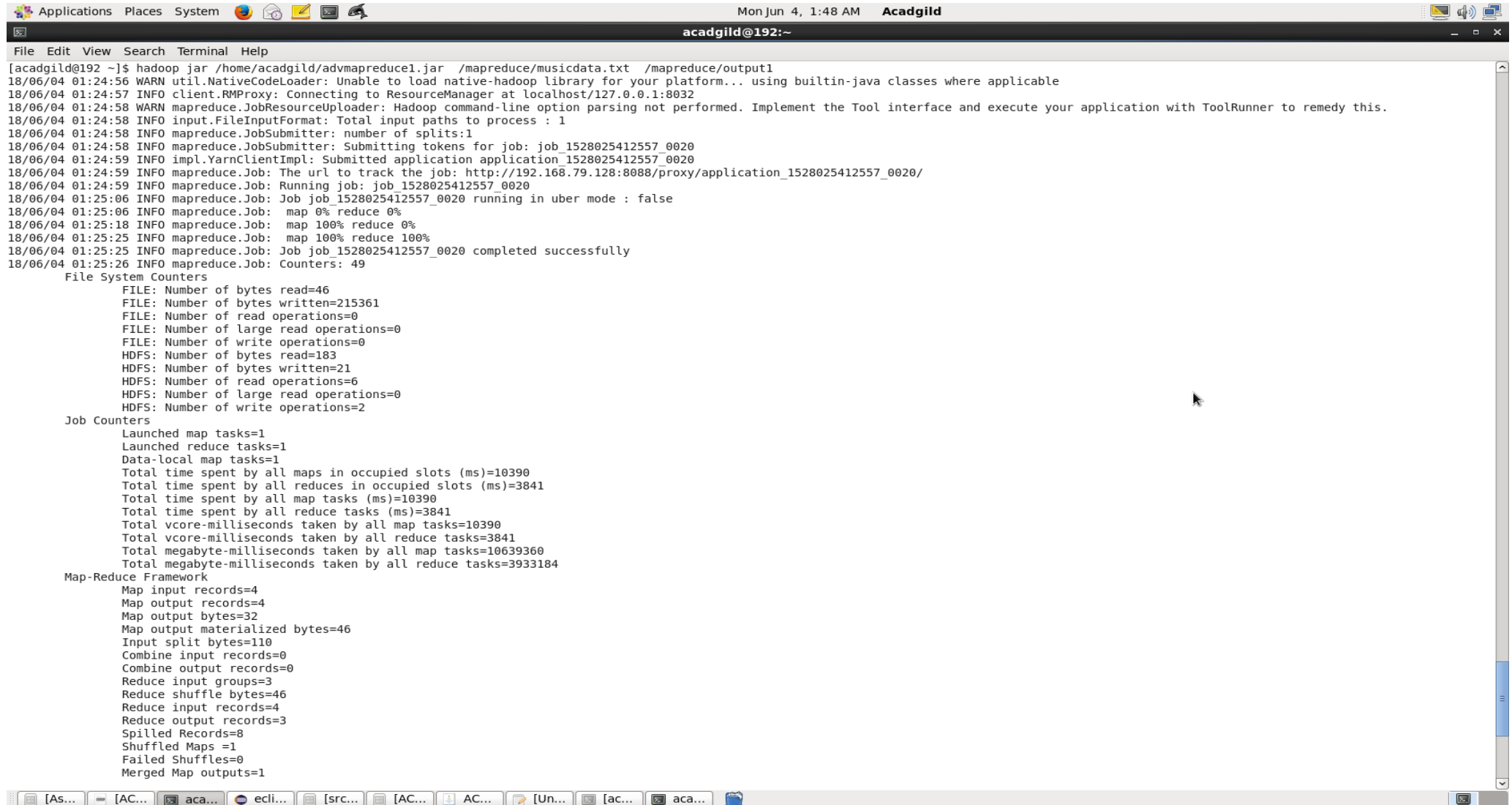


Advanced MapReduce

Task 1:

Find the number of unique listeners in the data set.

