

Mapreduce

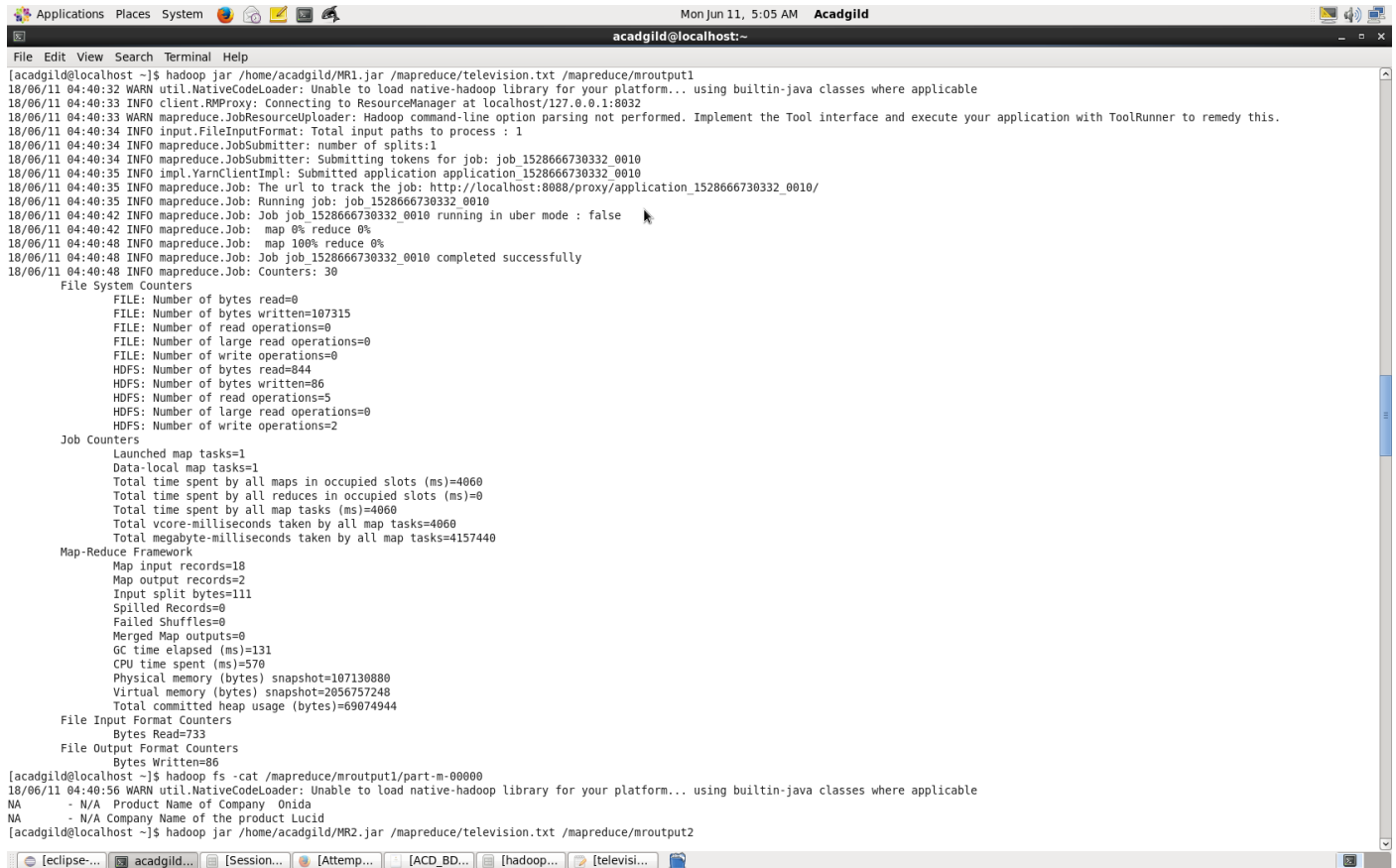
I have written Mapreduce programs according to requirement, mentioned o/p format

& i/p format classes, created jar files & executed as below.

