**Input datasets:**

(base) anjaliv@anjali-pc:~/Desktop/Fall\_2021/BigData/Assignments/assignment9/datasets$ wc -l \*

5061 anthem.csv

6781 apple.csv

4356 google.csv

6781 humana.csv

6781 medtronic.csv

6781 pfizer.csv

2881 tesla.csv

6781 unitedHealth.csv

46203 total

(base) anjaliv@anjali-pc:~/Desktop/Fall\_2021/BigData/Assignments/assignment9/datasets$

**Create keyspace and table in cql:**

CREATE KEYSPACE BigDataProject WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication\_factor' : 1 };

describe keyspaces;

use bigdataproject;

CREATE TABLE StocksInformation(

CompanyName text,

Date date,

Open double,

High double,

Low double,

Close double,

AdjClose double,

Volume bigint,

PRIMARY KEY (CompanyName, Date)

);

**Start kafka as below:**

In kafka directory:

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

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

**Run consumer as below:**

python cassandraConsumer.py stockInformation 127.0.0.1:9092 bigdataproject stocksinformation 127.0.0.1

**Run producer.py for reading from yahoo finance api as below:**

Python producer.py

**Run producerCSV.py for reading from CSV data as below:**

python producerCSV.py

**Output:**

Total count in table using CSVs should be: 46195

Total count in table using direct link should be: 29314