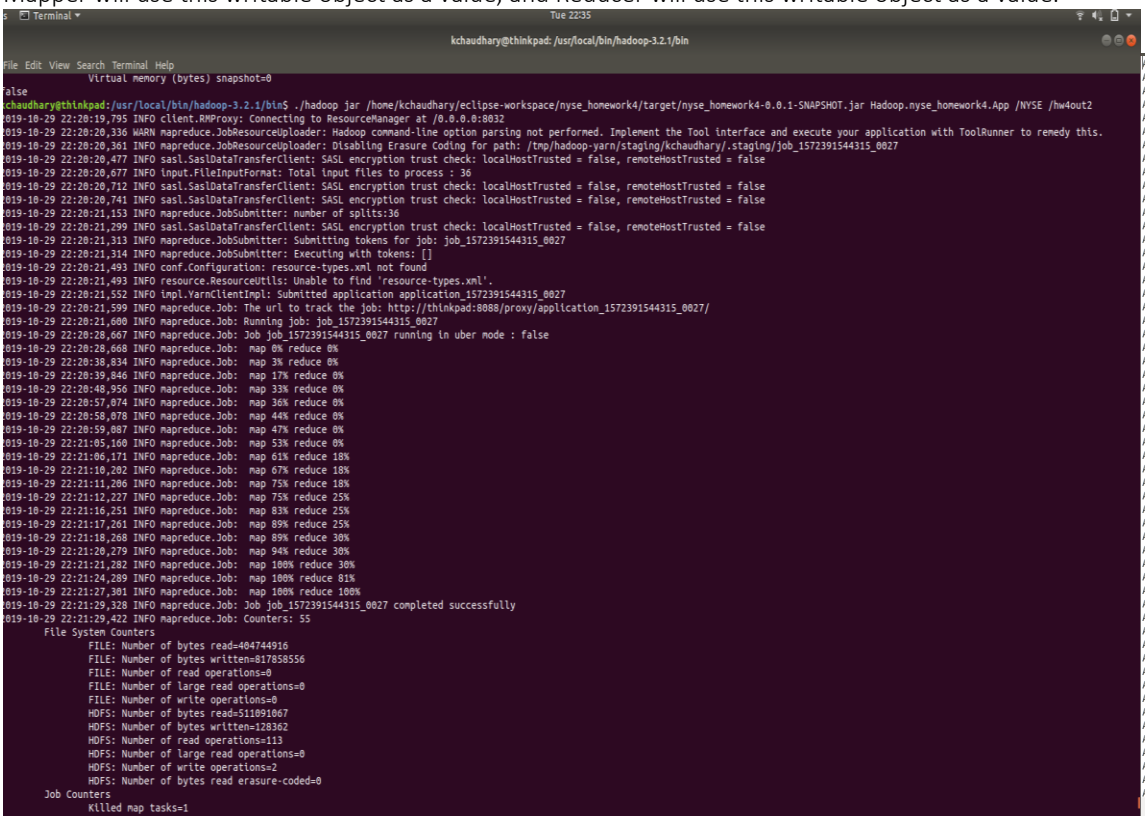
NAME : KAUSHAL CHAUDHARY
NUID : 001886763

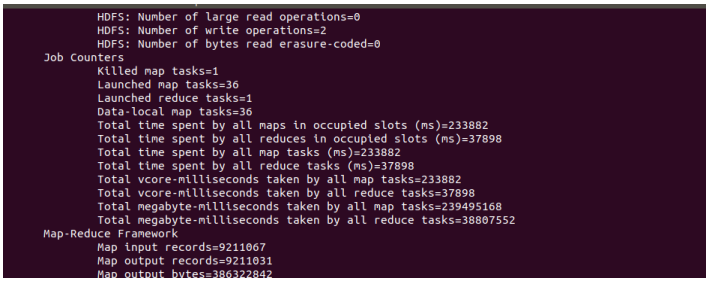
ASSIGNMENT 4

3. Create a Writable object that stores some fields from the the NYSE dataset to find
- the date of the max stock_volume
- the date of the min stock_volume
- the max stock_price_adj_close

This will be a custom writable class with the above fields.
Mapper will use this writable object as a value, and Reducer will use this writable object as a value.