Attaching Java programs & jar files in Github for reference

Task 1:

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



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/MR1.jar /mapreduce/television.txt /mapreduce/mroutput1
18/06/11 04:40:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/11 04:40:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/11 04:40:34 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/11 04:40:34 INFO mapreduce.JobSubmitter: number of splits:1
18/06/11 04:40:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528666730332_0010
18/06/11 04:40:35 INFO impl.YarnClientImpl: Submitted application application_1528666730332_0010
18/06/11 04:40:35 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528666730332_0010/
18/06/11 04:40:35 INFO mapreduce.Job: Running job: job_1528666730332_0010
18/06/11 04:40:42 INFO mapreduce.Job: Job job_1528666730332_0010 running in uber mode : false
18/06/11 04:40:42 INFO mapreduce.Job: map 0% reduce 0%
18/06/11 04:40:48 INFO mapreduce.Job: map 100% reduce 0%
18/06/11 04:40:48 INFO mapreduce.Job: Job job_1528666730332_0010 completed successfully
18/06/11 04:40:48 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=107315
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=844
    HDFS: Number of bytes written=86
    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)=4060
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=4060
    Total vcore-milliseconds taken by all map tasks=4060
    Total megabyte-milliseconds taken by all map tasks=4157440
  Map-Reduce Framework
    Map input records=18
    Map output records=2
    Input split bytes=111
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=131
    CPU time spent (ms)=570
    Physical memory (bytes) snapshot=107130880
    Virtual memory (bytes) snapshot=2056757248
    Total committed heap usage (bytes)=69074944
  File Input Format Counters
    Bytes Read=733
  File Output Format Counters
    Bytes Written=86
[acadgild@localhost ~]$ hadoop fs -cat /mapreduce/mroutput1/part-m-00000
18/06/11 04:40:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
NA      - N/A Product Name of Company Onida
NA      - N/A Company Name of the product Lucid
[acadgild@localhost ~]$ hadoop jar /home/acadgild/MR2.jar /mapreduce/television.txt /mapreduce/mroutput2
```

Mapreduce

Task 2:

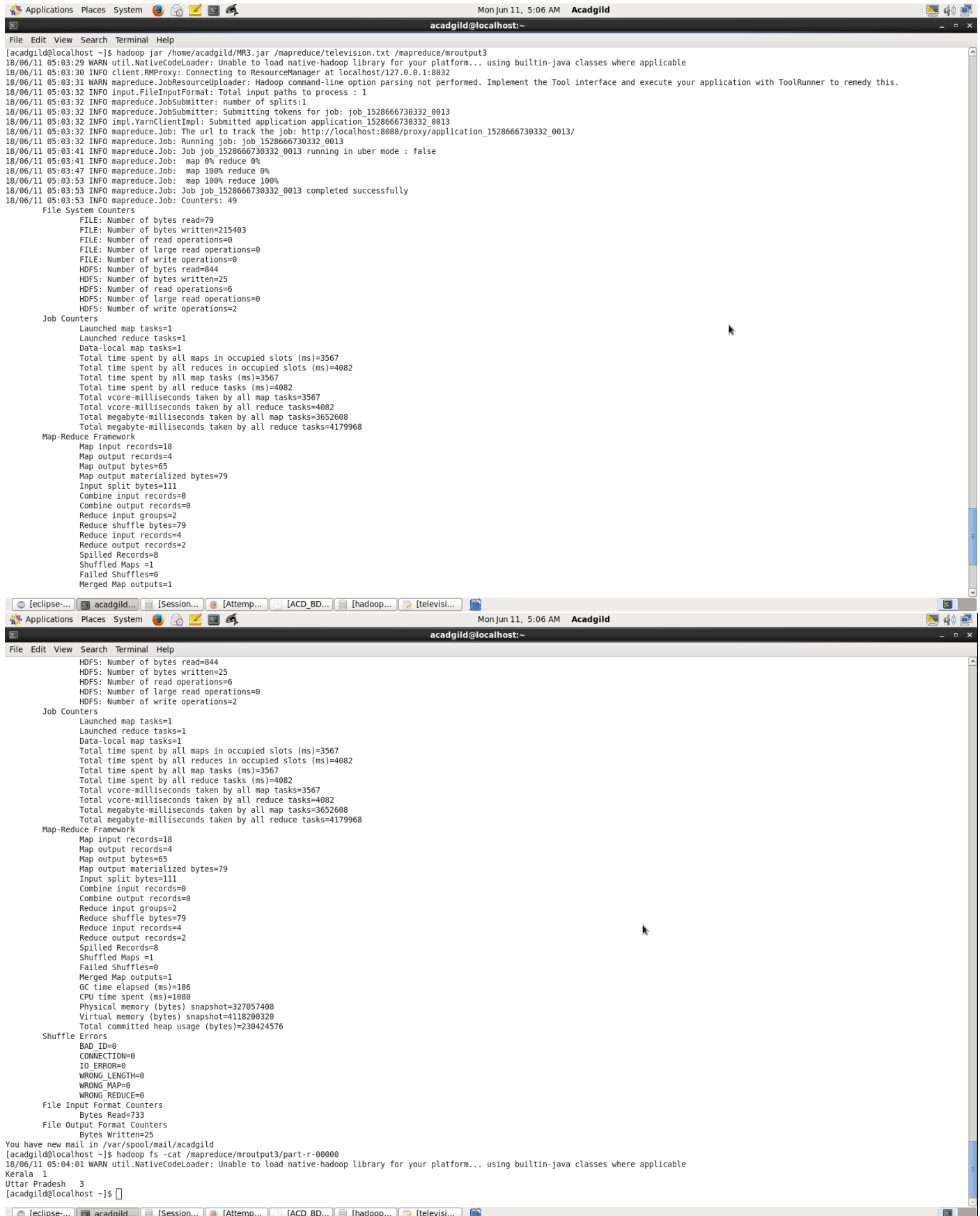
Write a Map Reduce program to calculate the total units sold for each Company.

