

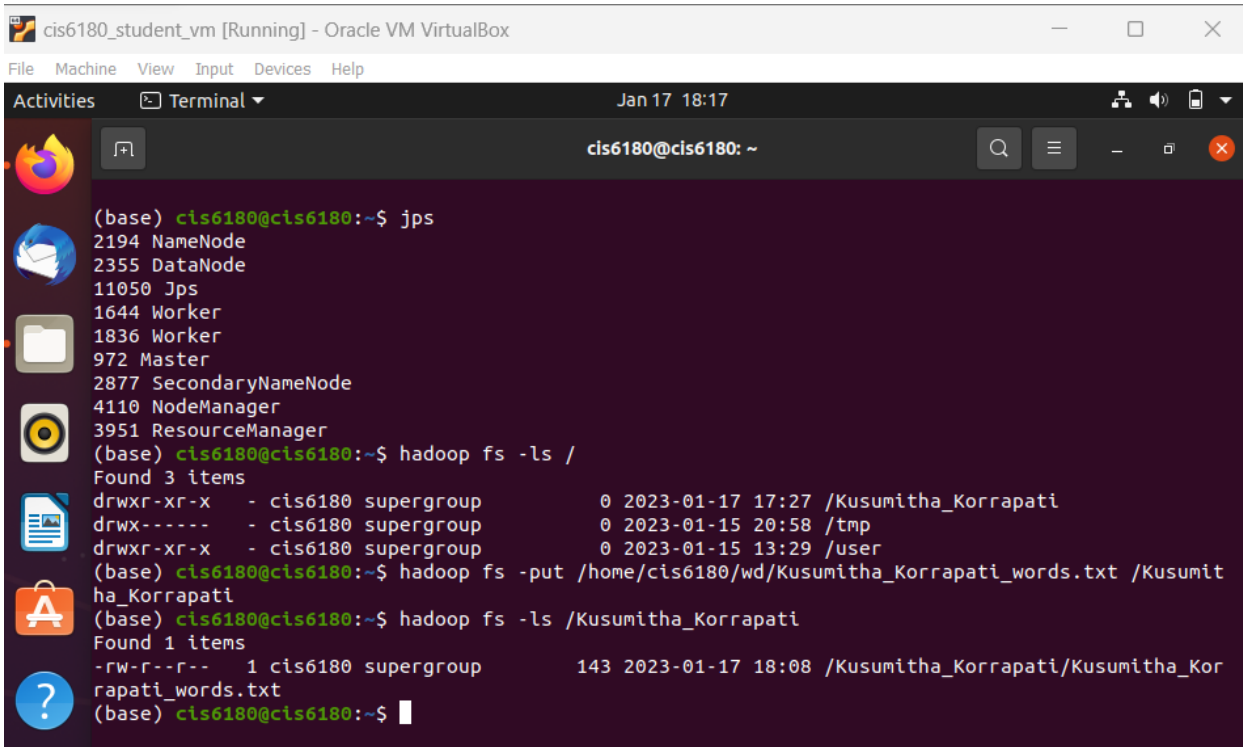
# Assignment 1 - CIS\*6180/DATA\*6300

Student ID – 1238693

Kusumitha Korrapati

1. Make a text file and copy into Hadoop HDFS [1 point]

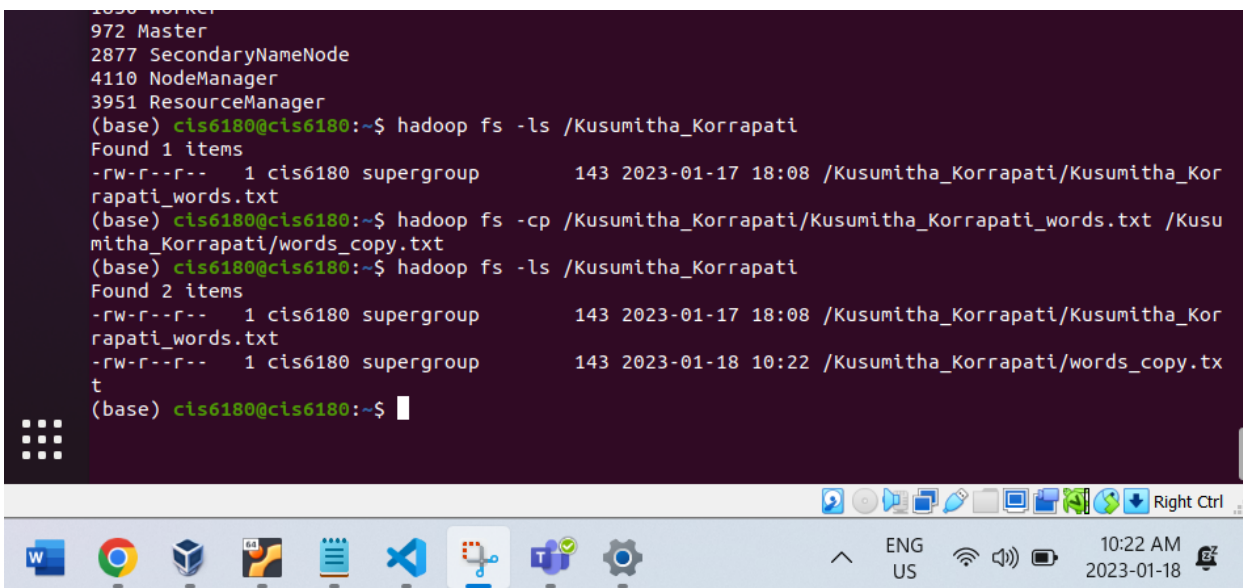
verify the copied file by using the ls command in the hdfs directory, take a screenshot and save it as directory\_file.png.



```
cis6180_student_vm [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Jan 17 18:17
cis6180@cis6180: ~
(base) cis6180@cis6180:~$ jps
2194 NameNode
2355 DataNode
11050 Jps
1644 Worker
1836 Worker
972 Master
2877 SecondaryNameNode
4110 NodeManager
3951 ResourceManager
