**ASSIGNMENT 3**

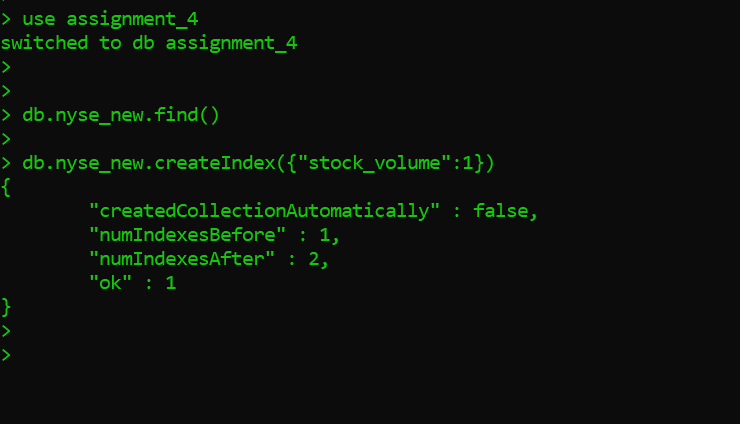
**Q2**

**ANS-**

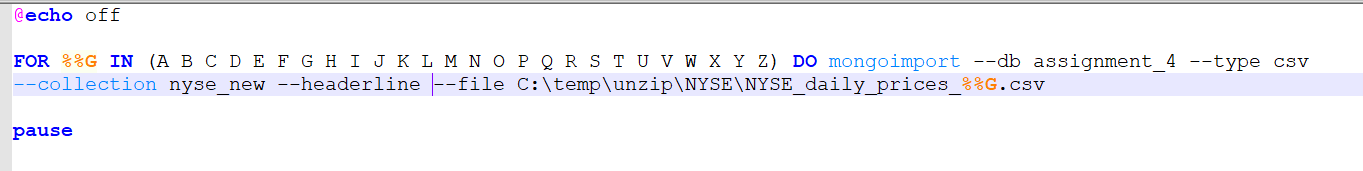
Created new DB- Assignment\_4

Then created a collections nyse\_new

Created index on key: stock\_volume in ascending order



**Then imported the nyse data to this collection using bat file**



**The indexes are correctly created.**



**Q3**

**Ans—**

**Imported all the nyse data to new collection stocks**

**Then after that created index** on key: stock\_symbol in ascending order



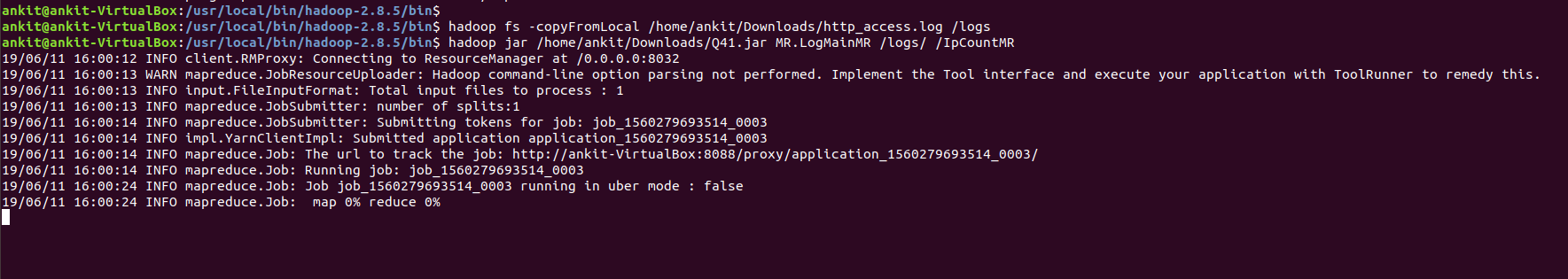
