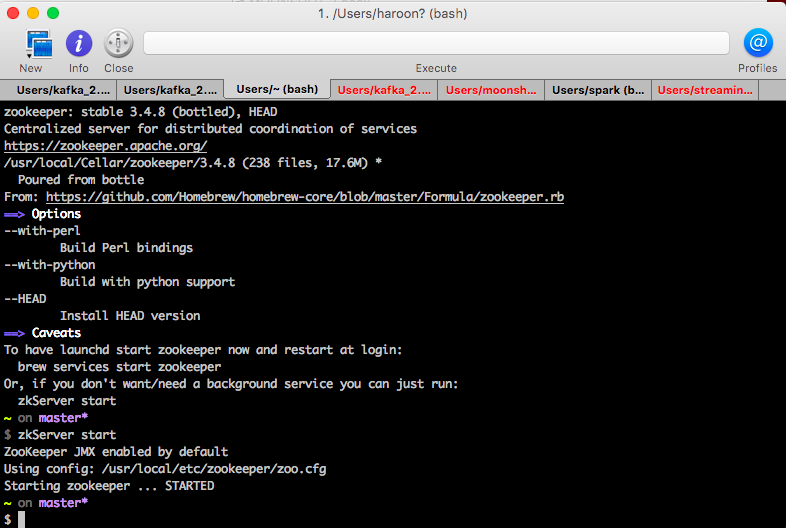
Real-time Streaming data: **Republicans vs Democrats**

**Steps:**

1) Create Twitter API account to get the API key and secret.

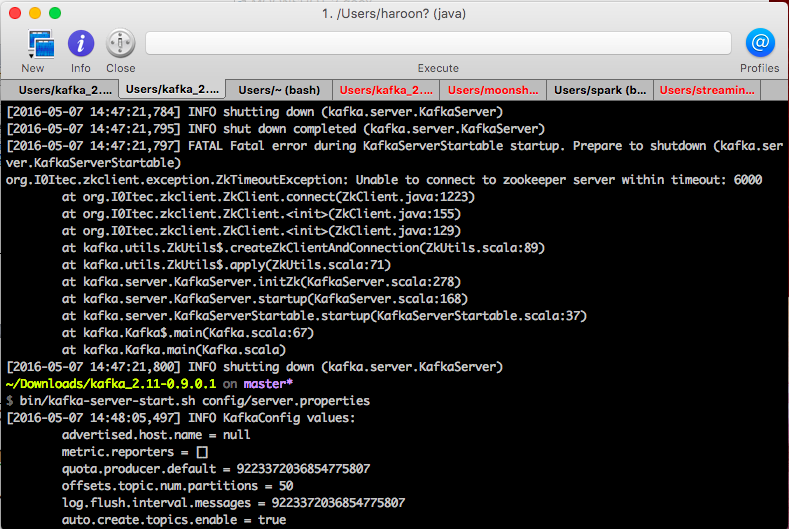
2) Start Zookeeper installed using brew

zkServer start



3) Start Kafka

bin/kafka-server-start.sh config/server.properties

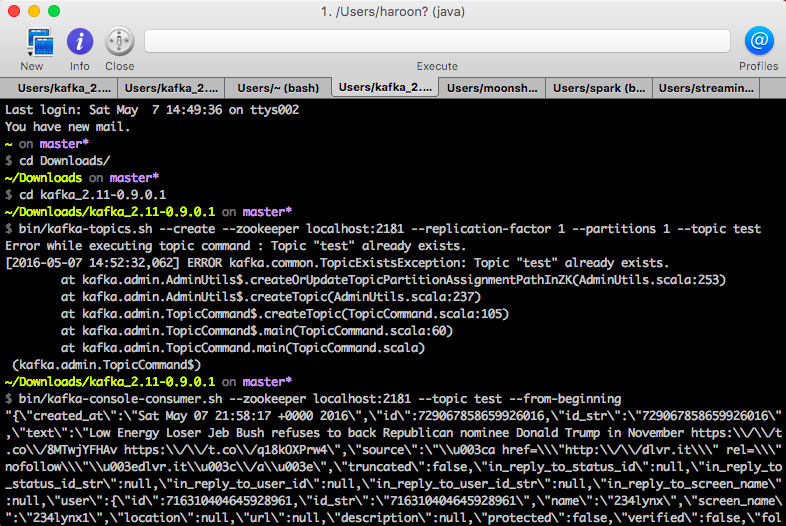


4) Create Kafka topic if necessary

bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic test

5) Create kafka Consumer

bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic test --from-beginning



6) Run the python script to use twitter API and get tweets and send it to kafka broker

python twitter\_streaming.py

**twitter\_streaming.py**

#Import the necessary methods from tweepy library

from tweepy.streaming import StreamListener

from tweepy import OAuthHandler

from tweepy import Stream

from kafka import KafkaProducer

import json

producer = KafkaProducer(value\_serializer=lambda v: json.dumps(v).encode('utf-8'))

#Variables that contains the user credentials to access Twitter API

access\_token = "4176278119-OkdJTZgtdJh6qZYz483HTdMjcDAVbJQL45CoSgZ"

access\_token\_secret = "Al6PWC6EXWNPUvvmP964kVDfZivlCyVEEdIUaCQYoecxx"

consumer\_key = "j19KmpBmWHDZVbORuUgXj3aMF"

consumer\_secret = "MrzG1HCftg9lgEu93QMnjtPfo4AxFb8YkzEXe4HvZZGbFcm4NO"

