

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.

```

Terminal
Tue 22:35
kchaudhary@thinkpad: /usr/local/bin/hadoop-3.2.1/bin

File Edit View Search Terminal Help
Virtual memory (bytes) snapshot=0
else
kchaudhary@thinkpad: /usr/local/bin/hadoop-3.2.1/bin$ ./hadoop jar /home/kchaudhary/eclipse-workspace/nyse_homework4/target/nyse_homework4-0.0.1-SNAPSHOT.jar Hadoop.nyse_homework4.App /NYSE /hw4out2
2019-10-29 22:20:19,795 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-10-29 22:20:20,336 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
2019-10-29 22:20:20,361 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kchaudhary/.staging/job_1572391544315_0027
2019-10-29 22:20:20,477 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2019-10-29 22:20:20,577 INFO InputFileInputFormat: Total input files to process = 36
2019-10-29 22:20:20,712 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2019-10-29 22:20:20,741 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2019-10-29 22:20:21,153 INFO mapreduce.JobSubmitter: number of splits:36
2019-10-29 22:20:21,299 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2019-10-29 22:20:21,313 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1572391544315_0027
2019-10-29 22:20:21,314 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-10-29 22:20:21,493 INFO conf.Configuration: resource-types.xml not found
2019-10-29 22:20:21,493 INFO resource.ResourceUtil: Unable to find 'resource-types.xml'.
2019-10-29 22:20:21,552 INFO Impl.VarnClientImpl: Submitted application application_1572391544315_0027
2019-10-29 22:20:21,599 INFO mapreduce.Job: The url to track the job: http://thinkpad:8088/proxy/application_1572391544315_0027/
2019-10-29 22:20:21,600 INFO mapreduce.Job: Running job: job_1572391544315_0027
2019-10-29 22:20:28,607 INFO mapreduce.Job: Job job_1572391544315_0027 running in uber mode : false
2019-10-29 22:20:28,668 INFO mapreduce.Job: map 0% reduce 0%
2019-10-29 22:20:38,834 INFO mapreduce.Job: map 3% reduce 0%
2019-10-29 22:20:39,846 INFO mapreduce.Job: map 17% reduce 0%
2019-10-29 22:20:40,956 INFO mapreduce.Job: map 33% reduce 0%
2019-10-29 22:20:57,074 INFO mapreduce.Job: map 36% reduce 0%
2019-10-29 22:20:58,078 INFO mapreduce.Job: map 44% reduce 0%
2019-10-29 22:20:59,087 INFO mapreduce.Job: map 47% reduce 0%
2019-10-29 22:21:05,160 INFO mapreduce.Job: map 53% reduce 0%
2019-10-29 22:21:06,171 INFO mapreduce.Job: map 61% reduce 18%
2019-10-29 22:21:10,202 INFO mapreduce.Job: map 67% reduce 18%
2019-10-29 22:21:11,206 INFO mapreduce.Job: map 75% reduce 18%
2019-10-29 22:21:12,227 INFO mapreduce.Job: map 75% reduce 25%
2019-10-29 22:21:16,251 INFO mapreduce.Job: map 83% reduce 25%
2019-10-29 22:21:17,261 INFO mapreduce.Job: map 89% reduce 25%
2019-10-29 22:21:18,268 INFO mapreduce.Job: map 89% reduce 30%
2019-10-29 22:21:20,279 INFO mapreduce.Job: map 94% reduce 30%
2019-10-29 22:21:21,282 INFO mapreduce.Job: map 100% reduce 30%
2019-10-29 22:21:24,289 INFO mapreduce.Job: map 100% reduce 81%
2019-10-29 22:21:27,301 INFO mapreduce.Job: map 100% reduce 100%
2019-10-29 22:21:29,328 INFO mapreduce.Job: Job job_1572391544315_0027 completed successfully
2019-10-29 22:21:29,422 INFO mapreduce.Job: Counters: 55

File System Counters
  FILE: Number of bytes read=404744916
  FILE: Number of bytes written=817858556
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=511891067
  HDFS: Number of bytes written=128362
  HDFS: Number of read operations=113
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0

Job Counters
  Killed map tasks=1

  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0

Job Counters
  Killed map tasks=1
  Launched map tasks=36
  Launched reduce tasks=1
  Data-Local map tasks=36
  Total time spent by all maps in occupied slots (ms)=233882
  Total time spent by all reduces in occupied slots (ms)=37898
  Total time spent by all map tasks (ms)=233882
  Total time spent by all reduce tasks (ms)=37898
  Total vcore-millisecods taken by all map tasks=233882
  Total vcore-millisecods taken by all reduce tasks=37898
  Total megabyte-millisecods taken by all map tasks=239495168
  Total megabyte-millisecods taken by all reduce tasks=38807552

Map-Reduce Framework
  Map input records=9211067
  Map output records=9211031
  Map output bytes=186322842
  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 63.08
  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