Approach for this is Userid is Unique so assigning this as KEY, There is no value column which can be assigned as nullable



```
[acadgild@192 ~]$ hadoop jar /home/acadgild/advmapreduce1.jar /mapreduce/musicdata.txt /mapreduce/output1
18/06/04 01:24:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/04 01:24:57 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 01:24:58 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/04 01:24:58 INFO input.FileInputFormat: Total input paths to process : 1
18/06/04 01:24:58 INFO mapreduce.JobSubmitter: number of splits:1
18/06/04 01:24:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528025412557_0020
18/06/04 01:24:59 INFO impl.YarnClientImpl: Submitted application application_1528025412557_0020
18/06/04 01:24:59 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1528025412557_0020/
18/06/04 01:24:59 INFO mapreduce.Job: Running job: job_1528025412557_0020
18/06/04 01:25:06 INFO mapreduce.Job: Job job_1528025412557_0020 running in uber mode : false
18/06/04 01:25:06 INFO mapreduce.Job: map 0% reduce 0%
18/06/04 01:25:18 INFO mapreduce.Job: map 100% reduce 0%
18/06/04 01:25:25 INFO mapreduce.Job: map 100% reduce 100%
18/06/04 01:25:25 INFO mapreduce.Job: Job job_1528025412557_0020 completed successfully
18/06/04 01:25:26 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=46
    FILE: Number of bytes written=215361
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=103
    HDFS: Number of bytes written=21
    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)=10390
    Total time spent by all reduces in occupied slots (ms)=3841
    Total time spent by all map tasks (ms)=10390
    Total time spent by all reduce tasks (ms)=3841
    Total vcore-milliseconds taken by all map tasks=10390
    Total vcore-milliseconds taken by all reduce tasks=3841
    Total megabyte-milliseconds taken by all map tasks=10639360
    Total megabyte-milliseconds taken by all reduce tasks=3933184
  Map-Reduce Framework
    Map input records=4
    Map output records=4
    Map output bytes=32
    Map output materialized bytes=46
    Input split bytes=110
    Combine input records=0
    Combine output records=0
    Reduce input groups=3
    Reduce shuffle bytes=46
    Reduce input records=4
    Reduce output records=3
    Spilled Records=8
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
```

