

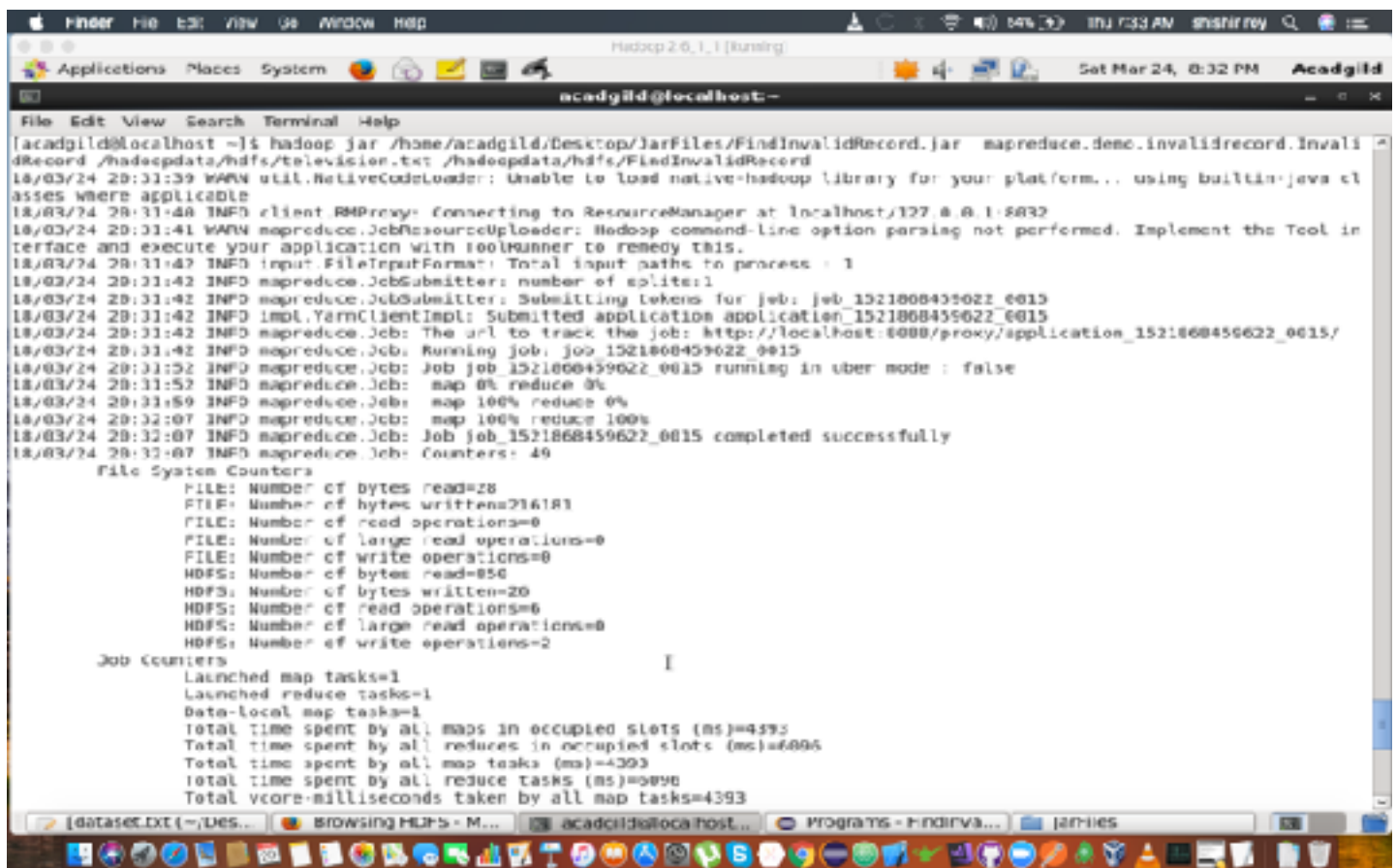
Task 1:

Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Created Jar file for executing Invalid Records

```
hadoop jar /home/acadgild/Desktop/JarFiles/FindInvalidRecord.jar  
mapreduce.demo.invalidrecord.InvalidRecord /hadoopdata/hdfs/television.txt /  
hadoopdata/hdfs/FindInvalidRecord
```

