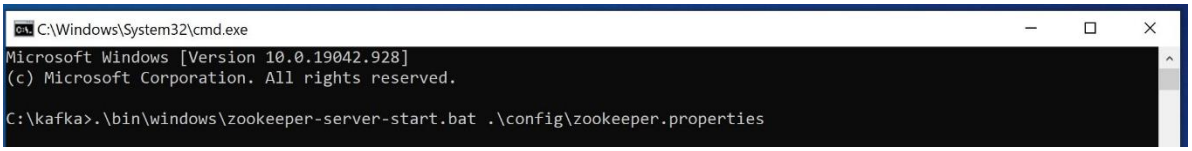
CS6350 Assignment 3 Part 1

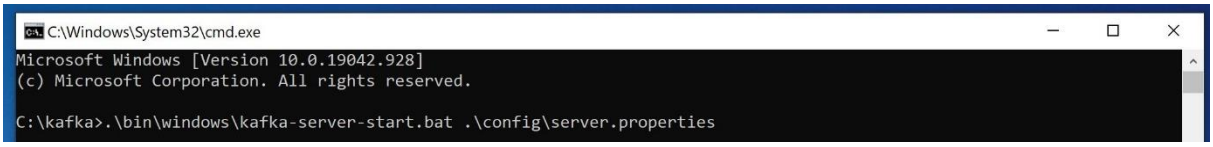
(Spark Streaming with Twitter and Kafka)

Steps to run the TwitterStreaming application:

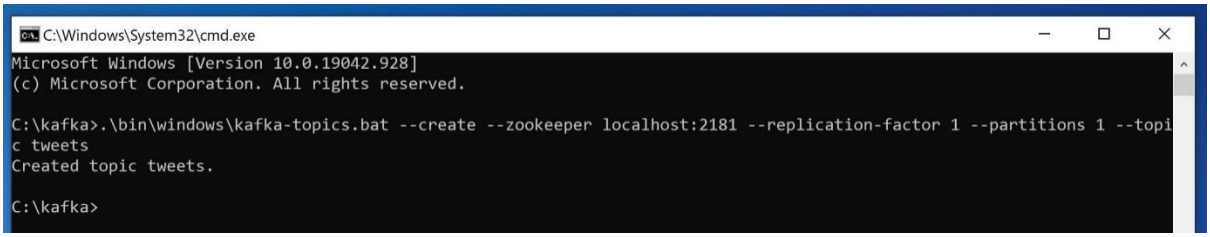
Step 1: Running Zookeeper Service



Step 2: Start Kafka Service



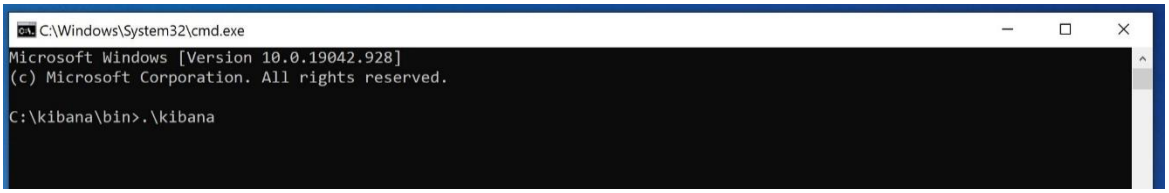
Step 3: Create kafka topic called as “tweets”.



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

Step 5: Starting Elastic Search Service

Step 6: Running Kibana Service

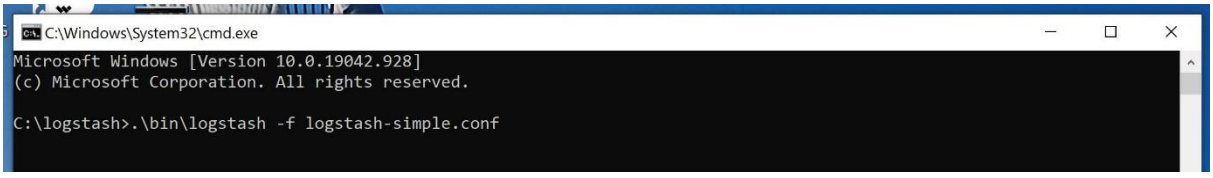


Step 7: Setting up Logstash config**.**

new 1 | logstash-simple.conf

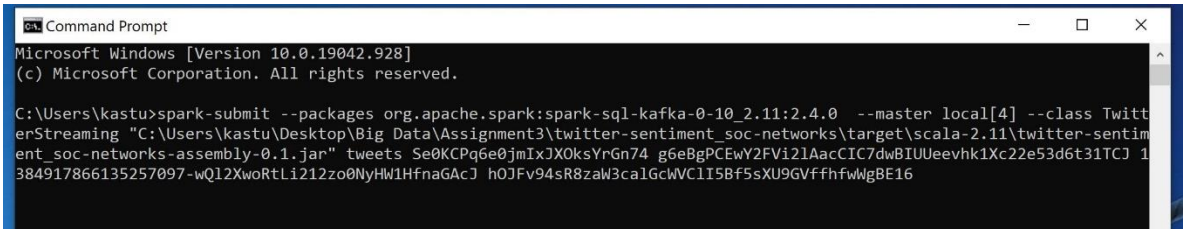
```
input {
    kafka {
        bootstrap_servers => "localhost:9092"
        topics => ["tweets"]
    }
}
filter {
    grok {
        match => { "message" => "%{WORD:Sentiment}:%{GREEDYDATA:Tweet}" }
    }
    grok {
        match => { "message" => "(?<Keyword>(?i)(Covid|POTUS|Vaccine|Bitcoin|India))" }
    }
    if "_grokparsefailure" in [tags] {
    drop {}
    }
}
output {
    elasticsearch {
        hosts => ["localhost:9200"]
        index => "tweets-index"
    }
}

