

CS6350 Assignment 3 Part 1

(Spark Streaming with Twitter and Kafka)

Steps to run the TwitterStreaming application:

Step 1: Running Zookeeper Service

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
```

Step 2: Start Kafka Service

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.\bin\windows\kafka-server-start.bat .\config\server.properties
```

Step 3: Create kafka topic called as "tweets".

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.\bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic tweets
Created topic tweets.

C:\kafka>
```

Step 4: Running producer and consumer and testing the connection.

```
C:\kafka>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic tweets
>POSITIVE:RT @Test_tweet: Covid Vaccine works
>

C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic tweets --from-beginning...
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

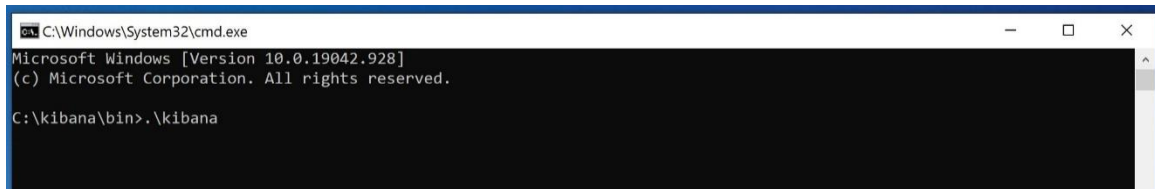
C:\kafka>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic tweets --from-beginning
POSITIVE:RT @Test_tweet: Covid Vaccine works
```

Step 5: Starting Elastic Search Service

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\elasticsearch\bin>.\elasticsearch
```

Step 6: Running Kibana Service



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

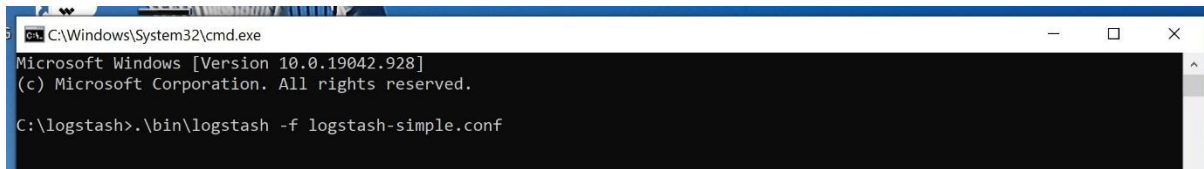
C:\kibana\bin>.\kibana
```

Step 7: Setting up Logstash config.



```
new 1 x logstash-simple.conf x
1 input {
2   kafka {
3     bootstrap_servers => "localhost:9092"
4     topics => ["tweets"]
5   }
6 }
7 filter {
8   grok {
9     match => { "message" => "%{WORD:Sentiment}:%{GREEDYDATA:Tweet}" }
10  }
11  grok {
12    match => { "message" => "(?<Keyword>(?!)(Covid|POTUS|Vaccine|Bitcoin|India))" }
13  }
14  if "_grokparsefailure" in [tags] {
15    drop {}
16  }
17 }
18 output {
19   elasticsearch {
20     hosts => ["localhost:9200"]
21     index => "tweets-index"
22   }
23 }
24 }
```

Step 8: Running Logstash Service.

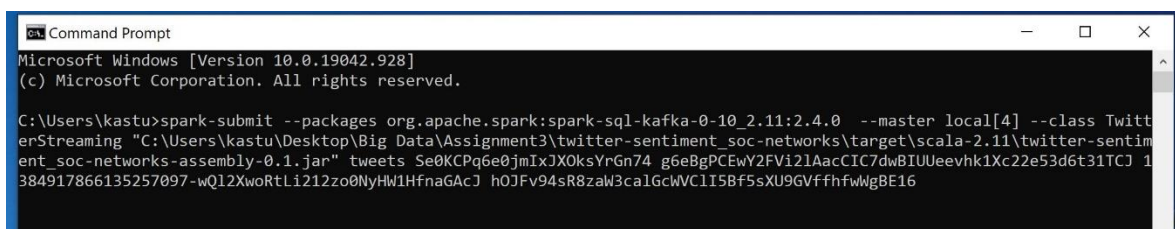