(base) cis6180@cis6180:~$ hadoop fs -ls /
Found 3 items
drwxr-xr-x - cis6180 supergroup 0 2023-01-17 17:27 /Kusumitha_Korrapati
drwx----- - cis6180 supergroup 0 2023-01-15 20:58 /tmp
drwxr-xr-x - cis6180 supergroup 0 2023-01-15 13:29 /user
(base) cis6180@cis6180:~$ hadoop fs -put /home/cis6180/wd/Kusumitha_Korrapati_words.txt /Kusumitha_Korrapati
(base) cis6180@cis6180:~$ hadoop fs -ls /Kusumitha_Korrapati
Found 1 items
-rw-r--r-- 1 cis6180 supergroup 143 2023-01-17 18:08 /Kusumitha_Korrapati/Kusumitha_Korrapati_words.txt
(base) cis6180@cis6180:~$
```

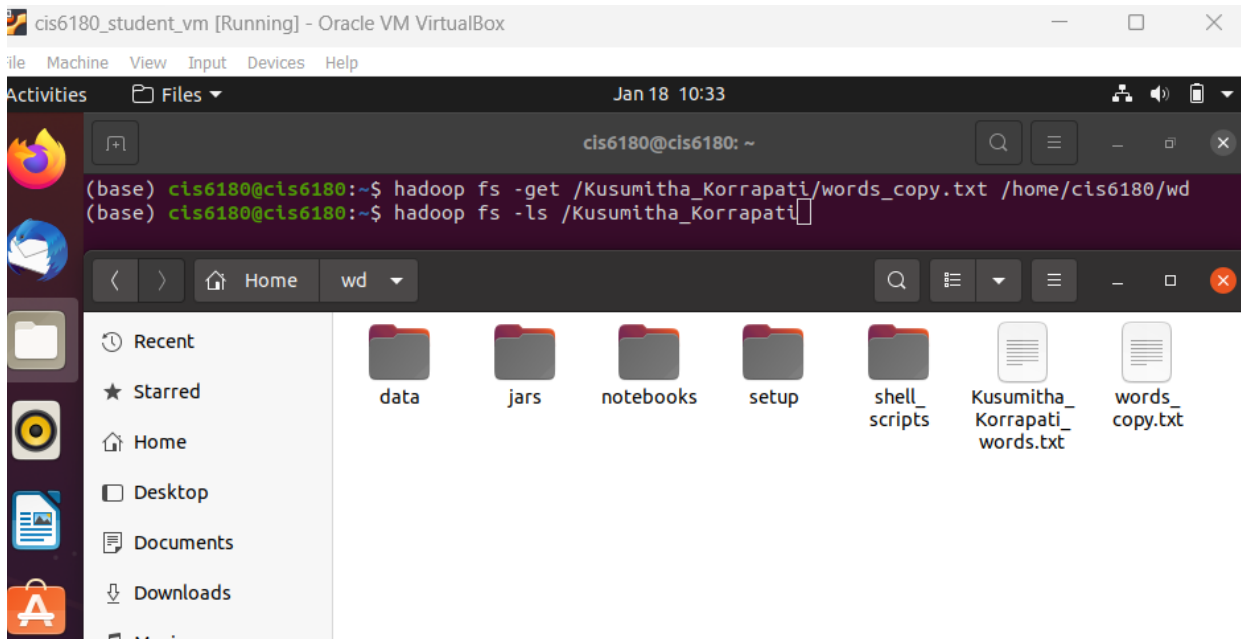
2. Copy file in HDFS [1 point]