```
Applications Places System Mon Jun 11, 5:05 AM Acadgild
acadgild@localhost:~$ hadoop jar /home/acadgild/MR2.jar /mapreduce/television.txt /mapreduce/mroutput2
18/06/11 04:52:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/11 04:52:12 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/11 04:52:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/11 04:52:13 INFO input.FileInputFormat: Total input paths to process : 1
18/06/11 04:52:13 INFO mapreduce.JobSubmitter: number of splits:1
18/06/11 04:52:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528666730332_0011
18/06/11 04:52:14 INFO impl.YarnClientImpl: Submitted application application_1528666730332_0011
18/06/11 04:52:14 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528666730332_0011/
18/06/11 04:52:14 INFO mapreduce.Job: Running job: job_1528666730332_0011
18/06/11 04:52:23 INFO mapreduce.Job: Job job_1528666730332_0011 running in uber mode : false
18/06/11 04:52:23 INFO mapreduce.Job: map 0% reduce 0%
18/06/11 04:52:30 INFO mapreduce.Job: map 100% reduce 0%
18/06/11 04:52:38 INFO mapreduce.Job: map 100% reduce 100%
18/06/11 04:52:38 INFO mapreduce.Job: Job job_1528666730332_0011 completed successfully
18/06/11 04:52:39 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=204
    FILE: Number of bytes written=215619
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=844
    HDFS: Number of bytes written=38
    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)=4819
    Total time spent by all reduces in occupied slots (ms)=4768
    Total time spent by all map tasks (ms)=4819
    Total time spent by all reduce tasks (ms)=4768
    Total vcore-millisecons taken by all map tasks=4819
    Total vcore-millisecons taken by all reduce tasks=4768
    Total megabyte-millisecons taken by all map tasks=4934656
    Total megabyte-millisecons taken by all reduce tasks=4882432
  Map-Reduce Framework
    Map input records=18
    Map output records=16
    Map output bytes=166
    Map output materialized bytes=204
    Input split bytes=111
    Combine input records=0
    Combine output records=0
    Reduce input groups=5
    Reduce shuffle bytes=204
    Reduce input records=16
    Reduce output records=5
    Spilled Records=32
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    Map output materialized bytes=204
    Input split bytes=111
    Combine input records=0
    Combine output records=0
    Reduce input groups=5
    Reduce shuffle bytes=204
    Reduce input records=16
    Reduce output records=5
    Spilled Records=32
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=318
    CPU time spent (ms)=1530
    Physical memory (bytes) snapshot=325165056
    Virtual memory (bytes) snapshot=418192128
    Total committed heap usage (bytes)=230424576
  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=38
You have new mail in /var/spool/mail/acadgild
acadgild@localhost ~$ hadoop fs -cat /mapreduce/mroutput2/part-m-00000
18/06/11 04:52:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: /mapreduce/mroutput2/part-m-00000: No such file or directory
acadgild@localhost ~$ hadoop fs -ls /mapreduce
18/06/11 04:53:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - acadgild supergroup          0 2018-06-11 04:40 /mapreduce/mroutput1
drwxr-xr-x - acadgild supergroup          0 2018-06-11 04:52 /mapreduce/mroutput2
-rw-r--r-- 1 acadgild supergroup          73 2018-06-03 23:28 /mapreduce/musicdata.txt
drwxr-xr-x - acadgild supergroup          0 2018-06-04 01:25 /mapreduce/output1
drwxr-xr-x - acadgild supergroup          0 2018-06-04 01:45 /mapreduce/output2
drwxr-xr-x - acadgild supergroup          0 2018-06-04 01:46 /mapreduce/output3
-rw-r--r-- 1 acadgild supergroup          733 2018-06-11 03:19 /mapreduce/television.txt
acadgild@localhost ~$ hadoop fs -ls /mapreduce/mroutput2
