**NSE Stock Exchange Analysis Using Bigdata**

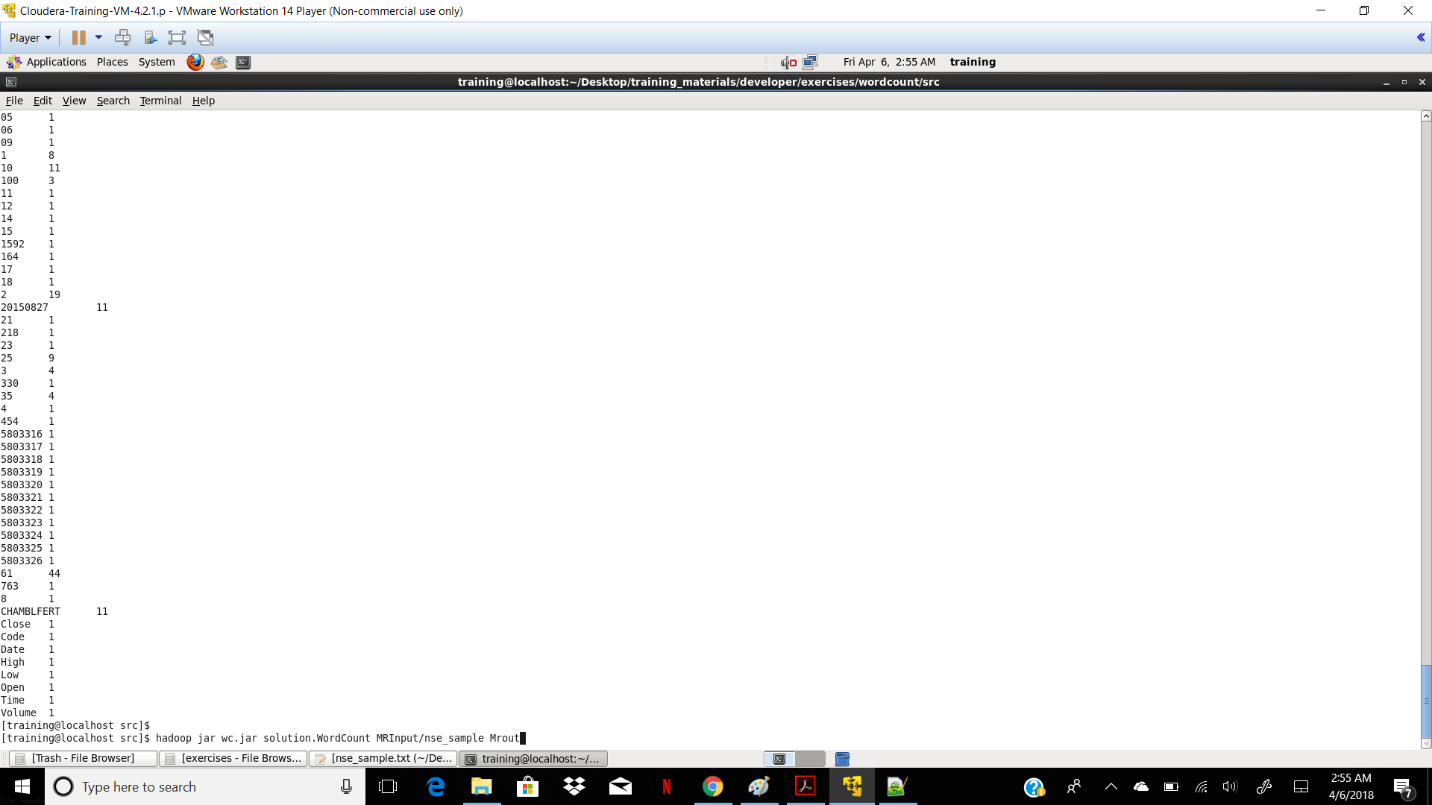
**Name: Aditya Shethiya**

**StudentId:1034791**

**Four parts Of the Project:**

**Part1: Map Reducing**

In the Map reducing program we use key on company Name and value is number of times of its occurrence so for that we take reference wordcount in of Map Reducing example and its source file is given in the folder and its results screen shot is given below:



After calculating the total number of record for specific company we go for the second part i.e Hive/pig

**Part 2: Analysis using Hive :**

In Hive we calculated the folling terms

1.Return

2.Persent return

3.Adjusted stock value 1

4.Adjusted stock value 2

5.Total number of share that owned by company in daily transection

6. Total number of share that owned by remining share holder

7.Avarage stock volume for daily transection

8.dailly turn over ratio

9.calculate the variance

10. calculated the deviance

Queries:

1.Create new folder on HDFS

hadoop fs -mkdir nse

2.loded on files on HDFS file system

hadoop fs -put /home/training/Desktop/nse-india-stocks-companies/\* /user/training/nsetest

3.Create New table in Hive and load data in that table:

create table nse\_stock\_data (ID INT, COMP\_NAME STRING, STOCK\_DATE TIMESTAMP, TIME TIMESTAMP, OPEN INT, HIGH INT, LOW INT, CLOSE INT, VOLUME INT) COMMENT 'NSE Data' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/training/nse/';

4.Genric Query for Our Project:

select

sum(close)/count(close) as avarageReturn,

sum(close)-sum(open)/sum(open)\*100 as persentreturn,

(sum(close)-sum(open))/sum(open) as return,sum(close)-(count(close)\*5) as Adj1,

sum(close)\*0.5 as Adj2,

sum(volume)/count(volume)\*0.5 as companyowned,

sum(volume)/count(volume)\*0.49 as FreeForHolder,

((sum(open)/count(open))+(sum(close)/count(close)))\*0.5 as AvgStoVal ,

(sum(high)/count(high))/((sum(open)/count(open))+(sum(close)/count(close)))\*0.5 as TurnOverRatio ,

variance(close) as variance,

Stddev(close) as StandredDevation

from Table where comp\_name='"CompanyName"';

5.Query after using Map Reducing Function In Hive:

select sum(close)/245 as avarageReturn,

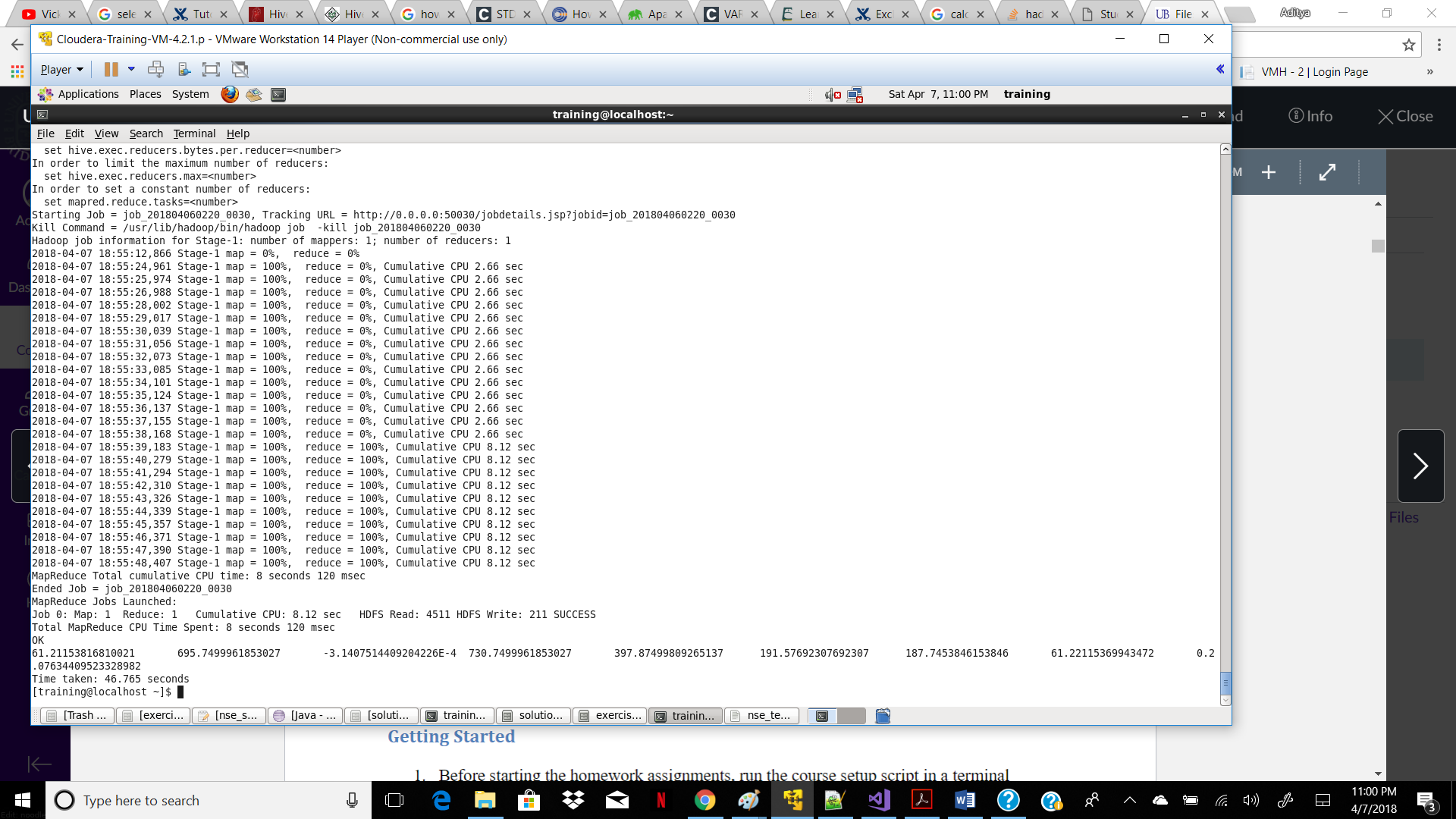
sum(close)-sum(open)/sum(open)\*100 as persentreturn,(sum(close)-sum(open))/sum(open) as return,sum(close)-( 245 \*5) as Adj1,sum(close)\*0.5 as Adj2,sum(volume)/ 245 \*0.5 as companyowned, sum(volume)/ 245 \*0.49 as FreeForHolder, ((sum(open)/ 245)+(sum(close)/ 245))\*0.5 as AvgStoVal ,

(sum(high)/ 245)/((sum(open)/ 245)+(sum(close)/ 245))\*0.5 as TurnOverRatio ,

variance(close) as variance, Stddev(close) as StandredDevation

from test2 where comp\_name='"CHAMBLFERT"';

Output of Hive:



6. for All company Data :

select

COMP\_NAME,

sum(close)/count(close) as avarageReturn,

sum(close)-sum(open)/sum(open)\*100 as persentreturn,

(sum(close)-sum(open))/sum(open) as return,sum(close)-(count(close)\*5) as Adj1,

sum(close)\*0.5 as Adj2,

sum(volume)/count(volume)\*0.5 as companyowned,

sum(volume)/count(volume)\*0.49 as FreeForHolder,

((sum(open)/count(open))+(sum(close)/count(close)))\*0.5 as AvgStoVal ,

(sum(high)/count(high))/((sum(open)/count(open))+(sum(close)/count(close)))\*0.5 as TurnOverRatio ,

variance(close) as variance,

Stddev(close) as StandredDevation

from test2 GROUP BY COMP\_NAME ;

Pig Query:

pig command:

A = LOAD '/home/training/Desktop/nse\_test\_data.csv' USING PigStorage(',');

B = FOREACH A GENERATE $0 AS ID,$1 AS COMPANY\_NAME,$2 AS DATE,$3 AS TIME,$4 AS OPEN, $5 AS HIGH,$6 AS LOW,$7 AS CLOSE,$8 AS VOLUME,((($7-$4)/$4)\*100) AS CHANGE,(((($7-$4)/$4)\*100)>0?'G':'L') AS GAINLOSS;

C = GROUP B BY COMPANY\_NAME;

D = FOREACH C GENERATE B.COMPANY\_NAME,AVG(B.CHANGE);

STORE D INTO '/home/training/pig/OUTPUT' USING PigStorage();