Advanced MapReduce

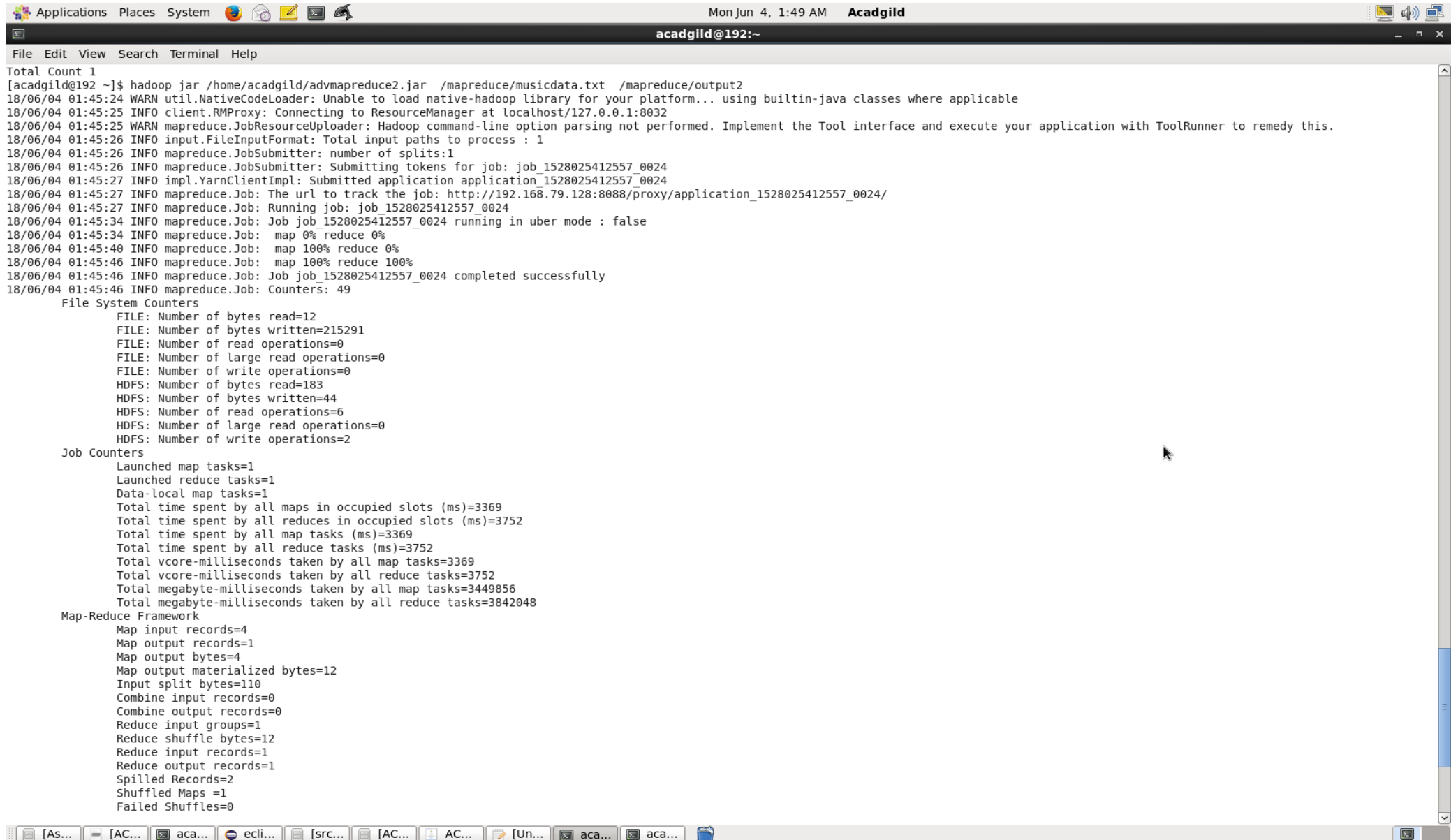
Answer for the requirement as show below, Output of Unique listeners

```
Applications Places System Mon Jun 4, 1:49 AM Acadgild
acadgild@192:~
File Edit View Search Terminal Help
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=10390
Total time spent by all reduces in occupied slots (ms)=3841
Total time spent by all map tasks (ms)=10390
Total time spent by all reduce tasks (ms)=3841
Total vcore-milliseconds taken by all map tasks=10390
Total vcore-milliseconds taken by all reduce tasks=3841
Total megabyte-milliseconds taken by all map tasks=10639360
Total megabyte-milliseconds taken by all reduce tasks=3933184
Map-Reduce Framework
Map input records=4
Map output records=4
Map output bytes=32
Map output materialized bytes=46
Input split bytes=110
Combine input records=0
Combine output records=0
Reduce input groups=3
Reduce shuffle bytes=46
Reduce input records=4
Reduce output records=3
Spilled Records=8
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=108
CPU time spent (ms)=5920
Physical memory (bytes) snapshot=324935680
Virtual memory (bytes) snapshot=4126842880
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=73
File Output Format Counters
Bytes Written=21
[acadgild@192 ~]$ hadoop fs -ls /mapreduce/output1
18/06/04 01:25:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
^[[Als: `/mapreduce/output1': No such file or directory
[acadgild@192 ~]$ hadoop fs -ls /mapreduce/output1
18/06/04 01:26:00 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 2018-06-04 01:25 /mapreduce/output1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 21 2018-06-04 01:25 /mapreduce/output1/part-r-00000
[acadgild@192 ~]$ hadoop fs -cat /mapreduce/output1/part-r-00000
18/06/04 01:26:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113
111115
111117
[acadgild@192 ~]$
```

