

TDSA (Twitter Distributed Sentiment Analysis)

How to run

1. Prepare environment variables for JAVA, SPARK, KAFKA, CASSANDRA and PYTHON.
2. Make sure to have zookeeper and kafka server up and running using commands such as these:

```
$KAFKA_HOME/bin/zookeeper-server-start.sh $KAFKA_HOME/config/zookeeper.properties  
> zookeeper.log $KAFKA_HOME/bin/kafka-server-start.sh  
$KAFKA_HOME/config/server.properties > kafka.log
```
3. Make sure to have Cassandra up and running: `$CASSANDRA_HOME/bin/cassandra -f`
4. execute the twitter stream by running: `python3 twitter-stream.py`
5. start the spark workers by running: `python3 pyspark-stream.py`
6. When you are happy with the data collected you can fetch and plot the results using: `python3 analyse_data.py`

It will produce an HTML file with the plot, it should be openable using any browser, we tested on Chrome and Firefox.