verify the copied file by using the ls command in the hdfs directory, take a screenshot and save it as directory\_file\_copy.png.

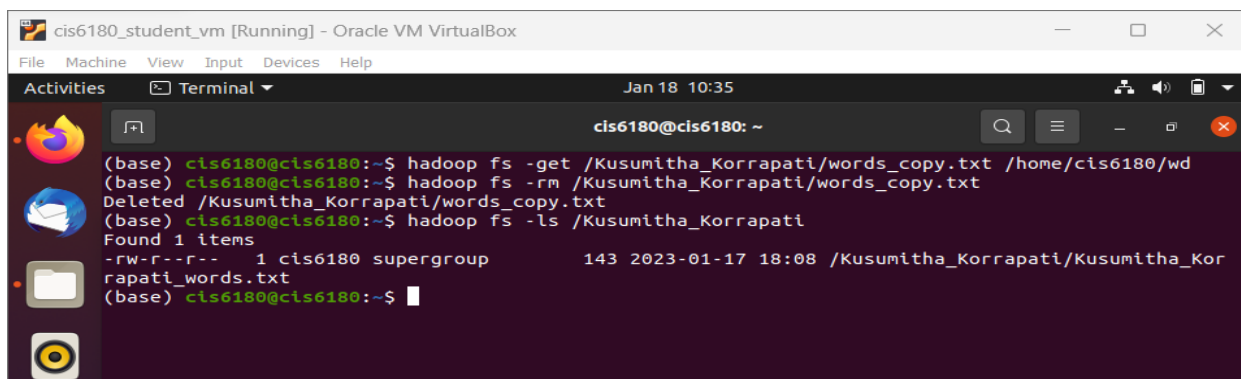


```
1836 Worker
972 Master
2877 SecondaryNameNode
4110 NodeManager
3951 ResourceManager
(base) cis6180@cis6180:~$ hadoop fs -ls /Kusumitha_Korrapati
Found 1 items
-rw-r--r-- 1 cis6180 supergroup 143 2023-01-17 18:08 /Kusumitha_Korrapati/Kusumitha_Korrapati_words.txt
(base) cis6180@cis6180:~$ hadoop fs -cp /Kusumitha_Korrapati/Kusumitha_Korrapati_words.txt /Kusumitha_Korrapati/words_copy.txt
(base) cis6180@cis6180:~$ hadoop fs -ls /Kusumitha_Korrapati
Found 2 items
-rw-r--r-- 1 cis6180 supergroup 143 2023-01-17 18:08 /Kusumitha_Korrapati/Kusumitha_Korrapati_words.txt
-rw-r--r-- 1 cis6180 supergroup 143 2023-01-18 10:22 /Kusumitha_Korrapati/words_copy.txt
(base) cis6180@cis6180:~$
```

