

You need to use input file "**wordcountproblem**" present at location **/home/data** in Jigsaw Lab

Questions:

Based on the counters and messages displayed while executing your mapreduce job on-screen and answer the following questions

Score for this quiz: 27 out of 30  
Submitted Nov 29, 2017, 12:05pm  
This attempt took 158 minutes.

⋮

Question 1

2 / 2 pts

You need to execute wc\_0reducer.jar for input file wordcountproblem and below questions accordingly:

**Q. Total input paths to process:**

1.0000

1 (with margin: 0)

⋮

Question 2

2 / 2 pts

You need to execute wc\_0reducer.jar for input file wordcountproblem and below questions accordingly:

**Q. HDFS: Number of read operations:**

5.0000

5 (with margin: 0)

⋮

Question 3

2 / 2 pts

---

You need to execute wc\_0reducer.jar for input file wordcountproblem and below questions accordingly:

**Q. Total number of Launched map tasks:**

1.0000

1 (with margin: 0)

⋮

Question 4

2 / 2 pts

You need to execute wc\_2reducer.jar for input file wordcountproblem and answer the below questions.

**Q. Total number of Launched reduce tasks:**

2.0000

2 (with margin: 0)

⋮

Question 5

2 / 2 pts

You need to execute wc\_2reducer.jar for input file wordcountproblem and answer the below questions.

**Q. Find the number of times word *Hadoop* is repeated in wordcountproblem file while executing wc\_2reducer.jar**

394.0000

394 (with margin: 0)

⋮

Question 6

2 / 2 pts

You need to execute wc\_defaultreducer.jar for input file wordcountproblem and answer below questions.

**Q. Total number of HDFS: Number of read operations?**

9.0000

9 (with margin: 0)

⋮

Question 7

2 / 2 pts

You need to execute wc\_defaultreducer.jar for input file wordcountproblem and answer below questions.

**Q. What is the total number of output files?**

2.0000

2 (with margin: 0)

⋮

Question 8

2 / 2 pts

You need to execute wc.jar for input file wordcountproblem and answer below questions

**Q. Based on counters, you need to tell Total Number of distinct word in file wordcountproblem:**

23.0000

351 (with margin: 0)

⋮

Question 10

2 / 2 pts

You need to execute `wc.jar` for input file `wordcountproblem` and answer below questions

**Q. How many times word *Execution* is repeated:**

6.0000

6 (with margin: 0)