```

```

HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0

Job Counters
  Killed map tasks=1
  Launched map tasks=36
  Launched reduce tasks=1
  Data-Local map tasks=36
  Total time spent by all maps in occupied slots (ms)=233882
  Total time spent by all reduces in occupied slots (ms)=37898
  Total time spent by all map tasks (ms)=233882
  Total time spent by all reduce tasks (ms)=37898
  Total vcore-millisecods taken by all map tasks=233882
  Total vcore-millisecods taken by all reduce tasks=37898
  Total megabyte-millisecods taken by all map tasks=239495168
  Total megabyte-millisecods taken by all reduce tasks=38807552

Map-Reduce Framework
  Map input records=9211067
  Map output records=9211031
  Map output bytes=186322842

```

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.

The screenshot shows a terminal window on the left and a web browser on the right. The terminal displays the output of the `hadoop job -list` command, showing the progress of a Hadoop job. The web browser shows the HDFS file information for the file `part-r-00000`.

File System Counters

```

FILE: Number of bytes read=809489862
FILE: Number of bytes written=1215365640
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=125509
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

```

Killed map tasks=1
Launched map tasks=4
Launched reduce tasks=1
Data-local map tasks=4
Total time spent by all maps in occupied slots (ms)=49092
Total time spent by all reduces in occupied slots (ms)=10822
Total time spent by all map tasks (ms)=49092
Total time spent by all reduce tasks (ms)=10822
Total vcore-millisecods taken by all map tasks=49092
Total vcore-millisecods taken by all reduce tasks=10822
Total megabyte-millisecods taken by all map tasks=50270208
Total megabyte-millisecods taken by all reduce tasks=11081728

```

Map-Reduce Framework

```

Map input records=9211057
Map output records=9211031
Map output bytes=386322842
Map output materialized bytes=404744928
Input split bytes=388
Combine input records=0
Combine output records=0
Reduce input groups=2853
Reduce shuffle bytes=404744928
Reduce input records=9211031
Reduce output records=2853
Spilled Records=27633093
Shuffled Maps=4
Failed Shuffles=0
Merged Map outputs=4
GC time elapsed (ms)=1181
CPU time spent (ms)=67410
Physical memory (bytes) snapshot=2793177888
Virtual memory (bytes) snapshot=1337973552
Total committed heap usage (bytes)=2609905664
Peak Map Physical memory (bytes)=501379072
Peak Map Virtual memory (bytes)=2678390784
Peak Reduce Physical memory (bytes)=808550912
Peak Reduce Virtual memory (bytes)=2676449280

```

Shuffle Errors

```

BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0

```

File information - part-r-00000

Download Head the file (first 32K) Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073743587
Block Pool ID: BP-1941831392-127.0.1.1-1571430186931
Generation Stamp: 2766
Size: 125509
Availability:
• thinkpad

File contents