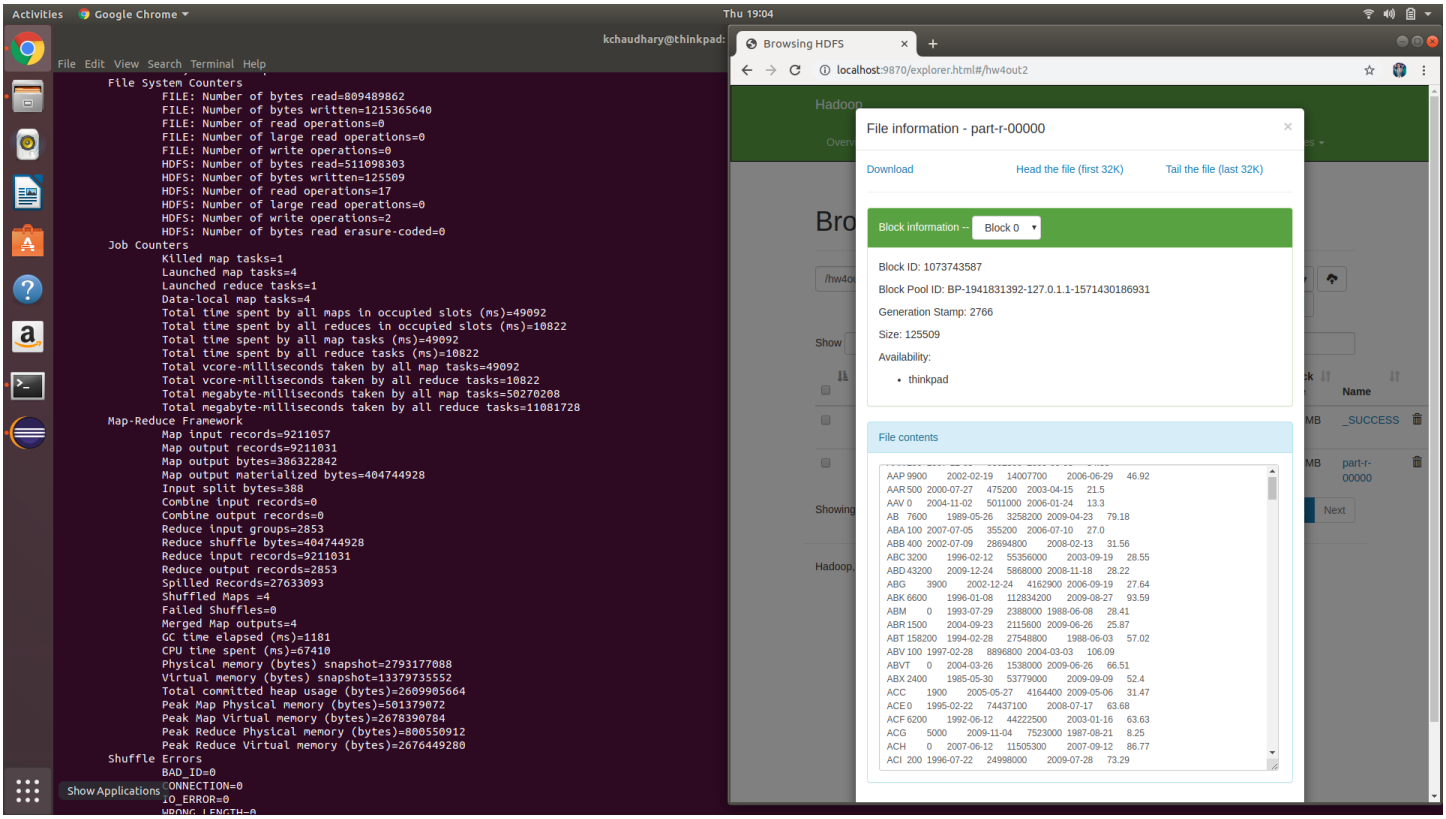



```
AA 1965-06-21 2009-03-19        44.18
AAI 1994-09-14 2009-10-08       33.5
AAN 1996-12-12 2009-06-08       34.36
AAP 2002-02-19 2006-06-29       46.92
AAR 2000-07-27 2003-04-15       21.5
AAV 2004-08-02 2006-01-24       13.3
AB 1989-05-26 2009-04-23        79.18
ABA 2007-07-05 2006-07-10       27.0
ABB 2002-07-09 2008-02-13       31.56
ABC 1996-02-12 2003-09-19       28.55
ABD 2009-12-24 2008-11-18       28.22
ABG 2002-12-24 2006-09-19       27.64
ABK 1996-01-08 2009-08-27       93.59
ABM 1993-07-29 1988-06-08       28.41
ABR 2004-09-23 2009-06-26       25.87
ABT 1994-02-28 1988-06-03       57.02
ABV 1997-04-02 2004-03-03       106.09
ABVT 2004-05-07 2009-06-26      66.51
ABX 1985-05-30 2009-09-09       52.4
ACC 2005-05-27 2009-05-06       31.47
ACE 1995-02-22 2008-07-17       63.68
ACF 1992-06-12 2003-01-16       63.63
ACG 2009-11-04 1987-08-21       8.25
ACH 2002-11-11 2007-09-12       86.77
ACI 1994-10-26 2009-07-28       73.29
ACL 2003-11-28 2010-01-04       169.14
ACM 2007-11-23 2007-05-10       37.25