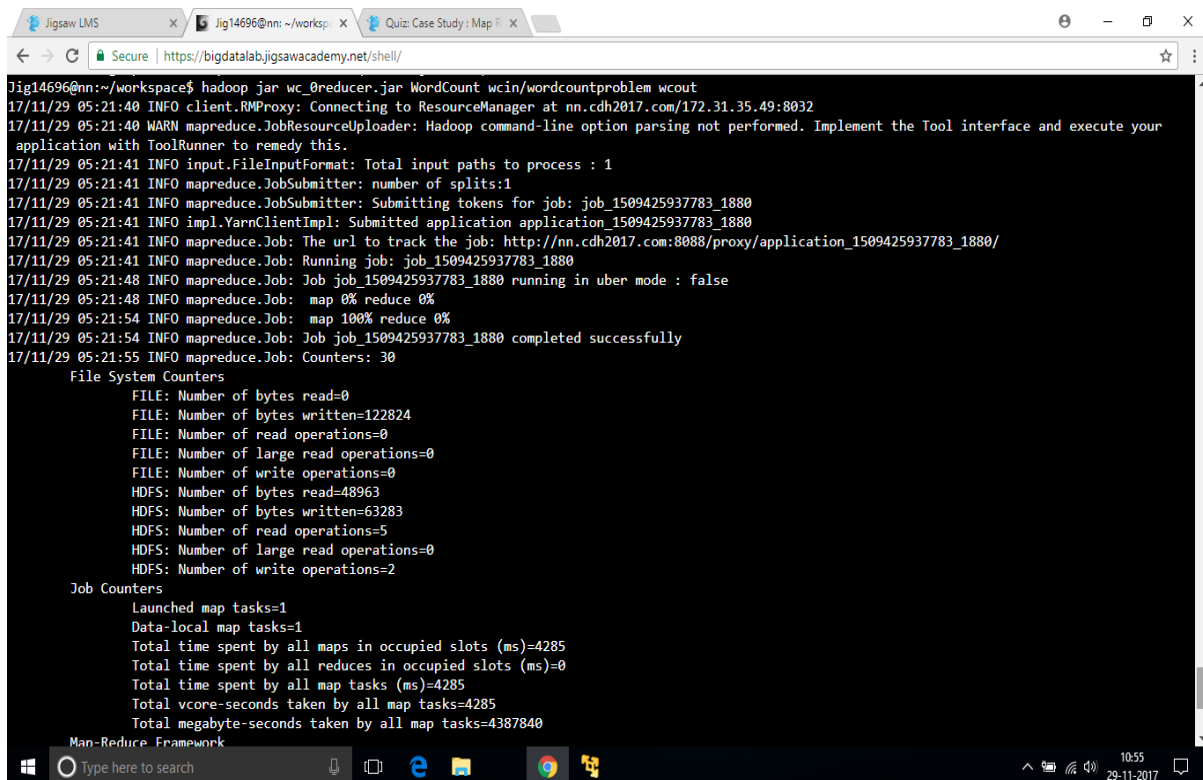
⋮

Question 11

Not yet graded / 10 pts

Please create a document with the screenshots of the output for the problem given in the description and upload the same in the option given below.

For `wc_0reducer.jar` questions(1,2,3):



```
Jigsaw LMS X Jig14696@nn: ~/workspace X Quiz: Case Study : Map X
Secure | https://bigdatalab.jigsawacademy.net/shell/
Jig14696@nn:~/workspace$ hadoop jar wc_0reducer.jar WordCount wcin/wordcountproblem wcout
17/11/29 05:21:40 INFO client.RMPProxy: Connecting to ResourceManager at nn.cdh2017.com/172.31.35.49:8032
17/11/29 05:21:40 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
17/11/29 05:21:41 INFO input.FileInputFormat: Total input paths to process : 1
17/11/29 05:21:41 INFO mapreduce.JobSubmitter: number of splits:1
17/11/29 05:21:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1509425937783_1880
17/11/29 05:21:41 INFO impl.YarnClientImpl: Submitted application application_1509425937783_1880
17/11/29 05:21:41 INFO mapreduce.Job: The url to track the job: http://nn.cdh2017.com:8088/proxy/application_1509425937783_1880/
17/11/29 05:21:41 INFO mapreduce.Job: Running job: job_1509425937783_1880
17/11/29 05:21:48 INFO mapreduce.Job: Job job_1509425937783_1880 running in uber mode : false
17/11/29 05:21:48 INFO mapreduce.Job: map 0% reduce 0%
17/11/29 05:21:54 INFO mapreduce.Job: map 100% reduce 0%
17/11/29 05:21:54 INFO mapreduce.Job: Job job_1509425937783_1880 completed successfully
17/11/29 05:21:55 INFO mapreduce.Job: Counters: 30

File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=122824
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=48963
  HDFS: Number of bytes written=63283
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4285
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4285
  Total vcore-seconds taken by all map tasks=4285
  Total megabyte-seconds taken by all map tasks=4387840

Man-Reduce Framework
```

```
Jigsaw LMS X Jig14696@nn: ~/worksp X Quiz: Case Study: Map X
Secure | https://bigdatalab.jigsawacademy.net/shell/
FILE: Number of bytes written=122824
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=48963
HDFS: Number of bytes written=63283
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4285
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4285
  Total vcore-seconds taken by all map tasks=4285
  Total megabyte-seconds taken by all map tasks=4387840
Map-Reduce Framework
  Map input records=103
  Map output records=7352
  Input split bytes=127
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=54
  CPU time spent (ms)=1010
  Physical memory (bytes) snapshot=182263808
  Virtual memory (bytes) snapshot=1370468352
  Total committed heap usage (bytes)=159383552
File Input Format Counters
  Bytes Read=48836
File Output Format Counters
  Bytes Written=63283
Jig14696@nn:~/workspace$
```