Advanced MapReduce

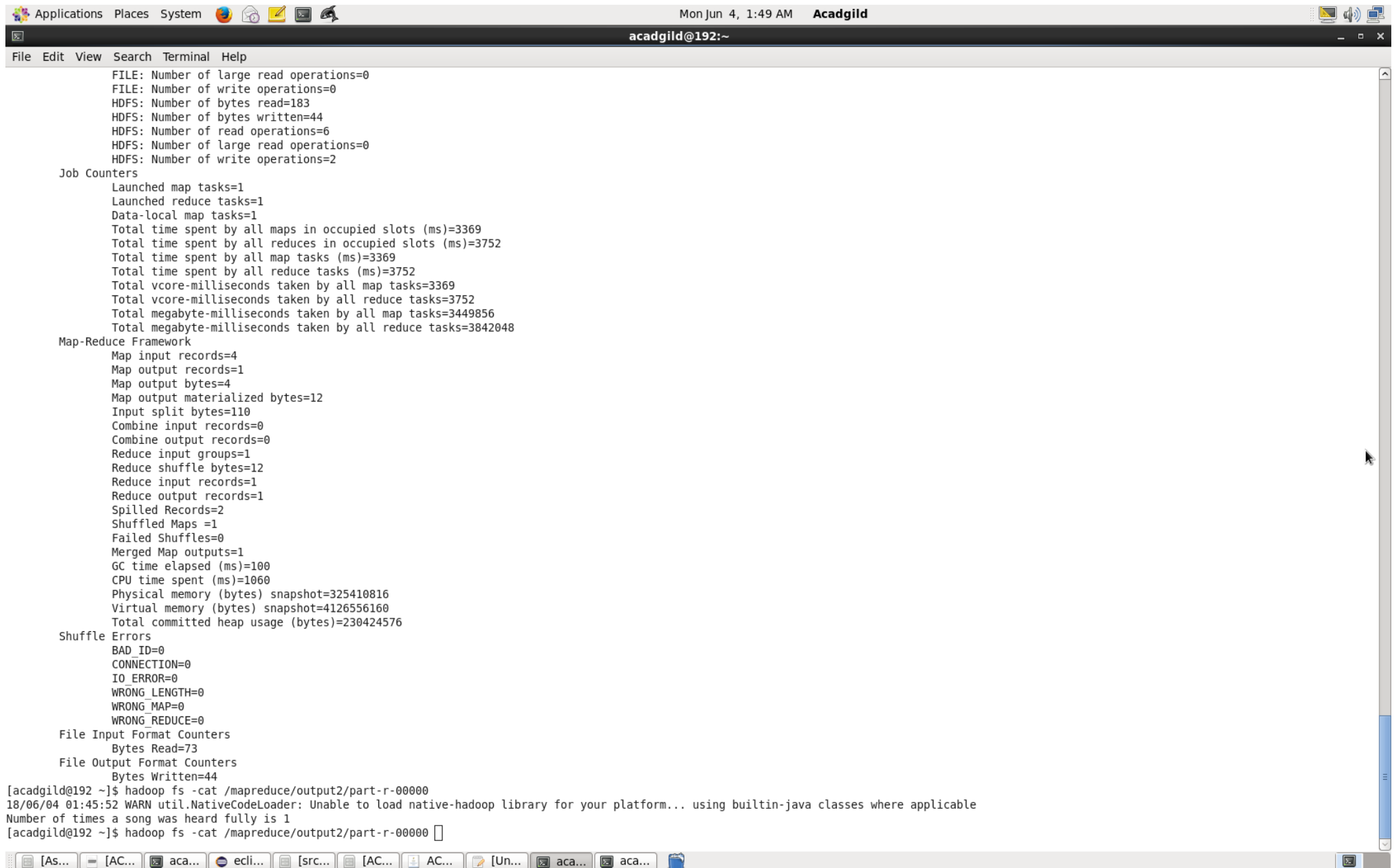
Task 2: What are the number of times a song was heard fully?