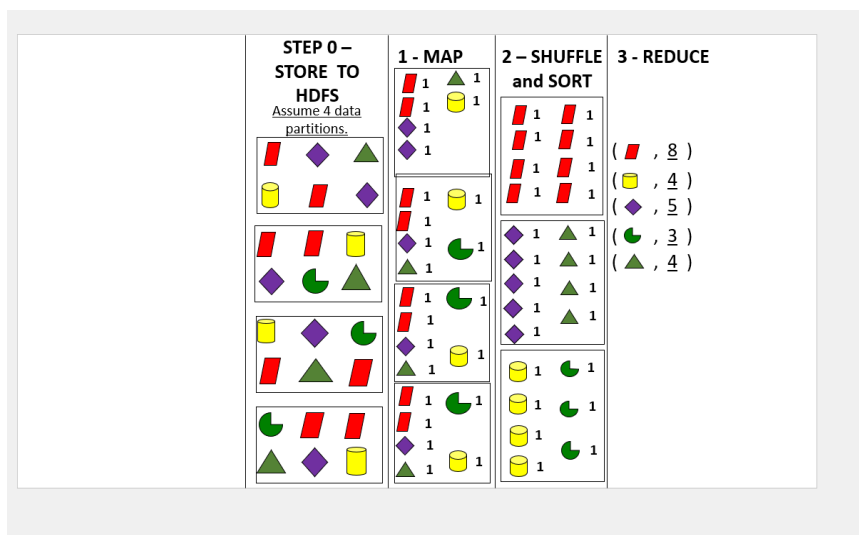
3. Copy file from HDFS [1 point]



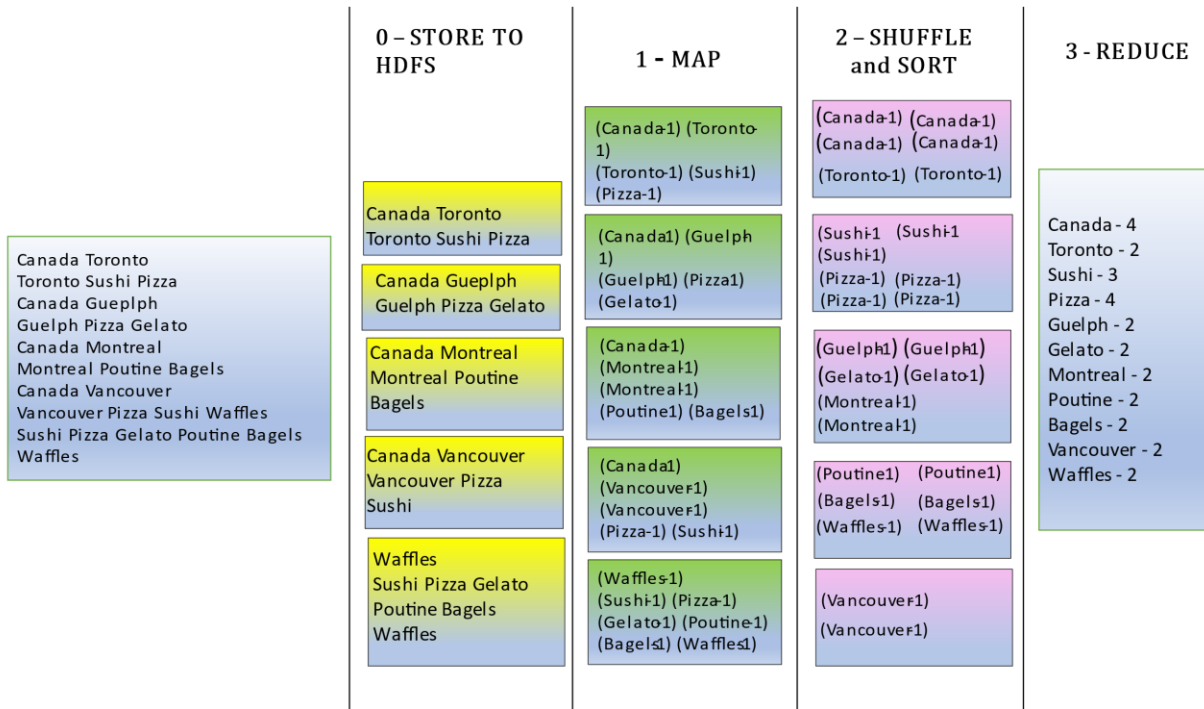
4. Delete file from HDFS [1 point]  
verify the deleted file by using the ls command in the hdfs directory, take a screenshot and save it as directory\_file\_delete.png.