For wc\_2reducer.jar questions(4,5)

Question 4:

```
Jigsaw LMS x Jig14696@nn: ~/workspace x Quiz: Case Study: Map F x
Secure | https://bigdatalab.jigsawacademy.net/shell/
Jig14696@nn:~/workspace$ hadoop jar wc_2reducer.jar WordCount wcin/wordcountproblem wcout1
17/11/29 05:31:29 INFO client.RMPProxy: Connecting to ResourceManager at nn.cdh2017.com/172.31.35.49:8032
17/11/29 05:31:29 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
17/11/29 05:31:29 INFO input.FileInputFormat: Total input paths to process : 1
17/11/29 05:31:29 INFO mapreduce.JobSubmitter: number of splits:1
17/11/29 05:31:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1509425937783_1883
17/11/29 05:31:30 INFO impl.YarnClientImpl: Submitted application application_1509425937783_1883
17/11/29 05:31:30 INFO mapreduce.Job: The url to track the job: http://nn.cdh2017.com:8088/proxy/application_1509425937783_1883/
17/11/29 05:31:30 INFO mapreduce.Job: Running job: job_1509425937783_1883
17/11/29 05:31:37 INFO mapreduce.Job: Job job_1509425937783_1883 running in uber mode : false
17/11/29 05:31:37 INFO mapreduce.Job: map 0% reduce 0%
17/11/29 05:31:43 INFO mapreduce.Job: map 100% reduce 0%
17/11/29 05:31:54 INFO mapreduce.Job: map 100% reduce 100%
17/11/29 05:31:54 INFO mapreduce.Job: Job job_1509425937783_1883 completed successfully
17/11/29 05:31:54 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=8346
  FILE: Number of bytes written=385565
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=48963
  HDFS: Number of bytes written=3799
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4

Job Counters
  Launched map tasks=1
  Launched reduce tasks=2
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4261
  Total time spent by all reduces in occupied slots (ms)=16457
  Total time spent by all map tasks (ms)=4261
  Total time spent by all reduce tasks (ms)=16457

Type here to search
```

```
Jigsaw LMS x Jig14696@nn: ~/workspace x Quiz: Case Study: Map F x