ACN 2001-12-24 2009-08-31       43.75
ACO 1988-07-26 2006-07-21       38.49
ACS 1996-01-08 2001-06-26       63.92
ACV 1984-12-24 2006-11-16       29.92
ADC 2000-12-12 2003-07-30       27.87
ADI 1990-03-13 2006-08-11       89.87
ADM 1984-06-11 2007-02-16       46.6
ADP 1983-10-10 2003-03-13       52.67
ADS 2001-08-06 2008-01-28       80.72
ADX 1984-10-17 2002-01-02       13.48
ADY 2002-03-25 2009-07-13       43.16
AEA 2005-01-07 2005-03-02       17.69
AEB 2007-03-30 2006-09-06       20.07
AEC 1994-11-25 1996-10-15       13.67
AED 2005-11-25 2005-11-30       19.59
AEE 1998-12-24 2009-09-10       47.91
AEF 2007-09-19 2007-09-20       21.52
AEG 1986-11-11 2008-09-18       44.33
AEH 2006-07-03 2005-06-15       19.62
AEL 2004-11-26 2004-12-01       13.96
AEM 1985-09-27 2009-10-30       79.74
AEO 1996-05-24 2001-08-21       31.92
AEP 1971-06-14 1988-08-03       44.6
AER 2007-03-27 2006-11-21       32.54
AES 1993-09-23 2001-09-26       70.62
AET 2000-12-14 1997-09-23       59.58
AEV 2006-11-24 2006-08-08       20.46
AF 1996-12-24 1993-11-18        27.16
```