**Q4**

**Ans:**

**Q5**

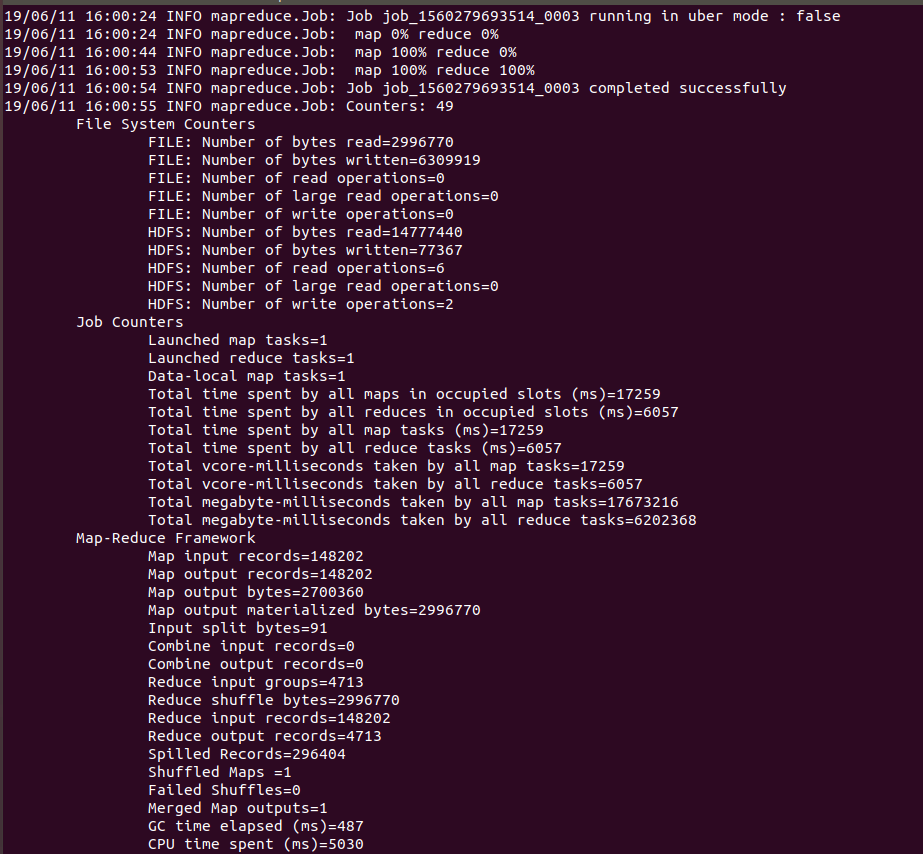
**Ans--**

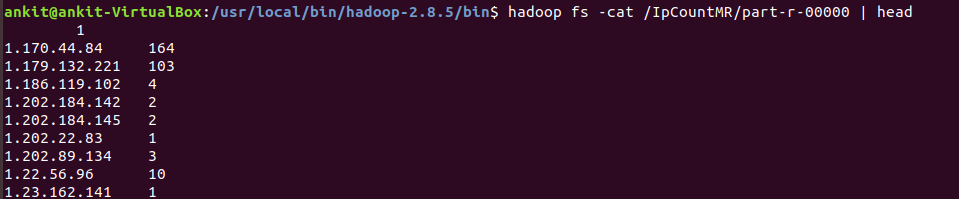
**copy http\_access.log to /logs**

****

Also ran MAP reduce on the logs to count the number of time ip is used

The program folder is inside the Assignment folder

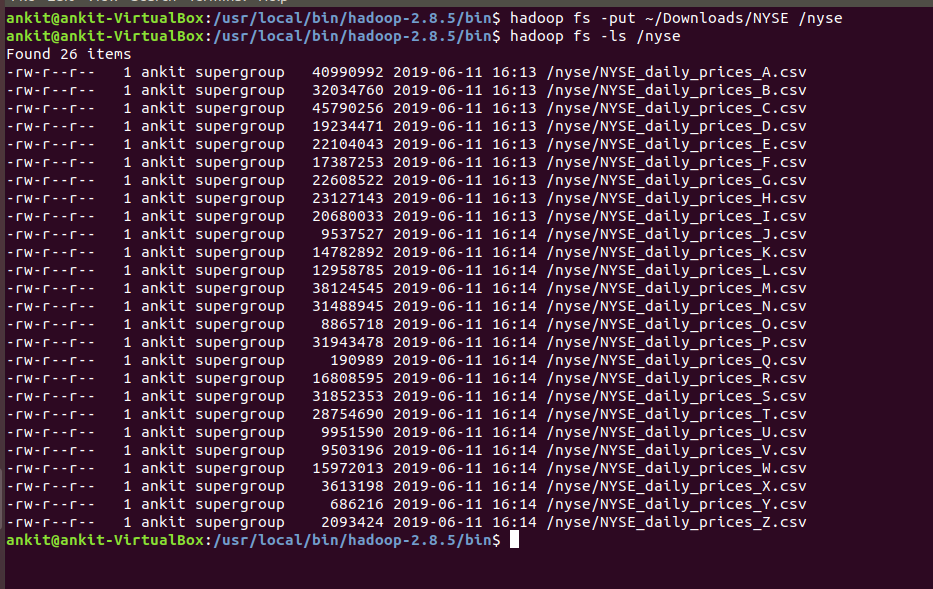
Output head is as follows



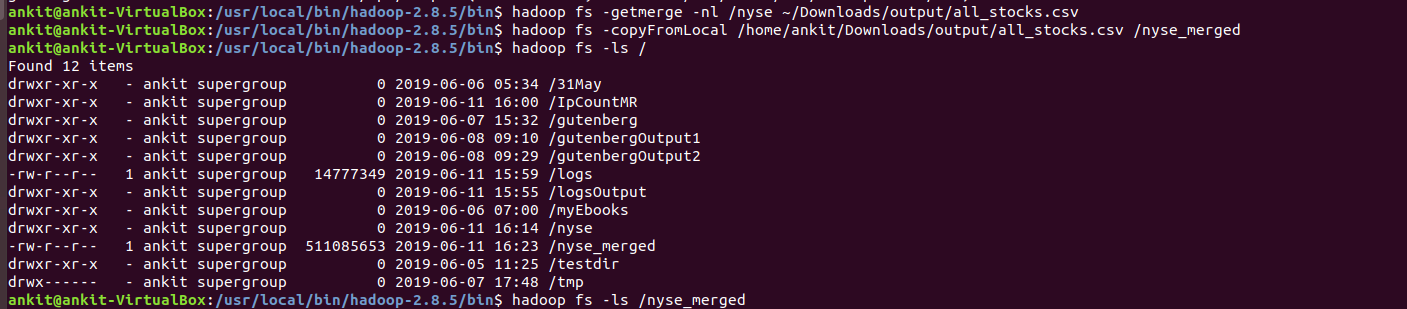
Q6

Ans--

Download and Copy all the files (http://msis.neu.edu/nyse/) (DailyPrices\_A to DailyPrices\_Z) to a folder in HDFS.