Secure | https://bigdatalab.jigsawacademy.net/shell/
Total time spent by all map tasks (ms)=4261
Total time spent by all reduce tasks (ms)=16457
Total vcore-seconds taken by all map tasks=4261
Total vcore-seconds taken by all reduce tasks=16457
Total megabyte-seconds taken by all map tasks=4363264
Total megabyte-seconds taken by all reduce tasks=16851968

Map-Reduce Framework
  Map input records=103
  Map output records=7352
  Map output bytes=77987
  Map output materialized bytes=8338
  Input split bytes=127
  Combine input records=0
  Combine output records=0
  Reduce input groups=351
  Reduce shuffle bytes=8338
  Reduce input records=7352
  Reduce output records=351
  Spilled Records=14704
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=225
  CPU time spent (ms)=3910
  Physical memory (bytes) snapshot=821014528
  Virtual memory (bytes) snapshot=4138467328
  Total committed heap usage (bytes)=747110400

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
```

```
Jigsaw LMS x Jig14696@nn: ~/worksp... Quiz: Case Study : Map...
Secure | https://bigdatalab.jigsawacademy.net/shell/

Total megabyte-seconds taken by all reduce tasks=16851968
Map-Reduce Framework
  Map input records=103
  Map output records=7352
  Map output bytes=77987
  Map output materialized bytes=8338
  Input split bytes=127
  Combine input records=0
  Combine output records=0
  Reduce input groups=351
  Reduce shuffle bytes=8338
  Reduce input records=7352
  Reduce output records=351
  Spilled Records=14704
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=225
  CPU time spent (ms)=3910
  Physical memory (bytes) snapshot=821014528
  Virtual memory (bytes) snapshot=4138467328
  Total committed heap usage (bytes)=747110400

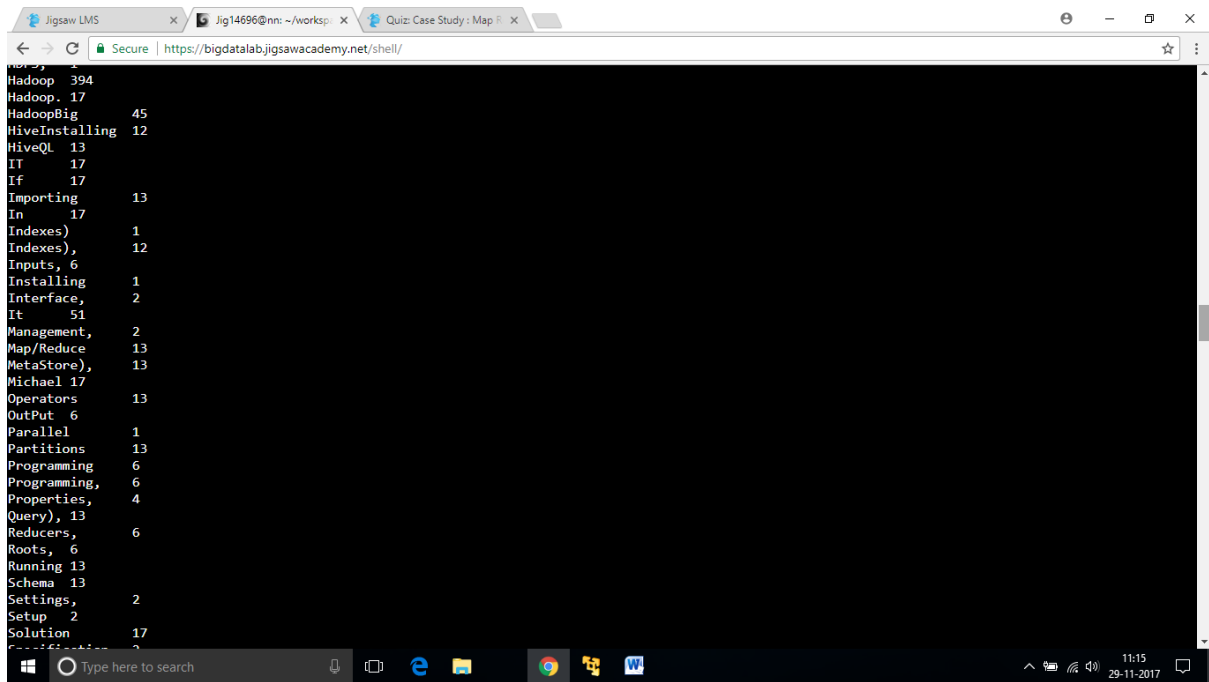
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=48836
File Output Format Counters
  Bytes Written=3799
Jig14696@nn:~/workspace$
```