Find Result below:



```
acadmild@localhost:~$ hadoop jar /home/acadgild/Desktop/JarFiles/FindInvalidRecord.jar mapreduce.demo.invalidrecord.InvalidRecord /hadoopdata/hdfs/television.txt /hadoopdata/hdfs/FindInvalidRecord
18/03/24 20:31:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/24 20:31:40 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/24 20:31:41 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/24 20:31:43 INFO input.FileInputFormat: Total input paths to process : 1
18/03/24 20:31:43 INFO mapreduce.JobSubmitter: number of splits=1
18/03/24 20:31:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521868459622_0015
18/03/24 20:31:43 INFO impl.YarnClientImpl: Submitted application application_1521868459622_0015
18/03/24 20:31:43 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1521868459622_0015/
18/03/24 20:31:43 INFO mapreduce.Job: Running job: job_1521868459622_0015
18/03/24 20:31:52 INFO mapreduce.Job: Job job_1521868459622_0015 running in uber mode : false
18/03/24 20:31:52 INFO mapreduce.Job: map 0% reduce 0%
18/03/24 20:31:59 INFO mapreduce.Job: map 100% reduce 0%
18/03/24 20:32:07 INFO mapreduce.Job: map 100% reduce 100%
18/03/24 20:32:07 INFO mapreduce.Job: Job job_1521868459622_0015 completed successfully
18/03/24 20:32:07 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=28
  FILE: Number of bytes written=216181
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=856
  HDFS: Number of bytes written=20
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

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)=4393
  Total time spent by all reduces in occupied slots (ms)=6895
  Total time spent by all map tasks (ms)=4393
  Total time spent by all reduce tasks (ms)=6895
  Total vcore-milliseconds taken by all map tasks=4393
```

```
acadgild@localhost:~$ cat /dev/null > /tmp/hadoop-hdfs/FindInvalidRecord/part-r-00000
acadgild@localhost:~$ hadoop fs -cat /tmp/hadoop-hdfs/FindInvalidRecord/part-r-00000
18/03/24 20:33:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
NA nullLucid
00100 nullNA
acadgild@localhost:~$
```

Map-Reduce Framework
Total time spent by all map tasks (ms)=4393
Total time spent by all reduce tasks (ms)=6095
Total vcore-milliseconds taken by all map tasks=4393
Total vcore-milliseconds taken by all reduce tasks=6095
Total megabyte-milliseconds taken by all map tasks=4498452
Total megabyte-milliseconds taken by all reduce tasks=6242394
Map input records=10
Map output records=2
Map output bytes=18
Map output materialized bytes=28
Input split bytes=137
Combine input records=0
Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=28
Reduce input records=2
Reduce output records=2
Spilled Records=4
Shuffled Hops =1
Failed shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=127
CPU time spent (ms)=1480
Physical memory (bytes) snapshot=292953888
Virtual memory (bytes) snapshot=4118192128
Total committed heap usage (bytes)=170604480