```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\logstash>.\bin\logstash -f logstash-simple.conf
```

Step 9: Running the fat JAR.

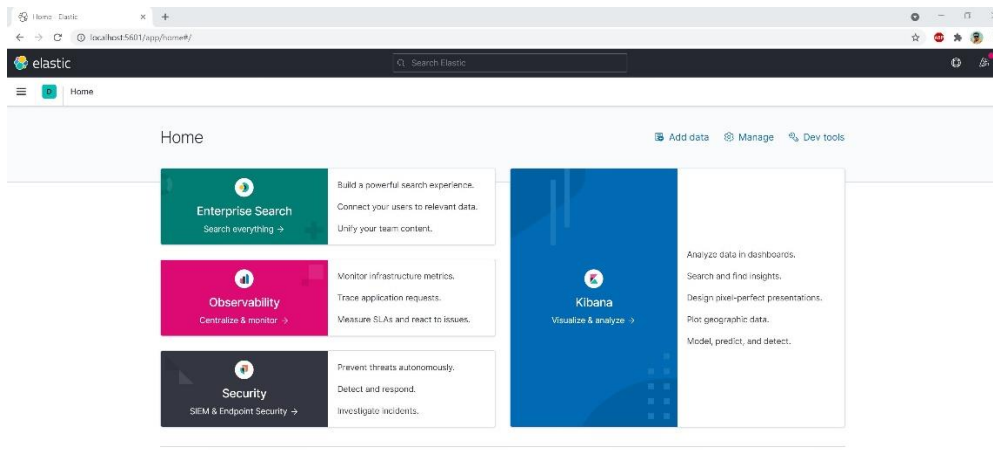
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.0 --master local[4] --class TwitterStreaming "C:\Users\kastu\Desktop\Big Data\Assignment3\twitter-sentiment_soc-networks\target\scala-2.11\twitter-sentiment_soc-networks-assembly-0.1.jar" tweets Se0KCPq6e0jmIxJX0ksYrGn74 g6BgPCEwY2FVi21AacCIC7dwBIUueevhk1Xc22e53d6t31TCJ 1384917866135257097-wQ12XwoRtLi212zo0NyHW1HfnaGAcJ h0JFv94sR8zaW3ca1GcWVC1I5Bf5sXU9GVffhfwWgBE16



```
Command Prompt
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\Users\kastu>spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.0 --master local[4] --class TwitterStreaming "C:\Users\kastu\Desktop\Big Data\Assignment3\twitter-sentiment_soc-networks\target\scala-2.11\twitter-sentiment_soc-networks-assembly-0.1.jar" tweets Se0KCPq6e0jmIxJX0ksYrGn74 g6BgPCEwY2FVi21AacCIC7dwBIUueevhk1Xc22e53d6t31TCJ 1384917866135257097-wQ12XwoRtLi212zo0NyHW1HfnaGAcJ h0JFv94sR8zaW3ca1GcWVC1I5Bf5sXU9GVffhfwWgBE16
```

Step 10: Kibana in (<http://localhost:5601>).



Step 11: Creating index pattern.

Create index pattern

An index pattern can match a single source, for example, `filebeat-4-3-22`, or **multiple** data sources, `filebeat-*`.
[Read documentation](#)

Step 1 of 2: Define an index pattern

Index pattern name

tweets-index*

Use an asterisk (*) to match multiple indices. Spaces and the characters `\,/,?,"',<,>,|` are not allowed.

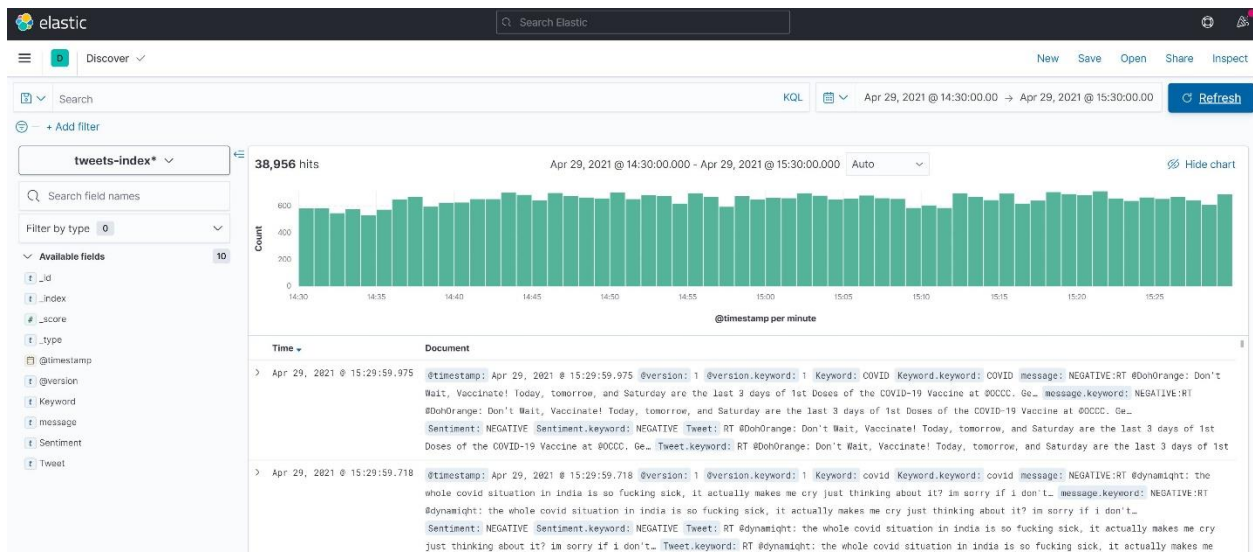
☐ Include system and hidden indices

✓ Your index pattern matches 1 source.

tweets-index Index

Rows per page: 10

Step 11: In discover tab we can see the data.



Final Output:

Created Dashboard with various visualizations for the twitter data with keywords Covid, India, POTUS, Vaccine and Bitcoin in the duration of

04/29/2021 14:30:00 to 04/29/2021 15:30:00