18/06/11 04:53:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-06-11 04:52 /mapreduce/mroutput2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup          38 2018-06-11 04:52 /mapreduce/mroutput2/part-r-00000
acadgild@localhost ~$ hadoop fs -cat /mapreduce/mroutput2/part-r-00000
18/06/11 04:53:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Aka1 1
Lava 3
Onida 3
Samsung 7
Zen 2
You have new mail in /var/spool/mail/acadgild
```

Mapreduce

Task 3:

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



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/MR3.jar /mapreduce/television.txt /mapreduce/mroutput3
18/06/11 05:03:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/11 05:03:30 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/11 05:03:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/11 05:03:32 INFO input.FileInputFormat: Total input paths to process : 1
18/06/11 05:03:32 INFO mapreduce.JobSubmitter: Number of splits:1
18/06/11 05:03:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528666730332_0013
18/06/11 05:03:32 INFO impl.YarnClientImpl: Submitted application application_1528666730332_0013
18/06/11 05:03:32 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528666730332_0013/
18/06/11 05:03:32 INFO mapreduce.Job: Running job: job_1528666730332_0013
18/06/11 05:03:41 INFO mapreduce.Job: Job job_1528666730332_0013 running in uber mode : false
18/06/11 05:03:41 INFO mapreduce.Job: map 0% reduce 0%
18/06/11 05:03:47 INFO mapreduce.Job: map 100% reduce 0%
18/06/11 05:03:53 INFO mapreduce.Job: map 100% reduce 100%
18/06/11 05:03:53 INFO mapreduce.Job: Job job_1528666730332_0013 completed successfully
18/06/11 05:03:53 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=79
  FILE: Number of bytes written=215403
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=844
  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)=3567
  Total time spent by all reduces in occupied slots (ms)=4082
  Total time spent by all map tasks (ms)=3567
  Total time spent by all reduce tasks (ms)=4082
  Total vcore-milliseonds taken by all map tasks=3567
  Total vcore-milliseonds taken by all reduce tasks=4082
  Total megabyte-milliseonds taken by all map tasks=3652608
  Total megabyte-milliseonds taken by all reduce tasks=4179968

Map-Reduce Framework
  Map input records=18
  Map output records=4
  Map output bytes=65
  Map output materialized bytes=79
  Input split bytes=111
  Combine input records=0
  Combine output records=0
  Reduce input groups=2
  Reduce shuffle bytes=79
  Reduce input records=4
  Reduce output records=2
  Spilled Records=8
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1

HDFS: Number of bytes read=844
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)=3567
  Total time spent by all reduces in occupied slots (ms)=4082
  Total time spent by all map tasks (ms)=3567
  Total time spent by all reduce tasks (ms)=4082
  Total vcore-milliseonds taken by all map tasks=3567
  Total vcore-milliseonds taken by all reduce tasks=4082
  Total megabyte-milliseonds taken by all map tasks=3652608
  Total megabyte-milliseonds taken by all reduce tasks=4179968

Map-Reduce Framework
  Map input records=18
  Map output records=4
  Map output bytes=65
  Map output materialized bytes=79
  Input split bytes=111
  Combine input records=0
  Combine output records=0
  Reduce input groups=2
  Reduce shuffle bytes=79
  Reduce input records=4
  Reduce output records=2
  Spilled Records=8
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=106
  CPU time spent (ms)=1080
  Physical memory (bytes) snapshot=327057408
  Virtual memory (bytes) snapshot=418208320
  Total committed heap usage (bytes)=230424576

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=25

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /mapreduce/mroutput3/part-r-00000
Kerala 1
Uttar Pradesh 3
[acadgild@localhost ~]$
```