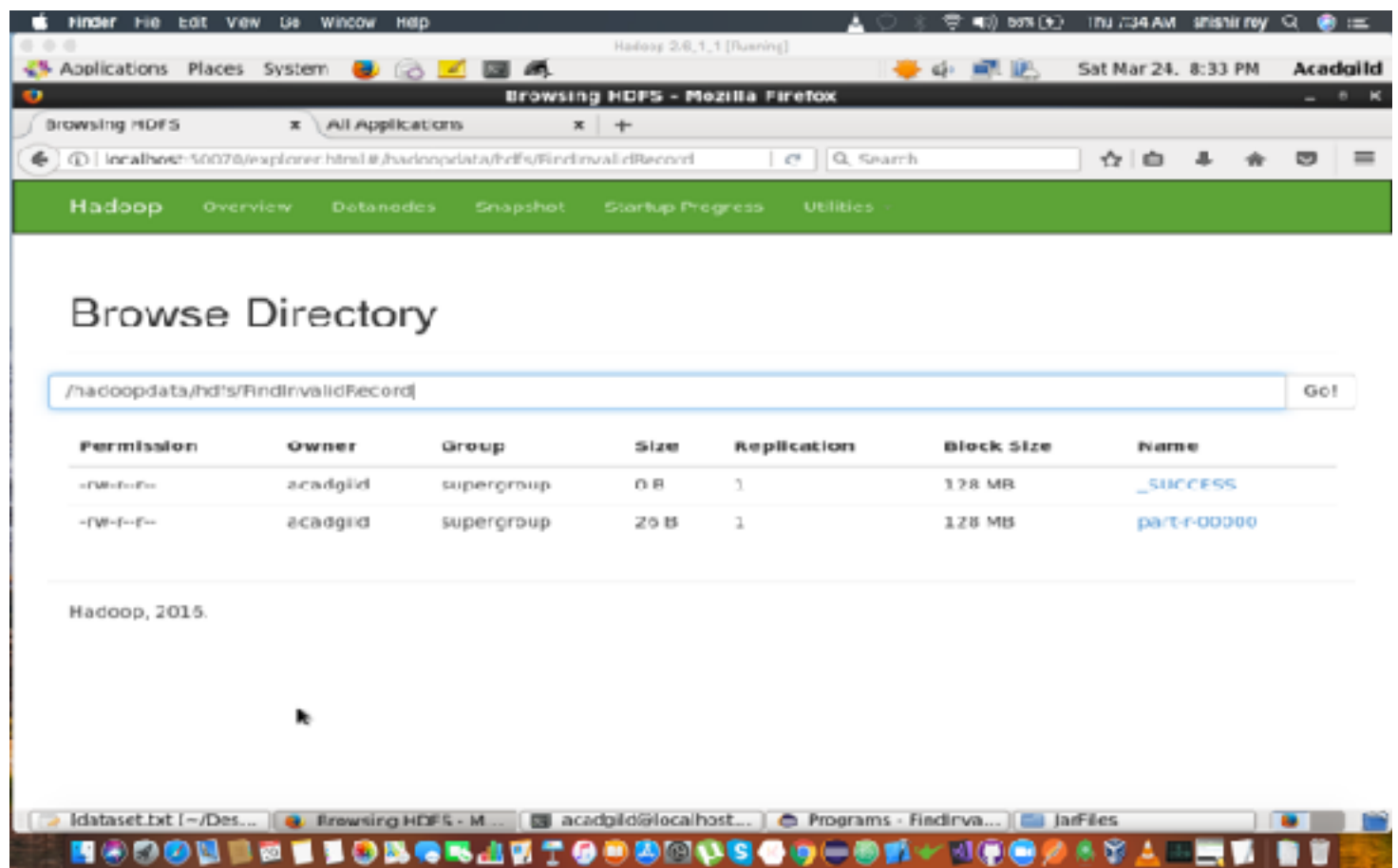
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=733

File Output Format Counters
Bytes Written=26

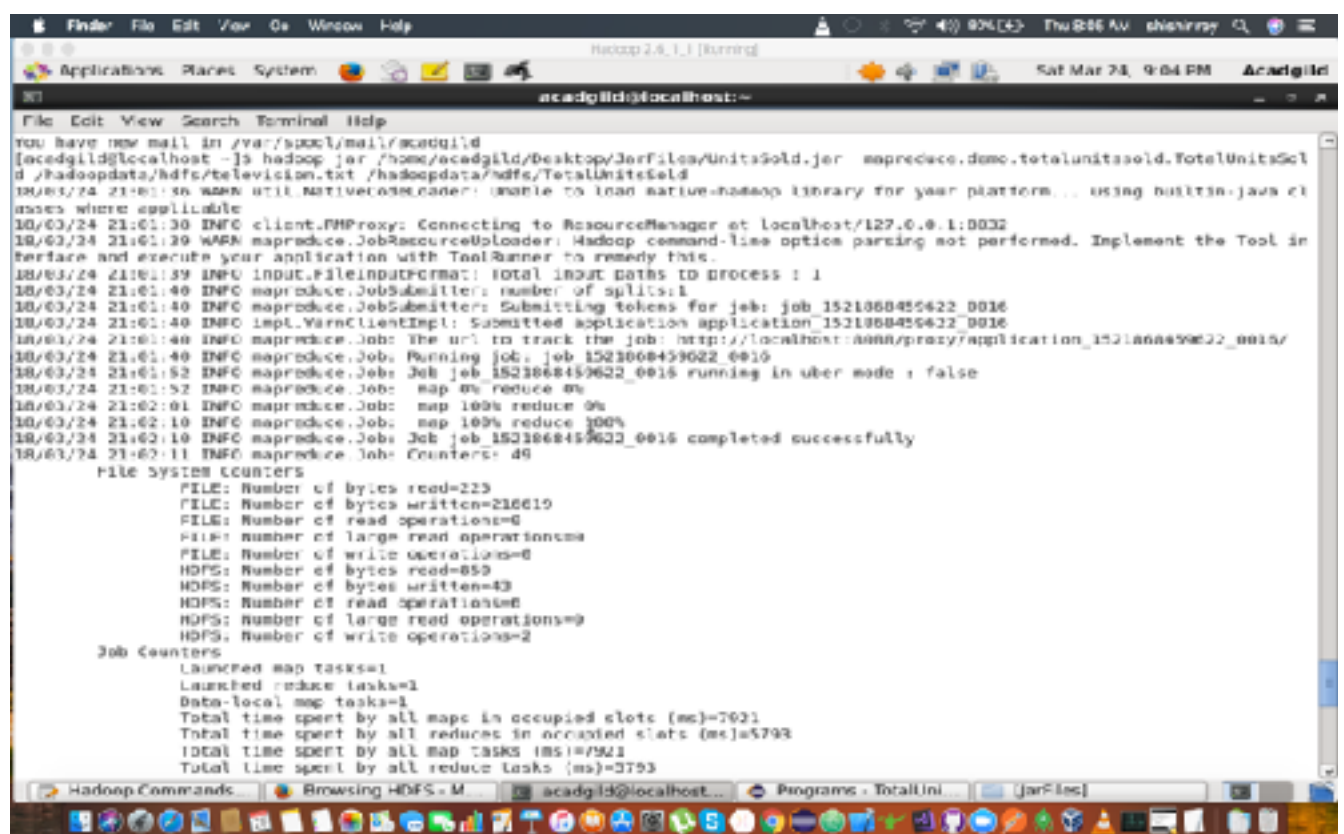
```
acadgild@localhost:~$
```

```
acadgild@localhost:~$ hadoop fs -cat /tmp/hadoop-hdfs/FindInvalidRecord/part-r-00000
18/03/24 20:33:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
NA nullLucid
00100 nullNA
acadgild@localhost:~$
```