```

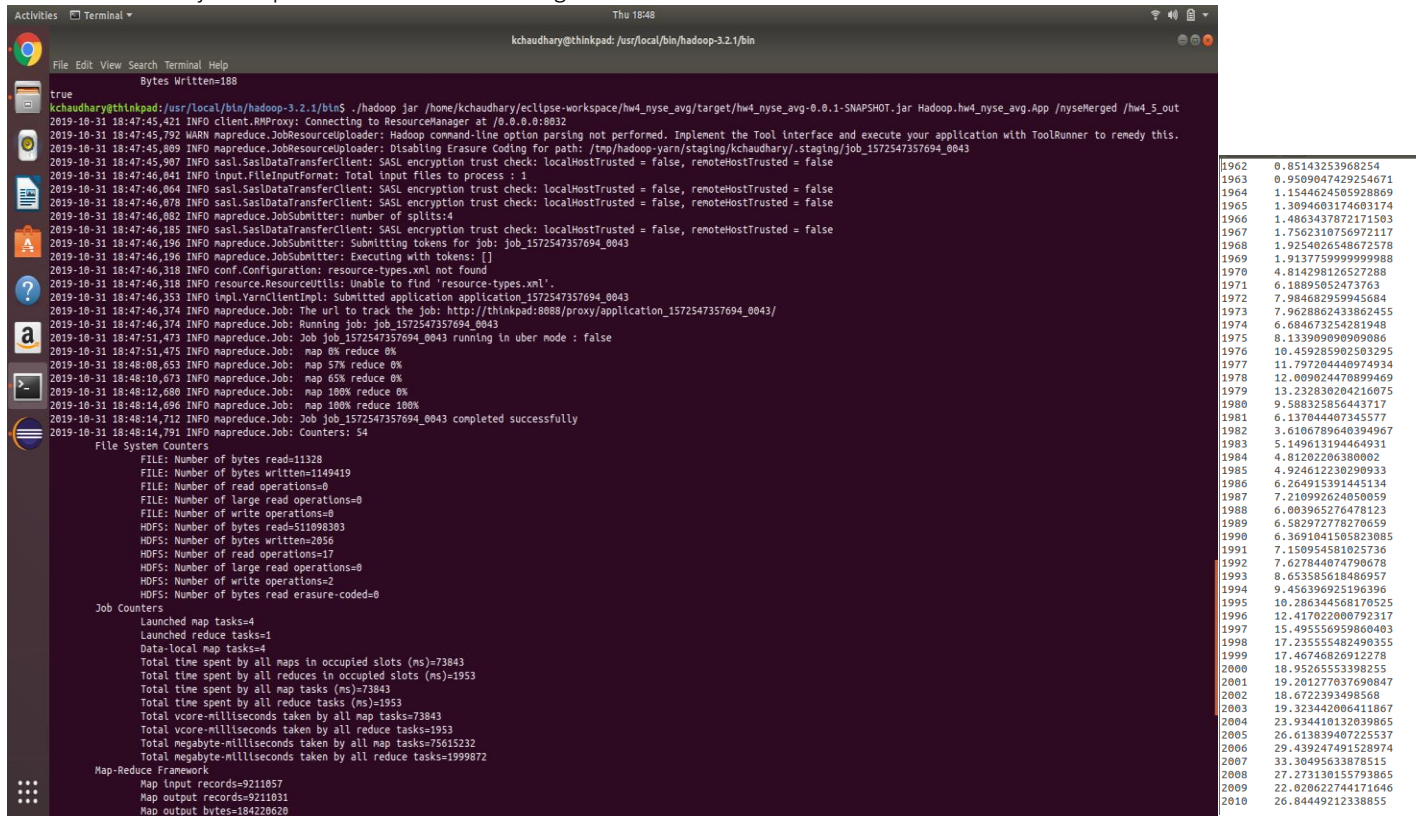
AAP 9900 2002-02-19 14007700 2006-06-29 46.92
AAR 500 2000-07-27 475200 2003-04-15 21.5
AAV 0 2004-11-02 5011000 2006-01-24 13.3
AB 7000 1989-05-26 3258200 2009-04-23 79.18
ABA 100 2007-07-05 355200 2006-07-10 27.0
ABB 400 2002-07-09 28694800 2008-02-13 31.56
ABC 3200 1996-02-12 55356000 2003-09-19 28.55
ABD 43200 2009-12-24 5868000 2008-11-18 28.22
ABG 3900 2002-12-24 4162900 2006-09-19 27.64
ABK 6800 1996-01-08 112834200 2009-08-27 93.59
ABM 0 1993-07-29 2388000 1988-06-08 28.41
ABR 1500 2004-09-23 2115600 2009-06-26 25.87
ABT 158200 1994-02-28 27548800 1988-06-03 57.02
ABV 100 1997-02-28 8896800 2004-03-03 106.09
ABVT 0 2004-03-26 1538000 2009-06-26 66.51
ABX 2400 1985-05-30 53779000 2009-09-09 52.4
ACC 1900 2005-05-27 4164400 2009-05-06 31.47
ACE 0 1995-02-22 74437100 2008-07-17 63.68
ACF 6200 1992-06-12 44222500 2003-01-16 63.63
ACG 5000 2009-11-04 7523000 1987-08-21 8.25
ACH 0 2007-06-12 11505300 2007-09-12 86.77
ACI 200 1996-07-22 24998000 2009-07-28 73.29

```

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.

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
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 /nyseHerged /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.VarnClientImpl: 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-milliseonds taken by all map tasks=73843
  Total vcore-milliseonds taken by all reduce tasks=1953
  Total megabyte-milliseonds taken by all map tasks=75615232
  Total megabyte-milliseonds 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 1.913775999999988
1970 4.814298126527288
1971 6.18895052473763
1972 7.984682959945684
1973 7.9628862433062455
1974 6.684673254281948
1975 8.133909090909086
1976 10.459285902503295
1977 11.797204440974934
1978 12.009024470899469
1979 13.232830204216075
1980 9.58832585643717
1981 6.137044447345577
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 6.3691041505823085
1991 7.150954581025736
1992 7.627844074790678
1993 8.653585618486957
1994 9.456396925196396
1995 10.286344568170525
1996 12.417022000792317
1997 15.495556959860403
1998 17.23555482490355
1999 17.46746826912278
2000 18.9526553398255
2001 19.201277037690847
2002 18.6722393498568
2003 19.323442006411867
2004 23.93441013209865
2005 26.61389407225537
2006 29.439247491528974
2007 33.36495633878515
2008 27.273130155793065
2009 22.020622744171646
2010 26.84449212338855
```