Indicator to listen song fully is 1 where 0 is for skipped so, here we need number of times i.e sum of fully heard songs approach is Key is Nullable, Value Will be Iterated and counted the time heard fully



```
Applications Places System Mon Jun 4, 1:49 AM Acadgild
acadgild@192:~
File Edit View Search Terminal Help
Total Count 1
[acadgild@192 ~]$ hadoop jar /home/acadgild/advmapreduce2.jar /mapreduce/musicdata.txt /mapreduce/output2
18/06/04 01:45:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/04 01:45:25 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 01:45:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/04 01:45:26 INFO input.FileInputFormat: Total input paths to process : 1
18/06/04 01:45:26 INFO mapreduce.JobSubmitter: number of splits:1
18/06/04 01:45:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528025412557_0024
18/06/04 01:45:27 INFO impl.YarnClientImpl: Submitted application application_1528025412557_0024
18/06/04 01:45:27 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1528025412557_0024/
18/06/04 01:45:27 INFO mapreduce.Job: Running job: job_1528025412557_0024
18/06/04 01:45:34 INFO mapreduce.Job: Job job_1528025412557_0024 running in uber mode : false
18/06/04 01:45:34 INFO mapreduce.Job:  map 0% reduce 0%
18/06/04 01:45:40 INFO mapreduce.Job:  map 100% reduce 0%
18/06/04 01:45:46 INFO mapreduce.Job:  map 100% reduce 100%
18/06/04 01:45:46 INFO mapreduce.Job: Job job_1528025412557_0024 completed successfully
18/06/04 01:45:46 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=12
    FILE: Number of bytes written=215291
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=183
    HDFS: Number of bytes written=44
    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)=3369
    Total time spent by all reduces in occupied slots (ms)=3752
    Total time spent by all map tasks (ms)=3369
    Total time spent by all reduce tasks (ms)=3752
    Total vcore-milliseconds taken by all map tasks=3369
    Total vcore-milliseconds taken by all reduce tasks=3752
    Total megabyte-milliseconds taken by all map tasks=3449856
    Total megabyte-milliseconds taken by all reduce tasks=3842048
  Map-Reduce Framework
    Map input records=4
    Map output records=1
    Map output bytes=4
    Map output materialized bytes=12
    Input split bytes=110
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=12
    Reduce input records=1
    Reduce output records=1
    Spilled Records=2
    Shuffled Maps =1
    Failed Shuffles=0
```