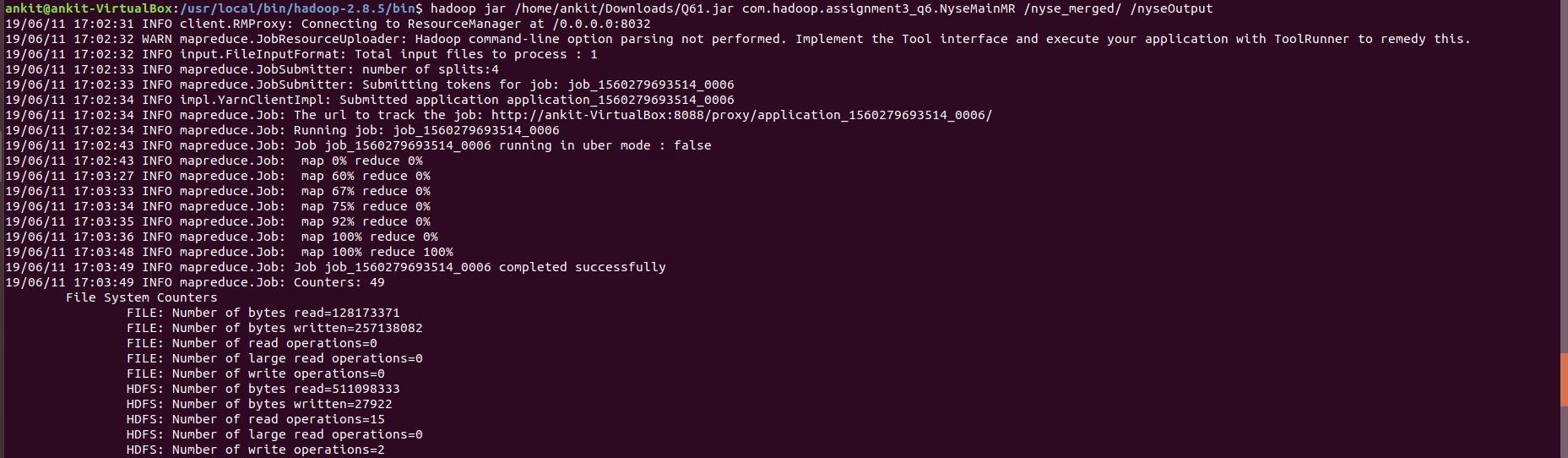


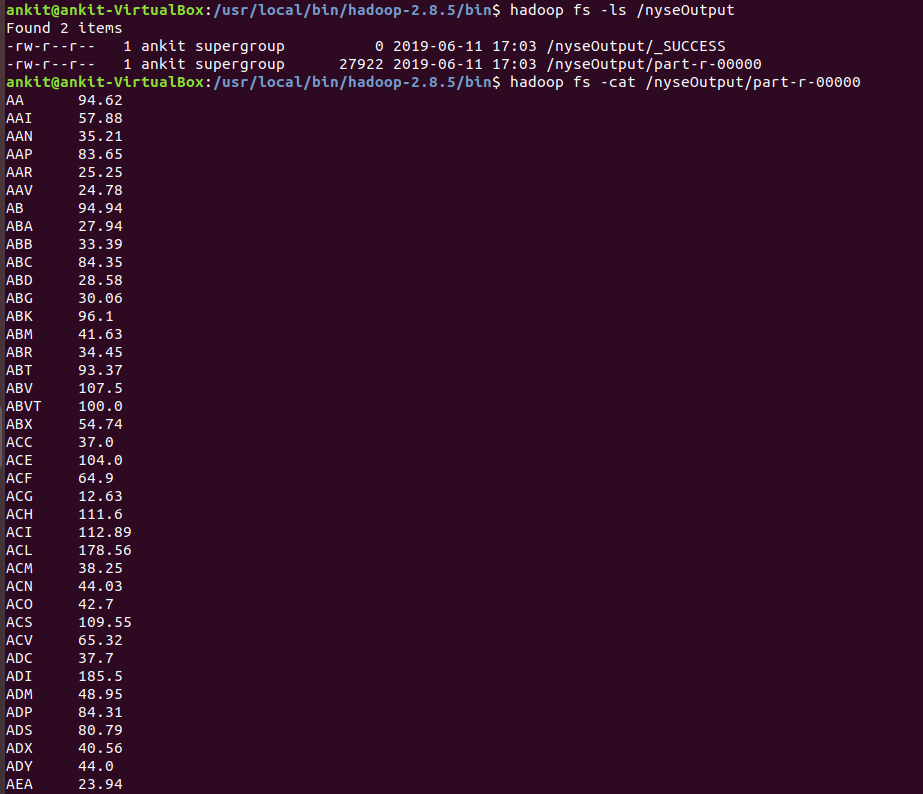
Merging all records in single file and saving in HDFS



The java program is in the assignment folder

Output of MapReduce for maximum stock\_price\_high for each stock is as follows:





Q7