NAME : KAUSHAL CHAUDHARY
NUID : 001886763

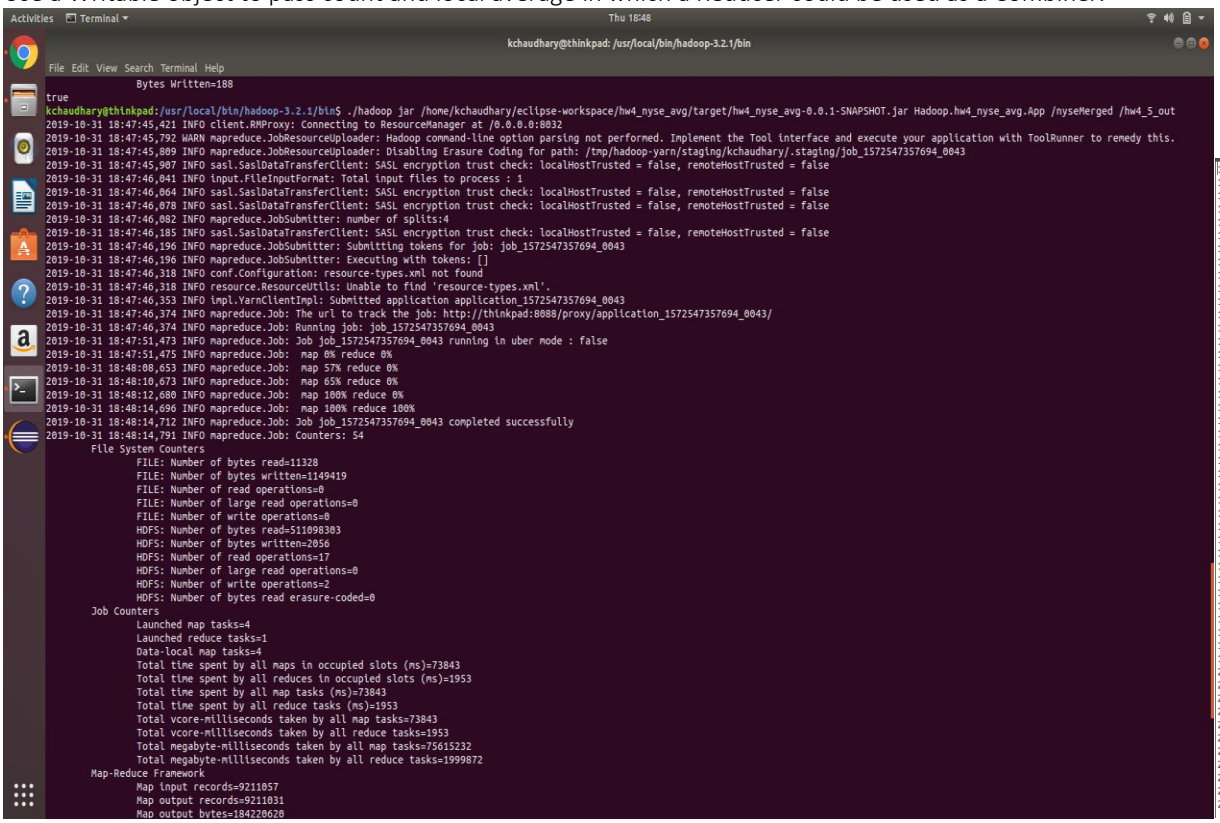
4. Redo Part3 of this assignment, but cram multiple values (max stock_volume, min stock_volume, max stock_price_adj_close) into a Text object with some delimiter. Use a Combiner. Compare the running time of Part 2 to Part 3. You could measure the running time programmatically, or use your smartphone's timer.



By comparing the time take by Part 3 and Part 4, we can clearly conclude that using combiner reduces optimizes the mapreduce job and takes comparatively less time if we use combiner along with reducer.

NAME : KAUSHAL CHAUDHARY
NUID : 001886763

5. Determine the average stock_price_adj_close value by the year.
Use a Writable object to pass count and local average in which a Reducer could be used as a Combiner.