## Question 5:

```
Jigsaw LMS x Jig14696@nn: ~/worksp... Quiz: Case Study : Map...
Secure | https://bigdatalab.jigsawacademy.net/shell/

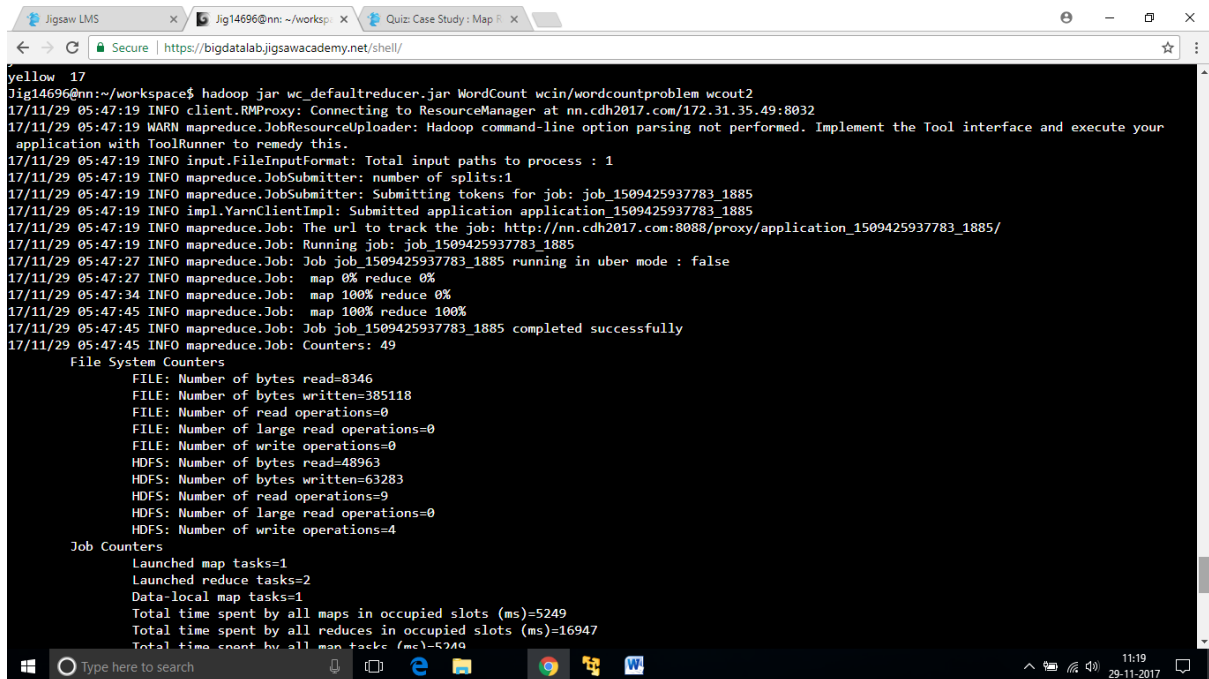
cat: wcout1: is a directory
Jig14696@nn:~/workspace$ hadoop fs -ls wcout1
Found 3 items
-rw-r--r-- 3 Jig14696 supergroup          0 2017-11-29 05:31 wcout1/_SUCCESS
-rw-r--r-- 3 Jig14696 supergroup    2059 2017-11-29 05:31 wcout1/part-r-00000
-rw-r--r-- 3 Jig14696 supergroup    1740 2017-11-29 05:31 wcout1/part-r-00001
Jig14696@nn:~/workspace$ hadoop fs -ls wcout1/part-r-00000
-rw-r--r-- 3 Jig14696 supergroup    2059 2017-11-29 05:31 wcout1/part-r-00000
Jig14696@nn:~/workspace$ hadoop fs -cat wcout1/part-r-00000
(Anatomy
1
(Configuring
13
(Managed
13
(Schema
13
-
24
A
34
Account
2
Anatomy
1
And
30
Architecture,
6
Archives
1
Benchmarking
2
Binary
6
Buckets,
13
Cafarella,
17
Command
1
Comparison
13
Concepts,
1
Concurrency
6
Copying
1
Custom
13
DISTCP,
1
Data!
51
Database
19
Databaseoutput)
1
Defined
13
```



```
1 Hadoop 394
2 Hadoop. 17
3 HadoopBig 45
4 HiveInstalling 12
5 HiveQL 13
6 IT 17
7 If 17
8 Importing 13
9 In 17
10 Indexes) 1
11 Indexes), 12
12 Inputs, 6
13 Installing 1
14 Interface, 2
15 It 51
16 Management, 2
17 Map/Reduce 13
18 MetaStore), 13
19 Michael 17
20 Operators 13
21 OutPut 6
22 Parallel 1
23 Partitions 13
24 Programming 6
25 Programming, 6
26 Properties, 4
27 Query), 13
28 Reducers, 6
29 Roots, 6
30 Running 13
31 Schema 13
32 Settings, 2
33 Setup 2
34 Solution 17
35 Configuration 2
```