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

The screenshot shows the execution of MapReduce Job 1. On the left, a terminal window displays the Hadoop command-line interface output. On the right, a web browser window shows the Hadoop Distributed File System (HDFS) Explorer interface.

Terminal Output:

```

Total vcore-millisecods taken by all map tasks=5595
Total megabyte-millisecods taken by all map tasks=5729280
false
kchaudhary@thinkpad: /usr/local/bin/hadoop-3.2.1/bin$ ./hadoop jar /home/kchaudhary/eclipse-works
space/hw4_chaining/target/hw4_chaining-0.1-SNAPSHOT.jar Hadoop.hw4_chaining.App /accesslogs /hw
4_chaining /hw4_chained
2019-10-31 17:43:57,259 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-10-31 17:43:57,693 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing n
ot performed. Implement the Tool interface and execute your application with ToolRunner to reme
dy this.
2019-10-31 17:43:57,707 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /
tmp/hadoop-yarn/staging/kchaudhary/.staging/job_1572547357694_0035
2019-10-31 17:43:57,769 INFO sasL.SaslDataTransferClient: SASL encryption trust check: localHost
Trusted = false, remoteHostTrusted = false
2019-10-31 17:43:57,922 INFO sasL.SaslDataTransferClient: SASL encryption trust check: localHost
Trusted = false, remoteHostTrusted = false
2019-10-31 17:43:57,926 INFO mapreduce.JobSubmitter: number of splits:1
2019-10-31 17:43:57,993 INFO sasL.SaslDataTransferClient: SASL encryption trust check: localHost
Trusted = false, remoteHostTrusted = false
2019-10-31 17:43:58,407 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_157254735769
4_0035
2019-10-31 17:43:58,407 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-10-31 17:43:58,532 INFO conf.Configuration: resource-types.xml not found
2019-10-31 17:43:58,532 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-10-31 17:43:58,564 INFO Impl.YarnClientImpl: Submitted application application_157254735769
4_0035
2019-10-31 17:43:58,583 INFO mapreduce.Job: The url to track the job: http://thinkpad:8088/proxy
/application_1572547357694_0035/
2019-10-31 17:43:58,583 INFO mapreduce.Job: Running job: job_1572547357694_0035
2019-10-31 17:44:03,661 INFO mapreduce.Job: Job job_1572547357694_0035 running in uber mode : fa
lse
2019-10-31 17:44:03,663 INFO mapreduce.Job: map 0% reduce 0%
2019-10-31 17:44:07,747 INFO mapreduce.Job: map 100% reduce 0%
2019-10-31 17:44:10,765 INFO mapreduce.Job: map 100% reduce 100%
2019-10-31 17:44:11,791 INFO mapreduce.Job: Job job_1572547357694_0035 completed successfully
2019-10-31 17:44:11,874 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=744624
  FILE: Number of bytes written=1940365
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=3497764
  HDFS: Number of bytes written=40880
  HDFS: Number of read operations=0
  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=1
  Launched reduce tasks=1