5. MapReduce practice [2 points]

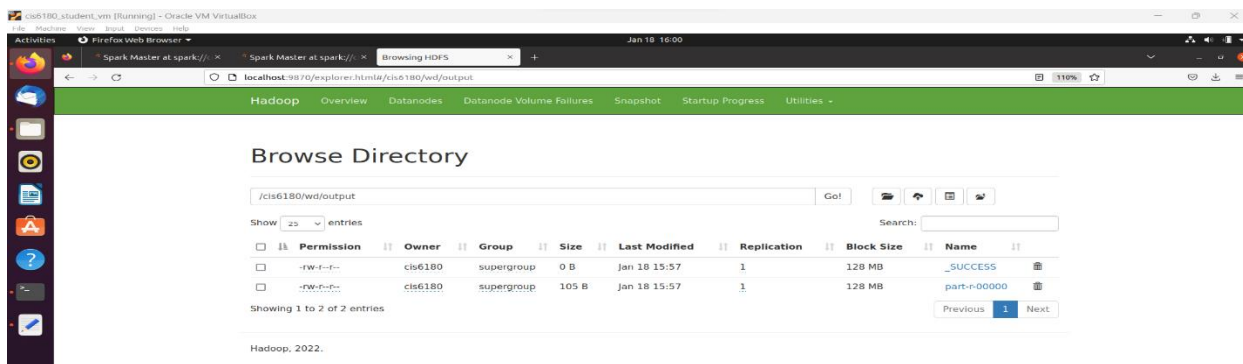


6. MapReduce word count [2 points]  
Apply the MapReduce manually on the following text.



1. Store the words into HDFS partitions
2. Map the words
3. Shuffle and sort
4. Reduce and write the sum for each item.

7. MapReduce word count and word mean using Hadoop [2 points]



To compare the text file with manually done map reduce, different text file is added to the Hadoop directory and those results are compared below

AutoSave On Kuru... Saved to this... Kuru... Kuru...

File Home Insert Draw Design Transition Animation Slide Show Record Review View Help

Undo Clipboard FS Slides Font Paragraph Drawing Editing Dictate Designer

1

0 - STORE TO HDFS	1 - MAP	2 - SHUFFLE and SORT	3 - REDUCE
Canada Toronto Toronto Sushi Pizza Canada Guelph Guelph Pizza Guelph Canada Montreal Montreal Poutine Bagels Canada Vancouver Vancouver Pizza Sushi Waffles Sushi Pizza Guelph Poutine Bagels Waffles	Canada-1 (Toronto-1) Toronto-1 (Sushi-1) Pizza-1 Canada-1 (Guelph-1) Guelph-1 (Pizza-1) Guelph-1 Canada-1 (Montreal-1) Montreal-1 (Poutine-1) Poutine-1 (Bagels-1) Canada-1 (Vancouver-1) Vancouver-1 (Sushi-1) Waffles-1 Sushi-1 (Pizza-1) Guelph-1 (Poutine-1) Bagels-1 (Waffles-1)	Canada-1 (Canada-1) Canada-1 (Canada-1) Toronto-1 (Toronto-1) Sushi-1 (Sushi-1) Pizza-1 (Pizza-1) Guelph-1 (Guelph-1) Poutine-1 (Poutine-1) Bagels-1 (Bagels-1) Waffles-1 (Waffles-1) Canada-1 (Canada-1) Toronto-1 (Toronto-1) Sushi-1 (Sushi-1) Pizza-1 (Pizza-1) Guelph-1 (Guelph-1) Poutine-1 (Poutine-1) Bagels-1 (Bagels-1) Waffles-1 (Waffles-1)	Canada-1 Toronto-1 Sushi-1 Pizza-1 Guelph-1 Poutine-1 Bagels-1 Waffles-1