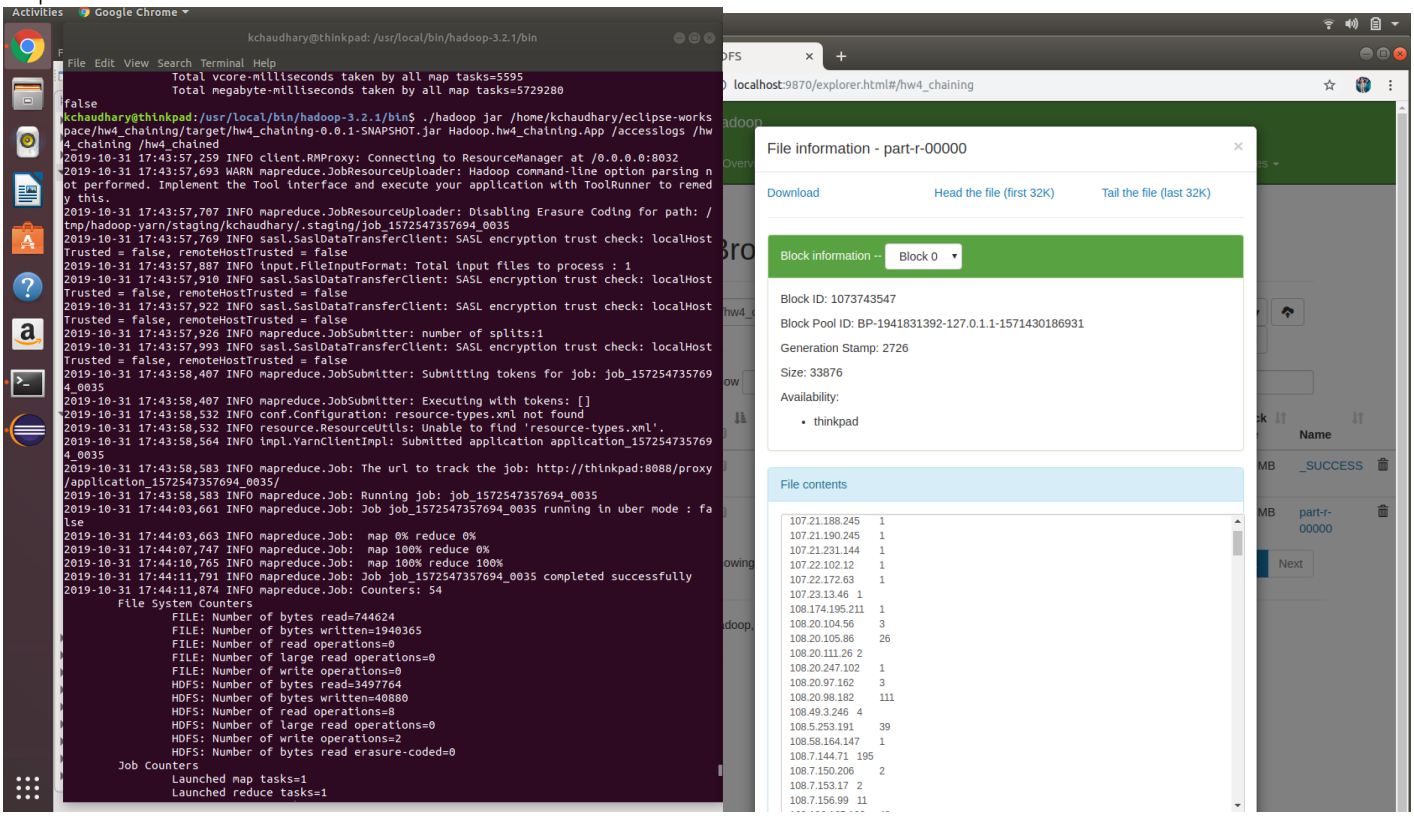
```
Activities  Terminal  Thu 18:48
kchaudhary@thinkpad: /usr/local/bin/hadoop-3.2.1/bin
File  Edit  View  Search  Terminal  Help
                Bytes Written=188
true
kchaudhary@thinkpad:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop jar /home/kchaudhary/eclipse-workspace/hw4_nyse_avg/target/hw4_nyse_avg-0.0.1-SNAPSHOT.jar Hadoop.hw4_nyse_avg.App /nyseMerged /hw4_5_out
2019-10-31 18:47:45,421 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-10-31 18:47:45,792 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2019-10-31 18:47:45,809 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kchaudhary/.staging/job_1572547357694_0043
2019-10-31 18:47:45,907 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2019-10-31 18:47:46,041 INFO input.FileInputFormat: Total input files to process : 1
2019-10-31 18:47:46,064 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2019-10-31 18:47:46,078 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2019-10-31 18:47:46,082 INFO mapreduce.JobSubmitter: number of splits:4
2019-10-31 18:47:46,185 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2019-10-31 18:47:46,196 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1572547357694_0043
2019-10-31 18:47:46,196 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-10-31 18:47:46,318 INFO conf.Configuration: resource-types.xml not found
2019-10-31 18:47:46,318 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-10-31 18:47:46,353 INFO impl.YarnClientImpl: Submitted application application_1572547357694_0043
2019-10-31 18:47:46,374 INFO mapreduce.Job: The url to track the job: http://thinkpad:8088/proxy/application_1572547357694_0043/
2019-10-31 18:47:46,374 INFO mapreduce.Job: Running job: job_1572547357694_0043
2019-10-31 18:47:51,473 INFO mapreduce.Job: Job job_1572547357694_0043 running in uber mode : false
2019-10-31 18:47:51,475 INFO mapreduce.Job:  map 0% reduce 0%
2019-10-31 18:48:08,653 INFO mapreduce.Job:  map 57% reduce 0%
2019-10-31 18:48:10,673 INFO mapreduce.Job:  map 65% reduce 0%
2019-10-31 18:48:12,680 INFO mapreduce.Job:  map 100% reduce 0%
2019-10-31 18:48:14,696 INFO mapreduce.Job:  map 100% reduce 100%
2019-10-31 18:48:14,712 INFO mapreduce.Job: Job job_1572547357694_0043 completed successfully
2019-10-31 18:48:14,791 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=11328
                FILE: Number of bytes written=1149419
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=511098303
                HDFS: Number of bytes written=2056
                HDFS: Number of read operations=17
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=4
