# Spark Streaming with Twitter and Kafka

## Running Environment

Spark Version: 2.4.5

Java Version: 1.8.0_241

Scala Version: 2.11.12

Kafka Version: 2.4.1

## Instruction

1. Create a fat jar file:
   In the Intellij, go to sbt task, double click on the "assembly" to create a fat jar file.
2. Start zookeeper server:
   Go to the kafka directory, run the command in terminal:

```
bin/zookeeper-server-start.sh config/zookeeper.properties
```

3. Start kafka server:
   Go to the kafka directory, run the command in new terminal tab:

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

4. Start elasticSearch:
   Go to the elasticSearch directory, run the command in new terminal tab:

```
bin/elasticSearch
```

5. Start kibana:
   Go to the kibana directory, run the command in new terminal tab:

```
bin/kibana
```

6. Create config file for logstash:
   Go to the logstash directory, create a config file with name "logstash-simple.conf" and the

following content (Analyze the tweets about test):

```
  input {
  kafka {
  bootstrap_servers => "localhost:9092"
  topics => ["test"]
  }
  }
  output {
  elasticsearch {
  hosts => ["localhost:9200"]
  index => "test-index"
  }
  }
```

7. Run logstash:
   Go to the logstash directory, run the command in new terminal tab:

```
  bin/logstash -f logstash-simple.conf
```

8. Run fat jar file:
   Go to the fat jar file directory, run the command in new terminal tab:

```
  spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.0 --
class TwitterSentiment TwitterSentiment-assembly-0.1.jar  test
```

9. Visualize the result:
   Go to http://localhost:5601 to visualize the results.