```

Web Browser Output:

The web browser shows the HDFS Explorer interface. A file information window is open for the file 'part-r-00000'. The file information includes:

- Block ID: 1073743547
- Block Pool ID: BP-1941831392-127.0.1.1-1571430186931
- Generation Stamp: 2726
- Size: 33876
- Availability: thinkpad

The file contents window shows the following data:

```

107.21.188.245 1
107.21.190.245 1
107.21.231.144 1
107.22.102.12 1
107.22.102.63 1
107.23.13.46 1
108.174.195.211 3
108.20.104.56 3
108.20.105.86 26
108.20.111.26 2
108.20.247.102 1
108.20.97.162 3
108.20.98.182 111
108.49.3.248 4
108.5.253.191 39
108.58.164.147 1
108.7.144.71 195
108.7.150.206 2
108.7.153.17 2
108.7.156.99 11

```

MapReduce Job 2

The screenshot shows the execution of MapReduce Job 2. On the left, a terminal window displays the Hadoop command-line interface output. On the right, a web browser window shows the HDFS Explorer interface.

Terminal Output:

```

HDFS: Number of read operations=0
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=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=1355
  Total time spent by all reduces in occupied slots (ms)=1477
  Total time spent by all map tasks (ms)=1355
  Total time spent by all reduce tasks (ms)=1477
  Total vcore-millisecods taken by all map tasks=1355
  Total vcore-millisecods taken by all reduce tasks=1477
  Total megabyte-millisecods taken by all map tasks=1387520
  Total megabyte-millisecods taken by all reduce tasks=1512448
Map-Reduce Framework
  Map input records=1945
  Map output records=10
  Map output bytes=222
  Map output materialized bytes=248
  Input split bytes=112
  Combine input records=0
  Combine output records=0
  Reduce input groups=10
  Reduce shuffle bytes=248
  Reduce input records=10
  Reduce output records=10
  Spilled Records=20
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=54
  CPU time spent (ms)=870
  Physical memory (bytes) snapshot=508289024
  Virtual memory (bytes) snapshot=5348540416
  Total committed heap usage (bytes)=509083648
  Peak Map Physical memory (bytes)=30327504
  Peak Map Virtual memory (bytes)=2672390144
  Peak Reduce Physical memory (bytes)=204861440
  Peak Reduce Virtual memory (bytes)=2676150272
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=33876
File Output Format Counters

```

Web Browser Output:

The web browser shows the HDFS Explorer interface. A file information window is open for the file 'part-r-00000'. The file information includes:

- Block ID: 1073743537
- Block Pool ID: BP-1941831392-127.0.1.1-1571430186931
- Generation Stamp: 2716
- Size: 188
- Availability: thinkpad

The file contents window shows the following data:

```

642 118.102.182.196
643 72.158.153.33
667 50.63.154.43
765 188.144.67.144
1279 107.20.213.124
1501 129.10.65.240
2108 10.15.10.135
2812 10.15.10.129
3724 207.248.55.246
4958 155.35.18.236

```