#This is a basic listener that just prints received tweets to stdout.

class StdOutListener(StreamListener):

def on\_data(self, data):

producer.send('test', data)

return True

def on\_error(self, status):

print status

if \_\_name\_\_ == '\_\_main\_\_':

#This handles Twitter authetification and the connection to Twitter Streaming API

l = StdOutListener()

auth = OAuthHandler(consumer\_key, consumer\_secret)

auth.set\_access\_token(access\_token, access\_token\_secret)

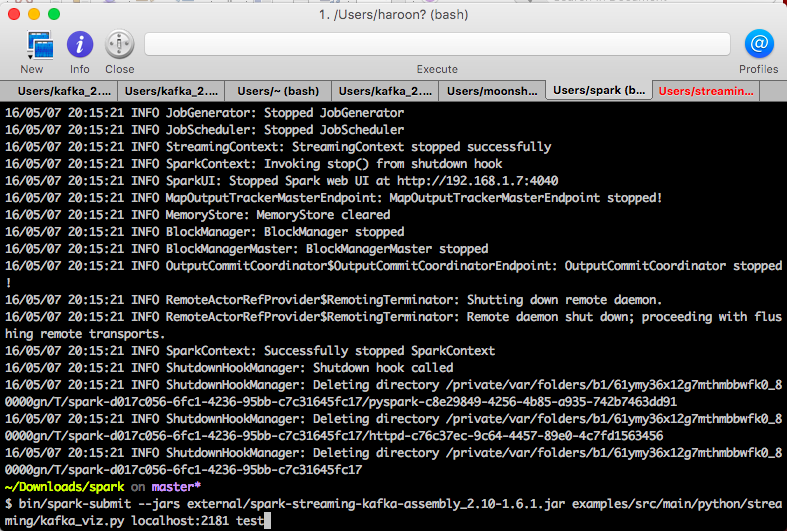
stream = Stream(auth, l)

#This line filter Twitter Streams to capture data by the keywords: 'python', 'javascript', 'ruby'

stream.filter(track=['democrat', 'republican'])

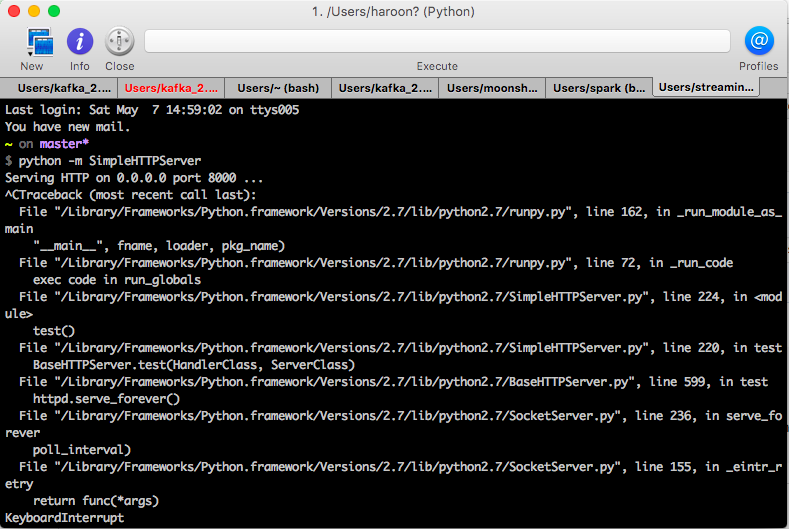
7) Run the spark kafka script to listen to kafka broker, receive the tweets from kafka, process the tweets and count the number of tweets for democrats and republicans. Save the information in a JSON file.

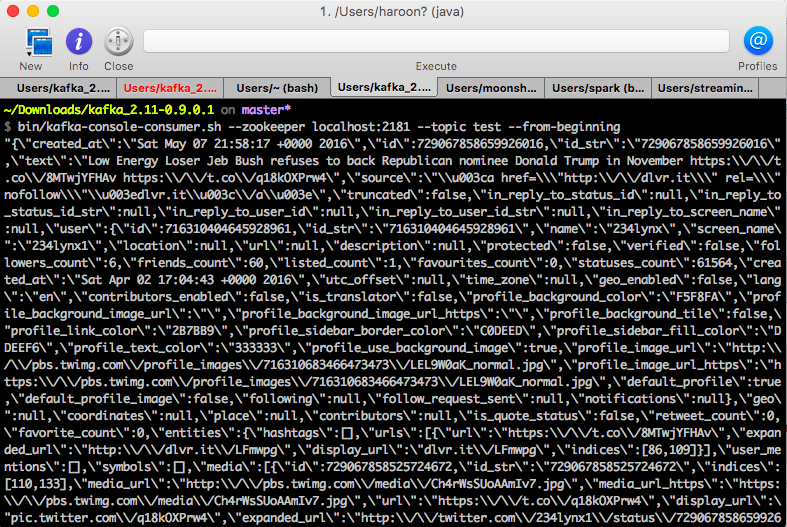
bin/spark-submit --jars external/spark-streaming-kafka-assembly\_2.10-1.6.1.jar examples/src/main/python/streaming/kafka\_viz.py localhost:2181 test



8) Use d3.js to read the JSON file in step 7 at regular intervals and plot the data in a bar chart and time series chart.

python -m SimpleHTTPServer





9) Run the URL for visualizing the flow.

<http://localhost:8000/streaming.html>

