[BD] Assignment 3-1 README

1. launch your kafka environment using the usual commands:
In one terminal window
bin/zookeeper-server-start.sh config/zookeeper.properties

In another:
bin/kafka-server-start.sh config/server.properties

2. Then create two topics. Call them test and finaltopic.
bin/kafka-topics.sh --create --topic test --bootstrap-server localhost:9092

bin/kafka-topics.sh --create --topic finaltopic --bootstrap-server localhost:9092

3. Create a directory named checkpoint at the same level as the programs you are running. I have already provided one, but in case running hw3-2 throws errors, create an empty directory named checkpoint and rerun the program twice.

4. Launch your ELK.
Logstash should be configured with the logstash.conf file given in the submission.

5. Then, run homework3-2 using:
sh <path_to_spark_submit> --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.1 <path_to_homework3-2.py> <bootstrap-servers> <checkpoint-dir> <input-topic> <output-topic>

for example for us:
sh /Users/Hema/Library/Python/3.8/bin/spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.1 /Users/Hema/bdhw3/partone/homework3-2.py localhost:9092 checkpoint test finaltopic

Similarly, run homework3_2_tweet
sh <path_to_spark_submit> --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.1 <path_to_homework3_2_tweet.py> <bootstrap-servers> <query> <topics>

For us it was:
sh /Users/Hema/Library/Python/3.8/bin/spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.1 /Users/Hema/bdhw3/partone/homework3_2_tweet.py localhost:9092 covid test

6. Go to Kibana, create an index pattern (hw3_final_index is what you need to match to) and visualize in the dashboard.