Task 2:

Write a Map Reduce program to calculate the total units sold for each Company.



```
Finder  File  Edit  View  Go  Window  Help
Hadoop 2.8.1_1 [Running]
Applications  Places  System
acadmild@localhost:~$
File  Edit  View  Search  terminal  Help
Map output materialized bytes=225
Input split bytes=117
Combine input records=0
Combine output records=0
Reduce input groups=6
Reduce shuffle bytes=225
Reduce input records=18
Reduce output records=6
Spilled Records=36
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=145
CPU time spent (ms)=1790
Physical memory (bytes) snapshot=263746304
Virtual memory (bytes) snapshot=4117963400
Total committed heap usage (bytes)=170004400

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=731
File Output Format Counters
Bytes Written=43

You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/TotalUnitsSold/part-r00000
18/03/24 21:03:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Akai 1
Lava 3
NA 1
Onida 4
Samsung 7
Zen 2
You have new mail in /var/spool/mail/acadmild
```

Finder File Edit View Go Window Help

Applications Places System

acadmild@localhost:~\$

Hadoop 2.8.1_1 [Running]

Sat Mar 24, 9:05 PM

Acadmild

Browsing HDFS - Mozilla Firefox

localhost:50070/explorer.html#/hadoopdata/hdfs/TotalUnitsSold

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/hadoopdata/hdfs/TotalUnitsSold