Advanced MapReduce



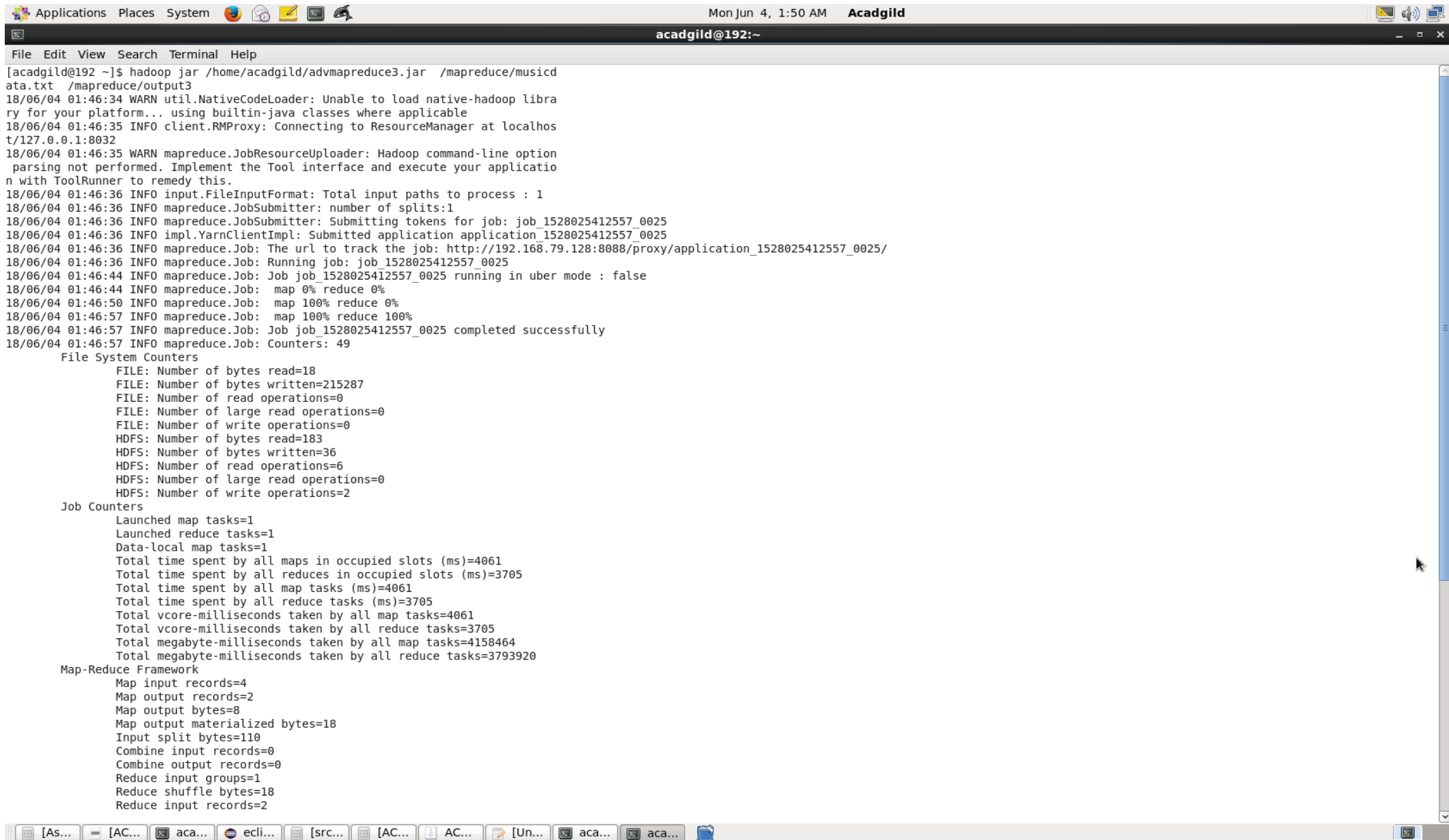
The screenshot shows a Linux desktop environment with a terminal window titled 'acadgild@192:~'. The terminal displays the output of a Hadoop MapReduce job. The output is organized into sections: FILE statistics, HDFS statistics, Job Counters, Map-Reduce Framework, Shuffle Errors, File Input Format Counters, and File Output Format Counters. At the bottom, the user runs the command 'hadoop fs -cat /mapreduce/output2/part-r-00000', which outputs the text 'Number of times a song was heard fully is 1'. The desktop background is dark, and the terminal window has a light gray border. The top of the window shows the system menu bar with 'Applications', 'Places', and 'System' menus, along with the date and time 'Mon Jun 4, 1:49 AM' and the username 'Acadgild'. The bottom of the window shows the taskbar with several open applications, including a file manager, a terminal, and a web browser.

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=183
HDFS: Number of bytes written=44
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)=3369
  Total time spent by all reduces in occupied slots (ms)=3752
  Total time spent by all map tasks (ms)=3369
  Total time spent by all reduce tasks (ms)=3752
  Total vcore-milliseconds taken by all map tasks=3369
  Total vcore-milliseconds taken by all reduce tasks=3752
  Total megabyte-milliseconds taken by all map tasks=3449856
  Total megabyte-milliseconds taken by all reduce tasks=3842048
Map-Reduce Framework
  Map input records=4
  Map output records=1
  Map output bytes=4
  Map output materialized bytes=12
  Input split bytes=110
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=12
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=100
  CPU time spent (ms)=1060
  Physical memory (bytes) snapshot=325410816
  Virtual memory (bytes) snapshot=4126556160
  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=73
File Output Format Counters
  Bytes Written=44
[acadgild@192 ~]$ hadoop fs -cat /mapreduce/output2/part-r-00000
18/06/04 01:45:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Number of times a song was heard fully is 1
[acadgild@192 ~]$ hadoop fs -cat /mapreduce/output2/part-r-00000
```

Advanced MapReduce

TASK 3: What are the number of times a song was shared?

Approach is same as above where we need to count the number of shares so, Key can be nullable, where we will iterate through Values clause for count



