|  |
| --- |
| Course: Big Data Technology  Final Project  Date: Sep 27th, 2020  Student ID: 987034 |

[Step 1: Create Hive table 2](#_Toc52136073)

[Step 2: Create partitions for Hive table 3](#_Toc52136074)

[Commands 3](#_Toc52136075)

[Note: 3](#_Toc52136076)

[Step 3: Create Kafka topic 4](#_Toc52136077)

[Step 4: Run SparkStreaming App 5](#_Toc52136078)

[Commands 5](#_Toc52136079)

[Notes: 5](#_Toc52136080)

[Step 5: Run Kafka Producer App 6](#_Toc52136081)

[Commands 6](#_Toc52136082)

[Notes 6](#_Toc52136083)

# Step 1: Create Hive table

|  |
| --- |
| --drop table StockDaily;  CREATE EXTERNAL TABLE StockDaily  (  Id STRING,  companyName STRING,  primaryExchange STRING,  openPrice FLOAT,  openTime BIGINT,  closePrice FLOAT,  closeTime BIGINT,  highPrice FLOAT,  highTime BIGINT,  lowPrice FLOAT,  lowTime BIGINT,  latestPrice FLOAT,  latestUpdate BIGINT,  receiveDate STRING,  latestVolume FLOAT,  previousClosePrice FLOAT,  previousVolume FLOAT,  change FLOAT,  changePercent FLOAT,  volume FLOAT,  avgTotalVolume FLOAT,  marketCap FLOAT,  peRatio FLOAT,  week52High FLOAT,  week52Low FLOAT,  ytdChange FLOAT,  lastTradeTime BIGINT  )  PARTITIONED BY (symbol STRING)  ROW FORMAT DELIMITED  FIELDS TERMINATED BY ','  LINES TERMINATED BY '\n'  LOCATION '/user/hive/BDT/StockDaily'; |

# Step 2: Create partitions for Hive table

## Commands

|  |
| --- |
| LOAD DATA INPATH  '/user/cloudera/BDT/FinalProject/Stock/AAPLSampleData.txt'  INTO TABLE stockdaily  PARTITION (symbol='AAPL');  LOAD DATA INPATH  '/user/cloudera/BDT/FinalProject/Stock/GOOGLSampleData.txt'  INTO TABLE stockdaily  PARTITION (symbol='GOOGL');  LOAD DATA INPATH  '/user/cloudera/BDT/FinalProject/Stock/MSFTSampleData.txt'  INTO TABLE stockdaily  PARTITION (symbol='MSFT'); |

## Note:

Before load & create partitions for the StockDaily table, please upload 3 sample output files into the folder below in the HDFS:

**/user/cloudera/BDT/FinalProject/Stock**/AAPLSampleData.txt'

'/user/cloudera/BDT/FinalProject/Stock/GOOGLSampleData.txt'

'/user/cloudera/BDT/FinalProject/Stock/MSFTSampleData.txt'

# Step 3: Create Kafka topic

|  |
| --- |
| kafka-topics --zookeeper quickstart.cloudera:2181 --create --topic AAPLTest --partitions 1 --replication-factor 1  kafka-topics --zookeeper quickstart.cloudera:2181 --create --topic GOOGLTest --partitions 1 --replication-factor 1  kafka-topics --zookeeper quickstart.cloudera:2181 --create --topic MSFTTest --partitions 1 --replication-factor 1 |

# Step 4: Run SparkStreaming App

## Commands

|  |
| --- |
| java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/stockSparkStreaming.jar' AAPLTest AAPL 20 /user/hive/BDT/StockDaily    java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/stockSparkStreaming.jar' GOOGLTest GOOGL 20 /user/hive/BDT/StockDaily  java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/stockSparkStreaming.jar' MSFTTest MSFT 20 /user/hive/BDT/StockDaily |

## Notes:

* + **Arguments**:

| **No** | **Arg** | **Description** | **Sample value** |
| --- | --- | --- | --- |
| 1 | Topic Name | Kafka topic name | AAPLTest |
| 2 | Symbol | Stock ticker/symbol | AAPL: Apple Inc, GOOGL: Google Inc |
| 3 | Duration | Timer Duration (seconds) | 10 |
| 4 | hdfs path | hdfs folder related to hive table | /user/hive/BDT/StockDaily |

# Step 5: Run Kafka Producer App

## Commands

|  |
| --- |
| java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/producer.jar' AAPLTest AAPL 10    java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/producer.jar' GOOGLTest GOOGL 10  java -jar '/home/cloudera/workspace/FinalProject/RunnableJarFiles/producer.jar' MSFTTest MSFT 10 |

## Notes

* + **Arguments**:

| **No** | **Arg** | **Description** | **Sample value** |
| --- | --- | --- | --- |
| 1 | Topic Name | Kafka topic name | AAPLTest |
| 2 | Symbol | Stock ticker/symbol | AAPL: Apple Inc, GOOGL: Google Inc |
| 3 | Duration | Timer Duration (seconds) | 5 |
| 4 | token | IEX API token | Bao's token |
| 5 | apiURL | IEX API URL | <https://sandbox.iexapis.com/stable/stock> |