                Launched reduce tasks=1
                Data-local map tasks=4
                Total time spent by all maps in occupied slots (ms)=73843
                Total time spent by all reduces in occupied slots (ms)=1953
                Total time spent by all map tasks (ms)=73843
                Total time spent by all reduce tasks (ms)=1953
                Total vcore-milliseconds taken by all map tasks=73843
                Total vcore-milliseconds taken by all reduce tasks=1953
                Total megabyte-milliseconds taken by all map tasks=75615232
                Total megabyte-milliseconds taken by all reduce tasks=1999872
        Map-Reduce Framework
                Map input records=9211057
                Map output records=9211031
                Map output bytes=184220620
```

```
1962    0.85143253968254
1963    0.9509047429254671
1964    1.1544624505928869
1965    1.3094603174603174
1966    1.4863437872171503
1967    1.7562310756972117
1968    1.9254026548672578
1969    1.9137759999999988
1970    4.814298126527288
1971    6.18895052473763
1972    7.9846829559945684
1973    7.962886243386245 5
1974    6.684673254281948
1975    8.133909090909086
1976    10.459285902503295
1977    11.797204440974934
1978    12.009024470899469
1979    13.232830204216075
1980    9.588325856443717
1981    6.137044407345577
1982    3.6106789640394967
1983    5.149613194464931
1984    4.81202206380002
1985    4.924612230290933
1986    6.264915391445134
1987    7.210992624050059
1988    6.003965276478123
1989    6.582972778270659
1990    6.369104150582308 5
1991    7.150954581025736
1992    7.627844074790678
1993    8.653585618486957
1994    9.456396925196396
1995    10.286344568170525
1996    12.417022600792317
1997    15.495556959860403
1998    17.235555482490355
1999    17.46746826912278
2000    18.95265553398255
2001    19.201277037690847
2002    18.6722393498568
2003    19.323442006411867
2004    23.934410132039865
2005    26.613839407225537
2006    29.439247491528974
2007    33.30495633878515
2008    27.273130155793865
2009    22.020622744171646
2010    26.84449212338855
```

NAME : KAUSHAL CHAUDHARY
NUID : 001886763

6. Top 10 most visiting IP Addresses (HW3-Part3). This can be achieved by 2 jobs :
   A map reduce job to find how many time each item is bought.
   The second job, sort items based on number of times it was bought, and get top 10 items.
   The second MapReduce job swaps key and value so that we get words sorted in descending order by frequency.
   Use Chaining to find the Top 10 IP addresses based on access count.

MapReduce Job 1



MapReduce Job 2