```
acadmild@192:~  
File Edit View Search Terminal Help  
[acadmild@192 ~]$ hadoop jar /home/acadmild/advmapreduce3.jar /mapreduce/musicdata.txt /mapreduce/output3  
18/06/04 01:46:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
18/06/04 01:46:35 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/06/04 01:46:35 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
18/06/04 01:46:36 INFO input.FileInputFormat: Total input paths to process : 1  
18/06/04 01:46:36 INFO mapreduce.JobSubmitter: number of splits:1  
18/06/04 01:46:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528025412557_0025  
18/06/04 01:46:36 INFO impl.YarnClientImpl: Submitted application application_1528025412557_0025  
18/06/04 01:46:36 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1528025412557_0025/  
18/06/04 01:46:36 INFO mapreduce.Job: Running job: job_1528025412557_0025  
18/06/04 01:46:44 INFO mapreduce.Job: Job job_1528025412557_0025 running in uber mode : false  
18/06/04 01:46:44 INFO mapreduce.Job:  map 0% reduce 0%  
18/06/04 01:46:50 INFO mapreduce.Job:  map 100% reduce 0%  
18/06/04 01:46:57 INFO mapreduce.Job:  map 100% reduce 100%  
18/06/04 01:46:57 INFO mapreduce.Job: Job job_1528025412557_0025 completed successfully  
18/06/04 01:46:57 INFO mapreduce.Job: Counters: 49  
  File System Counters  
    FILE: Number of bytes read=18  
    FILE: Number of bytes written=215287  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=183  
    HDFS: Number of bytes written=36  
    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)=4061  
    Total time spent by all reduces in occupied slots (ms)=3705  
    Total time spent by all map tasks (ms)=4061  
    Total time spent by all reduce tasks (ms)=3705  
    Total vcore-milliseconds taken by all map tasks=4061  
    Total vcore-milliseconds taken by all reduce tasks=3705  
    Total megabyte-milliseconds taken by all map tasks=4158464  
    Total megabyte-milliseconds taken by all reduce tasks=3793920  
  Map-Reduce Framework  
    Map input records=4  
    Map output records=2  
    Map output bytes=8  
    Map output materialized bytes=18  
    Input split bytes=110  
    Combine input records=0  
    Combine output records=0  
    Reduce input groups=1  
    Reduce shuffle bytes=18  
    Reduce input records=2
```

Advanced MapReduce

```
Applications Places System Mon Jun 4, 1:50 AM Acadgild
acadgild@192:~
File Edit View Search Terminal Help
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=183
HDFS: Number of bytes written=36
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)=4061
  Total time spent by all reduces in occupied slots (ms)=3705
  Total time spent by all map tasks (ms)=4061
  Total time spent by all reduce tasks (ms)=3705
  Total vcore-milliseconds taken by all map tasks=4061
  Total vcore-milliseconds taken by all reduce tasks=3705
  Total megabyte-milliseconds taken by all map tasks=4158464
  Total megabyte-milliseconds taken by all reduce tasks=3793920
Map-Reduce Framework
  Map input records=4
  Map output records=2
  Map output bytes=8
  Map output materialized bytes=18
  Input split bytes=110
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=18
  Reduce input records=2
  Reduce output records=1
  Spilled Records=4
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=101
  CPU time spent (ms)=1380
  Physical memory (bytes) snapshot=326213632
  Virtual memory (bytes) snapshot=4126855168
  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=73
File Output Format Counters
  Bytes Written=36
[acadgild@192 ~]$ hadoop fs -cat /mapreduce/output3/part-r-00000
18/06/04 01:47:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Number of time song was shared is 2
[acadgild@192 ~]$
```