For wc\_defaultreducer.jar questions(6,7)

## Question 6



```
yellow 17
jig14696@nn:~/workspace$ hadoop jar wc_defaultreducer.jar WordCount wcin/wordcountproblem wcout2
17/11/29 05:47:19 INFO client.RMProxy: Connecting to ResourceManager at nn.cdh2017.com/172.31.35.49:8032
17/11/29 05:47:19 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
17/11/29 05:47:19 INFO input.FileInputFormat: Total input paths to process : 1
17/11/29 05:47:19 INFO mapreduce.JobSubmitter: number of splits:1
17/11/29 05:47:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1509425937783_1885
17/11/29 05:47:19 INFO impl.YarnClientImpl: Submitted application application_1509425937783_1885
17/11/29 05:47:19 INFO mapreduce.Job: The url to track the job: http://nn.cdh2017.com:8088/proxy/application_1509425937783_1885/
17/11/29 05:47:19 INFO mapreduce.Job: Running job: job_1509425937783_1885
17/11/29 05:47:27 INFO mapreduce.Job: Job job_1509425937783_1885 running in uber mode : false
17/11/29 05:47:27 INFO mapreduce.Job: map 0% reduce 0%
17/11/29 05:47:34 INFO mapreduce.Job: map 100% reduce 0%
17/11/29 05:47:45 INFO mapreduce.Job: map 100% reduce 100%
17/11/29 05:47:45 INFO mapreduce.Job: Job job_1509425937783_1885 completed successfully
17/11/29 05:47:45 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=8346
  FILE: Number of bytes written=385118
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=48963
  HDFS: Number of bytes written=63283
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Job Counters
  Launched map tasks=1
  Launched reduce tasks=2
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=5249
  Total time spent by all reduces in occupied slots (ms)=16947
  Total time spent by all map tasks (ms)=5249
```

## Question 7

```
Jigsaw LMS x Jig14696@nn: ~/workspace x Quiz: Case Study : Map F x
Secure | https://bigdatalab.jigsawacademy.net/shell/

Reduce output records=7352
Spilled Records=14704
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=217
CPU time spent (ms)=4040
Physical memory (bytes) snapshot=810778624
Virtual memory (bytes) snapshot=4162191360
Total committed heap usage (bytes)=747110400

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=48836
File Output Format Counters
Bytes Written=63283
Jig14696@nn:~/workspace$ hadoop fs -ls
Found 5 items
drwxr-xr-x - Jig14696 supergroup 0 2017-11-29 05:47 .staging
drwxr-xr-x - Jig14696 supergroup 0 2017-11-29 05:00 wcin
drwxr-xr-x - Jig14696 supergroup 0 2017-11-29 05:21 wcout
drwxr-xr-x - Jig14696 supergroup 0 2017-11-29 05:31 wcout1
drwxr-xr-x - Jig14696 supergroup 0 2017-11-29 05:47 wcout2
Jig14696@nn:~/workspace$ hadoop fs -ls wcout2
Found 3 items
-rw-r--r-- 3 Jig14696 supergroup 0 2017-11-29 05:47 wcout2/_SUCCESS
-rw-r--r-- 3 Jig14696 supergroup 38554 2017-11-29 05:47 wcout2/part-r-00000
-rw-r--r-- 3 Jig14696 supergroup 24729 2017-11-29 05:47 wcout2/part-r-00001
Jig14696@nn:~/workspace$
```

For wc.jar questions(8,9)