Ans--

CombineFileInputFormat

FixedLengthInputFormat

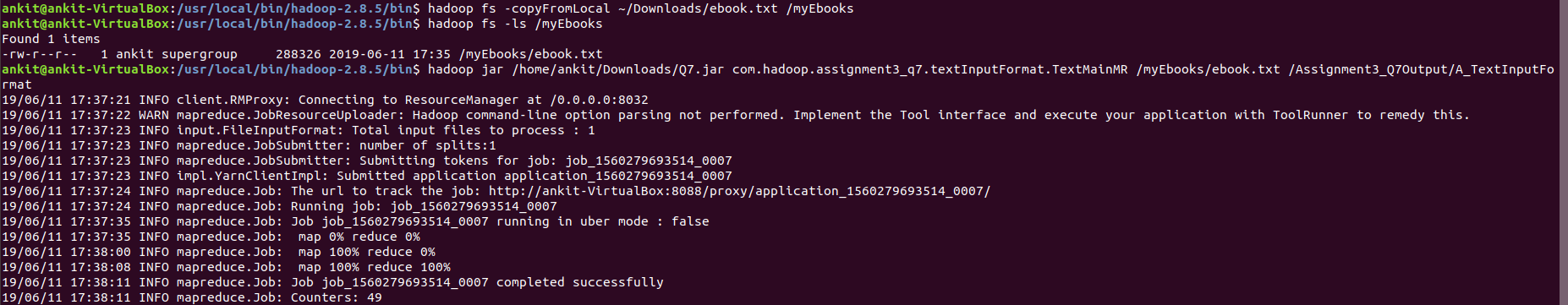
KeyValueTextInputFormat

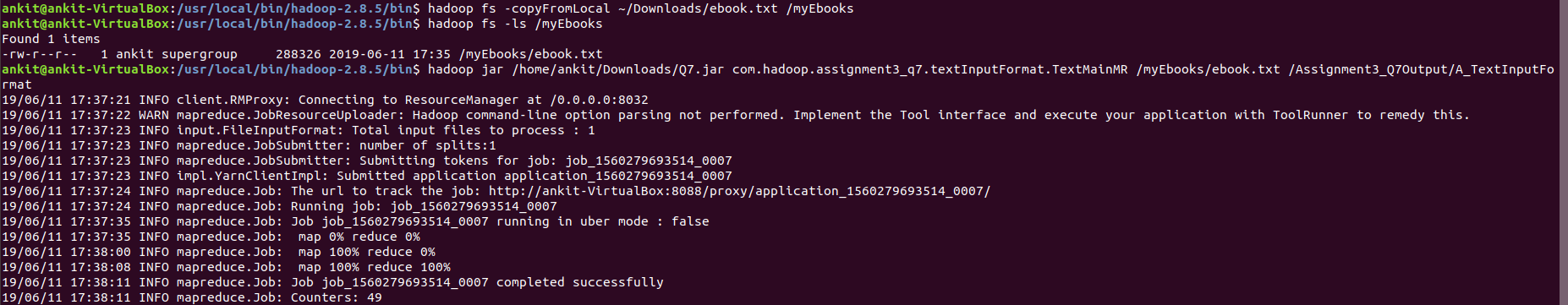
NLineInputFormat

**1> TextInputFormat**

Code in package textInputFormat

Output





**2> SequenceFileInputFormat**