```

Step 8: Running Logstash Service**.**

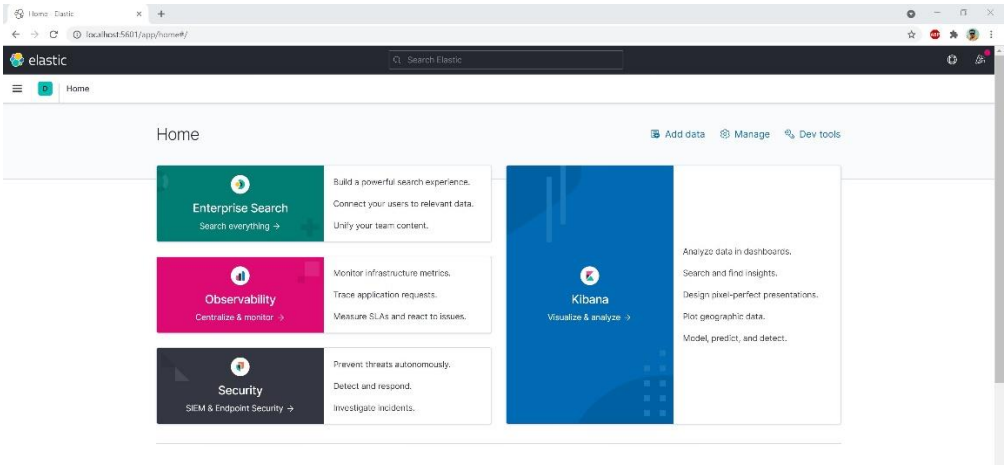


Step 9: Running the fat JAR**.**

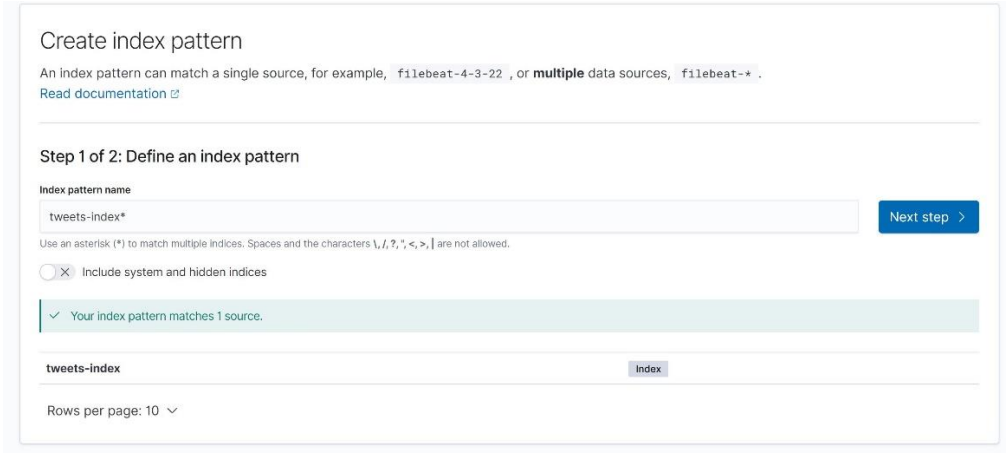
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.0 --master local[4] --class <classname> <path-of-jar-location> <topicname> <API-key> <API-secret> <Access-token> <Access-token-secret>



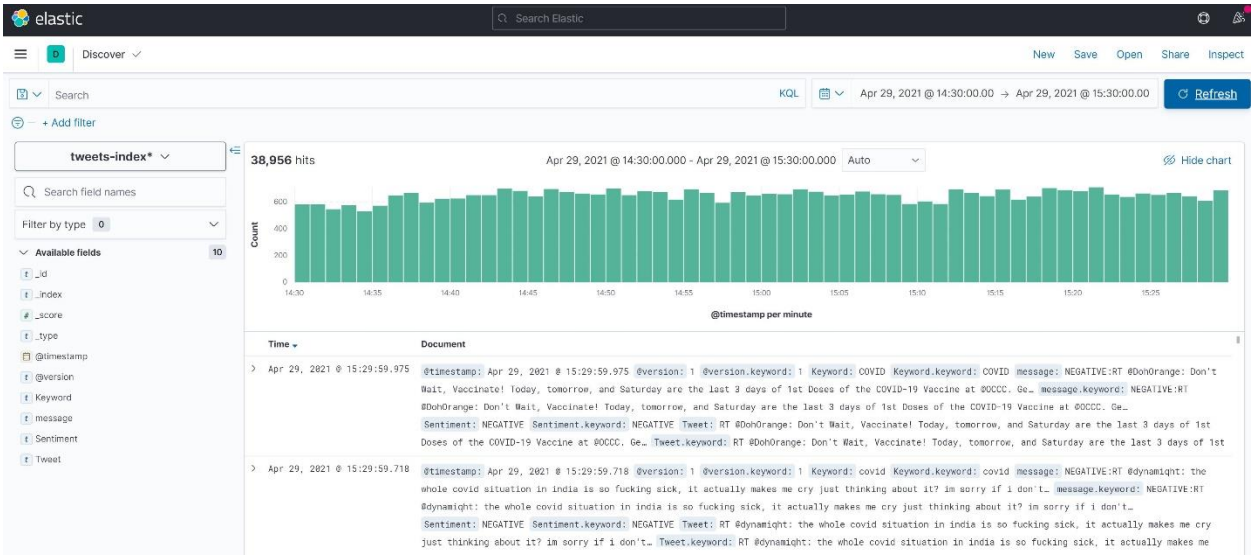
Step 10: Kibana in (http://localhost:5601).



Step 11: Creating index pattern.



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