Question 9:

```
Jigsaw LMS x Jig14696@nn: ~/workspace x Quiz: Case Study : Map F x
Secure | https://bigdatalab.jigsawacademy.net/shell/

Jig14696@nn:~/workspace$ hadoop jar wc.jar WordCount wcin/wordcountproblem wcout3
17/11/29 05:51:43 INFO client.RMPProxy: Connecting to ResourceManager at nn.cdh2017.com/172.31.35.49:8032
17/11/29 05:51:43 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
17/11/29 05:51:44 INFO input.FileInputFormat: Total input paths to process : 1
17/11/29 05:51:44 INFO mapreduce.JobSubmitter: number of splits:1
17/11/29 05:51:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1509425937783_1886
17/11/29 05:51:44 INFO impl.YarnClientImpl: Submitted application application_1509425937783_1886
17/11/29 05:51:44 INFO mapreduce.Job: The url to track the job: http://nn.cdh2017.com:8088/proxy/application_1509425937783_1886/
17/11/29 05:51:44 INFO mapreduce.Job: Running job: job_1509425937783_1886
17/11/29 05:51:50 INFO mapreduce.Job: Job job_1509425937783_1886 running in uber mode : false
17/11/29 05:51:50 INFO mapreduce.Job: map 0% reduce 0%
17/11/29 05:51:56 INFO mapreduce.Job: map 100% reduce 0%
17/11/29 05:52:04 INFO mapreduce.Job: map 100% reduce 100%
17/11/29 05:52:05 INFO mapreduce.Job: Job job_1509425937783_1886 completed successfully
17/11/29 05:52:05 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=8346
FILE: Number of bytes written=385538
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=48963
HDFS: Number of bytes written=3799
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Job Counters
Launched map tasks=1
Launched reduce tasks=2
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4380
Total time spent by all reduces in occupied slots (ms)=8673
Total time spent by all map tasks (ms)=4380
Total time spent by all reduce tasks (ms)=8673
```

```
Jigsaw LMS x Jig14696@nnc: ~/worksp... Quiz: Case Study: Map...
Secure | https://bigdatalab.jigsawacademy.net/shell/

Launched map tasks=1
Launched reduce tasks=2
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4380
Total time spent by all reduces in occupied slots (ms)=8673
Total time spent by all map tasks (ms)=4380
Total time spent by all reduce tasks (ms)=8673
Total vcore-seconds taken by all map tasks=4380
Total vcore-seconds taken by all reduce tasks=8673
Total megabyte-seconds taken by all map tasks=4485120
Total megabyte-seconds taken by all reduce tasks=8881152

Map-Reduce Framework
  Map input records=103
  Map output records=7352
  Map output bytes=77987
  Map output materialized bytes=8338
  Input split bytes=127
  Combine input records=0
  Combine output records=0
  Reduce input groups=351
  Reduce shuffle bytes=8338
  Reduce input records=7352
  Reduce output records=351
  Spilled Records=14704
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=130
  CPU time spent (ms)=3640
  Physical memory (bytes) snapshot=845230080
  Virtual memory (bytes) snapshot=4171251712
  Total committed heap usage (bytes)=881328128

Shuffle Errors
BAD_ID=0
CONNECTION=0
```

## Question 10:

```
Jigsaw LMS x Jig14696@nnc: ~/worksp... Quiz: Case Study: Map...
Secure | https://bigdatalab.jigsawacademy.net/shell/

BIG 51
Big 1
Big 57
BinaryOutput, 6
Cluster 10
ClusterCluster 1
Coherency 1
Combiners, 6
Concept 24
Configuration 2
Configuration, 2
Creation), 2
Cutting 17
DATA 17
Daemon 4
Data 118
Data, 13
Data/Code 6
Databaseoutput)Hadoop 5
Design 1
Error 6
Execution 6
Fault 6
File 9
Flow 1
Formats 12
Formats, 13
Framework 6
Functions), 13
Hadoop! 17
Handling), 6
Hive 40
Hive, 26
However 17
Imperative 6
```