Kafka install

- Transfer the following file to the ***/home/hadoop/lab/downloads*** directory using winscp or filezilla. Then untar the file using the following command.

cd /home/hadoop/lab/software

tar -xvf /home/hadoop/lab/downloads/kafka_2.11-2.0.0.tgz

Start Kafka Services

- Change directory to kafka install directory

cd /home/hadoop/lab/software/kafka_2.11-2.0.0

- Start zookeeper server

nohup bin/zookeeper-server-start.sh config/zookeeper.properties &

- Start kafka server

nohup bin/kafka-server-start.sh config/server.properties &

- Verify if the kafka services are running

Jps

It should list the processes: *kafka* and *QuorumPeerMain.*

```
[hadoop@hadooplab sbin]$ jps
10865 SecondaryNameNode
7233 Kafka
13506 ResourceManager
13620 NodeManager
10549 NameNode
6934 QuorumPeerMain
15574 SparkSubmit
14472 Jps
10687 DataNode
```

Create Kafka Topic

- Create a topic called tweets_topic

bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic tweets_topic --config retention.ms=60000

- List topics

bin/kafka-topics.sh --list --zookeeper localhost:2181

Write the spark-kafka program

- Use the Spark kafka Tutorial.pdf for writing the program

Note: Just write the program. But do not run the program yet.

- Write to kafka topic from the command line

cat /home/hadoop/lab/data/tweets | while read x; do echo "$x"; sleep 0.2; done | /home/hadoop/lab/software/kafka_2.11-2.0.0/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic tweets_topic

- Run the jupyter notebook program

Note: Use *Cell -> Run All* option on the jupyter notebook.