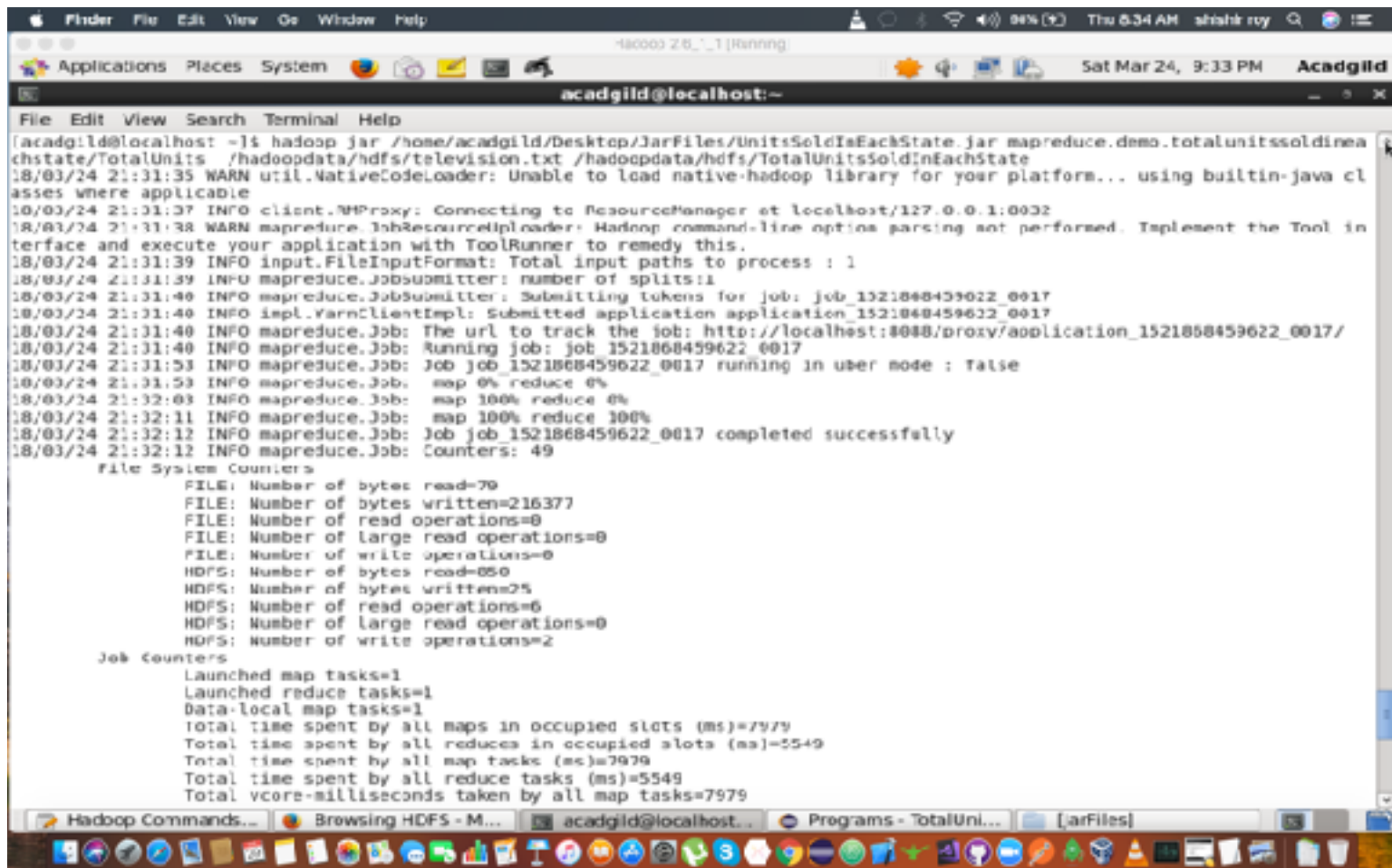
Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	acadmild	supergroup	0 B	1	128 MB	_SUCCESS
-rw-r--r--	acadmild	supergroup	43 B	1	128 MB	part-r00000

Hadoop, 2016

Task 3:

Write a Map Reduce program to calculate the total units sold in each state for Onida company.



```
acacgild@localhost:~$ hadoop jar /home/acacgild/Desktop/JarFiles/UnitsSoldInEachState.jar mapreduce.demo.totalunitssoldineachstate/TotalUnits /hadoopdata/hdfs/television.txt /hadoopdata/hdfs/TotalUnitsSoldInEachState
18/03/24 21:31:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/24 21:31:37 INFO client.NMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/24 21:31:38 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/24 21:31:39 INFO input.FileInputFormat: Total input paths to process : 1
18/03/24 21:31:39 INFO mapreduce.JobSubmitter: number of splits:1
18/03/24 21:31:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521868459622_0017
18/03/24 21:31:40 INFO impl.YarnClientImpl: Submitted application application_1521868459622_0017
18/03/24 21:31:40 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1521868459622_0017/
18/03/24 21:31:40 INFO mapreduce.Job: Running job: job_1521868459622_0017
18/03/24 21:31:53 INFO mapreduce.Job: Job job_1521868459622_0017 running in uber mode : false
18/03/24 21:31:53 INFO mapreduce.Job:  map 0% reduce 0%
18/03/24 21:32:03 INFO mapreduce.Job:  map 100% reduce 0%
18/03/24 21:32:11 INFO mapreduce.Job:  map 100% reduce 100%
18/03/24 21:32:12 INFO mapreduce.Job: Job job_1521868459622_0017 completed successfully
18/03/24 21:32:12 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=70
    FILE: Number of bytes written=216377
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=850
    HDFS: Number of bytes written=25
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  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)=7979
    Total time spent by all reduces in occupied slots (ms)=5549
    Total time spent by all map tasks (ms)=7979
    Total time spent by all reduce tasks (ms)=5549
    Total vcore-milliseconds taken by all map tasks=7979
```