Slide 1 of 1 English (Canada) Accessibility: Investigate Notes 47%

Activities Text Editor Jan 18 15:59

part-r-00000 Save

\*words\_copy (copy).txt part-r-00000

```
1 Bagels 2
2 Canada 4
3 Gelato 2
4 Guelph 2
5 Montreal 2
6 Pizza 4
7 Poutine 2
8 Sushi 3
9 Toronto 2
10 Vancouver 2
11 Waffles 2
```

Plain Text Tab Width: 8 Ln 1, Col 1 INS

cis6180\_student\_vm [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Terminal Jan 21 20:31

cis6180@cis6180: ~

```
(base) cis6180@cis6180:~$ hadoop fs -ls /cis6180/wd/output
Found 2 items
-rw-r--r-- 1 cis6180 supergroup 0 2023-01-18 15:57 /cis6180/wd/output/_SUCCESS
-rw-r--r-- 1 cis6180 supergroup 105 2023-01-18 15:57 /cis6180/wd/output/part-r-00000
(base) cis6180@cis6180:~$
```

Word count mean –

cis6180\_student\_vm [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Text Editor Jan 21 21:10

part-r-00000(1) ~/Downloads

```
1 count 27
2 length 171
```

File output Format Counters

```
Bytes Written:20
The mean is: 0.3333333333333333
(base) cis6180@cis6180:~$ hadoop fs -ls /wcout_mean
Found 2 items
-rw-r--r-- 1 cis6180 supergroup 0 2023-01-21 20:46 /wcout_mean/_SUCCESS
-rw-r--r-- 1 cis6180 supergroup 20 2023-01-21 20:46 /wcout_mean/part-r-00000
(base) cis6180@cis6180:~$
```

```
Activities Terminal Jan 21 21:13 cis6180@cis6180: ~

at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.base/java.lang.reflect.Method.invoke(Method.java:566)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
Caused by: java.io.IOException: Input path does not exist: hdfs://localhost:8000/user/cis6180/Kusunitha.Korrapati/words_map_reduce.txt
at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:313)
... 28 more
(base) cis6180@cis6180: ~$ hadoop jar /home/cis6180/udf/jars/hadoop-mapreduce-examples-3.3.4.jar wordmean /Kusunitha.Korrapati/words_map_reduce.txt /wcout_mean
2023-01-21 20:46:03.311 INFO client.DefaultHadoopHwProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-01-21 20:46:03.624 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/cis6180/.staging/job_1674351037203_0003
2023-01-21 20:46:03.830 INFO input.FileInputFormat: Total input files to process : 1
2023-01-21 20:46:03.950 INFO mapreduce.JobSubmitter: number of splits:1
2023-01-21 20:46:04.201 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1674351037203_0003
2023-01-21 20:46:04.201 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-01-21 20:46:04.379 INFO conf.Configuration: resource-types.xml not found
2023-01-21 20:46:04.380 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-01-21 20:46:04.646 INFO impl.YarnClientImpl: Submitted application application_1674351037203_0003
2023-01-21 20:46:04.688 INFO mapreduce.Job: The url to track the job: http://cis6180.myquest.virtualbox.org:8088/proxy/application_1674351037203_0003/
2023-01-21 20:46:04.689 INFO mapreduce.Job: Running job: job_1674351037203_0003
2023-01-21 20:46:11.810 INFO mapreduce.Job: Job job_1674351037203_0003 running in uber mode : false
2023-01-21 20:46:11.811 INFO mapreduce.Job: map 0% reduce 0%
2023-01-21 20:46:15.064 INFO mapreduce.Job: map 100% reduce 0%
2023-01-21 20:46:20.919 INFO mapreduce.Job: map 100% reduce 100%
2023-01-21 20:46:20.931 INFO mapreduce.Job: Job job_1674351037203_0003 completed successfully
2023-01-21 20:46:21.004 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=52797
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=225
  HDFS: Number of bytes written=20
  HDFS: Number of read operations=8
  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)=2260
  Total time spent by all reduces in occupied slots (ms)=1928
  Total time spent by all map tasks (ms)=2260
  Total time spent by all reduce tasks (ms)=1928
  Total vcore-millisecs taken by all map tasks=2260
  Total vcore-millisecs taken by all reduce tasks=1928
```