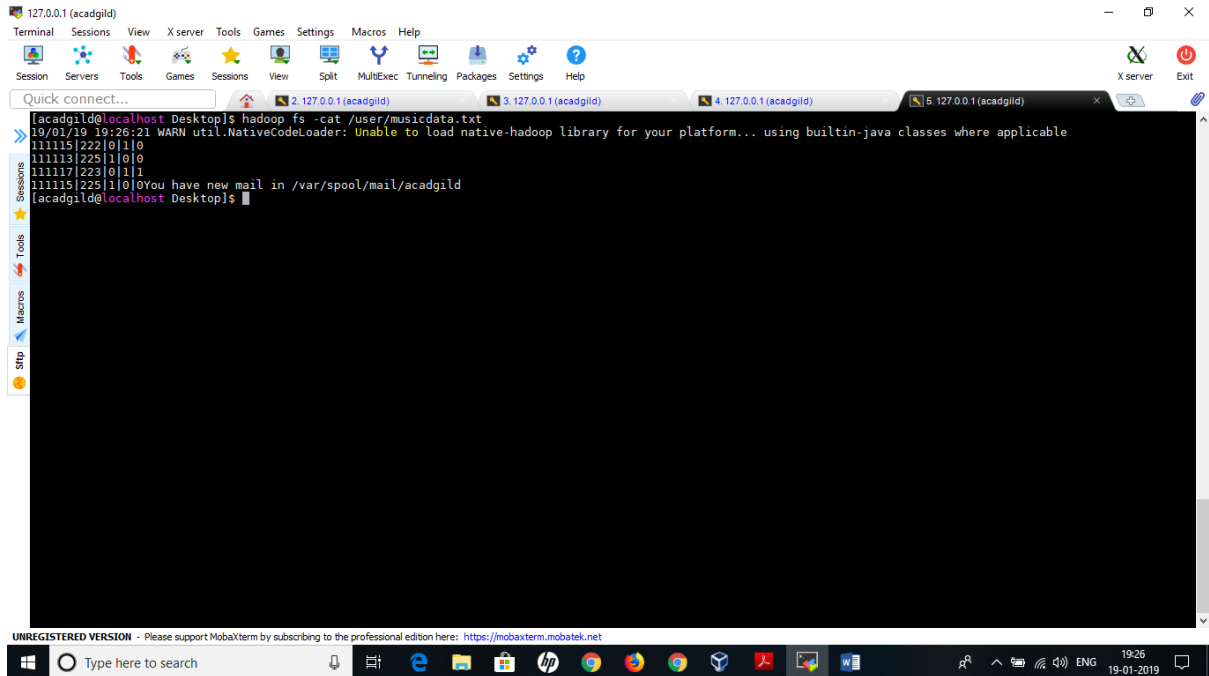


# Assignment 5.1 Advance Map Reduce.

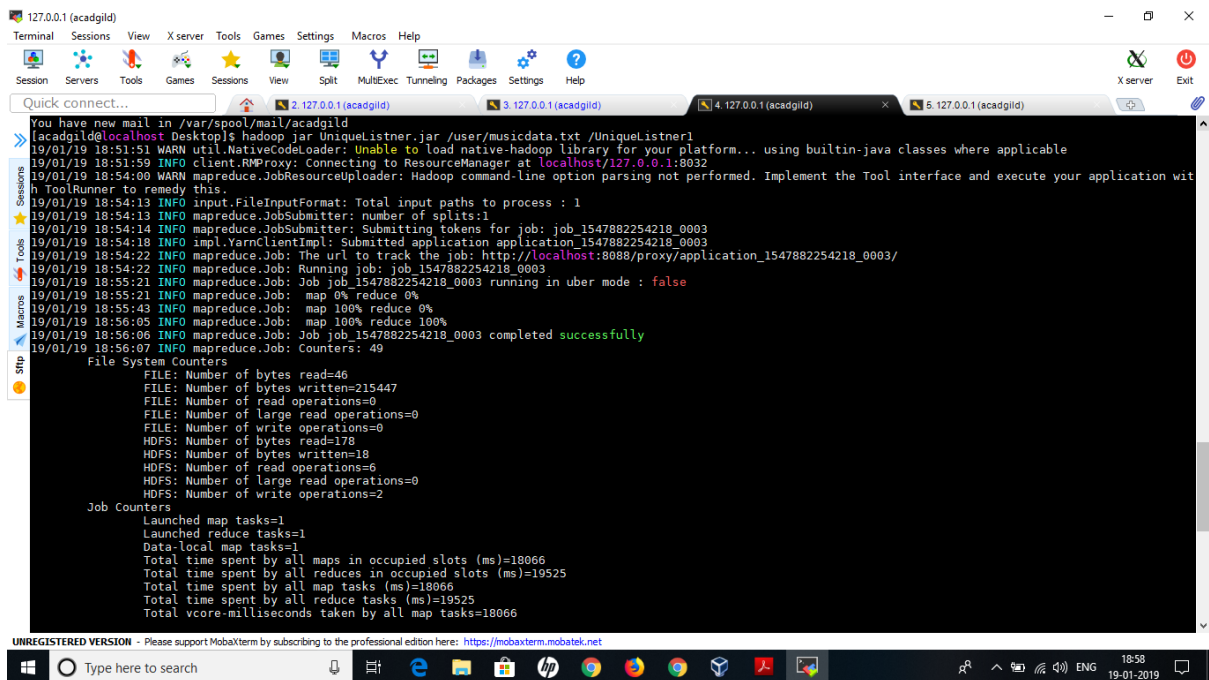
## Input Data Set:



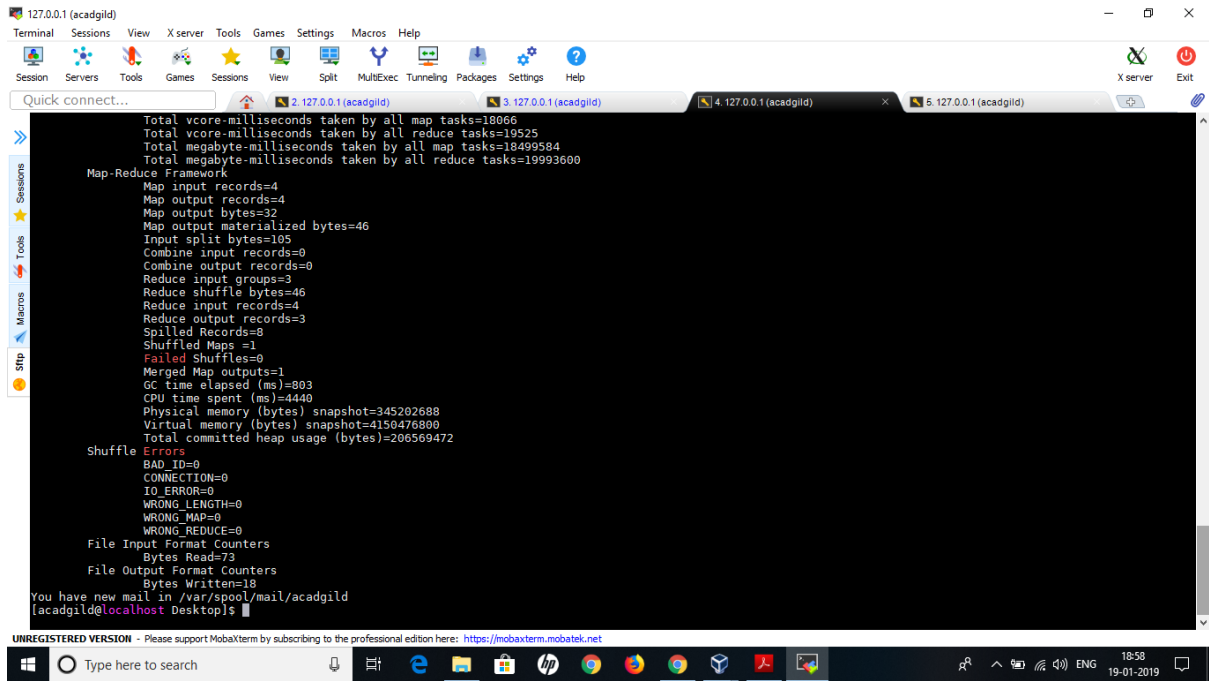
```
[acadgild@localhost Desktop]$ hadoop fs -cat /user/musicdata.txt
19/01/19 19:26:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

Task1: Find the number of unique listeners in the data set.

Solution:



```
[acadgild@localhost Desktop]$ hadoop jar UniqueListener.jar /user/musicdata.txt /Uniquelister1
19/01/19 18:51:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/19 18:51:59 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/19 18:54:00 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
h ToolRunner to remedy this.
19/01/19 18:54:13 INFO input.FileInputFormat: Total input paths to process : 1
19/01/19 18:54:13 INFO mapreduce.JobSubmitter: number of splits:1
19/01/19 18:54:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547882254218_0003
19/01/19 18:54:18 INFO impl.YarnClientImpl: Submitted application application_1547882254218_0003
19/01/19 18:54:22 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547882254218_0003/
19/01/19 18:54:22 INFO mapreduce.Job: Running job: job_1547882254218_0003
19/01/19 18:55:21 INFO mapreduce.Job: Job job_1547882254218_0003 running in uber mode : false
19/01/19 18:55:21 INFO mapreduce.Job: map 0% reduce 0%
19/01/19 18:55:43 INFO mapreduce.Job: map 100% reduce 0%
19/01/19 18:56:05 INFO mapreduce.Job: map 100% reduce 100%
19/01/19 18:56:06 INFO mapreduce.Job: Job job_1547882254218_0003 completed successfully
19/01/19 18:56:07 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=46
  FILE: Number of bytes written=215447
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=178
  HDFS: Number of bytes written=18
  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)=18066
  Total time spent by all reduces in occupied slots (ms)=19525
  Total time spent by all map tasks (ms)=18066
  Total time spent by all reduce tasks (ms)=19525
  Total vcore-milliseconds taken by all map tasks=18066
```



127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

2 127.0.0.1 (acadgild) 3 127.0.0.1 (acadgild) 4 127.0.0.1 (acadgild) 5 127.0.0.1 (acadgild)

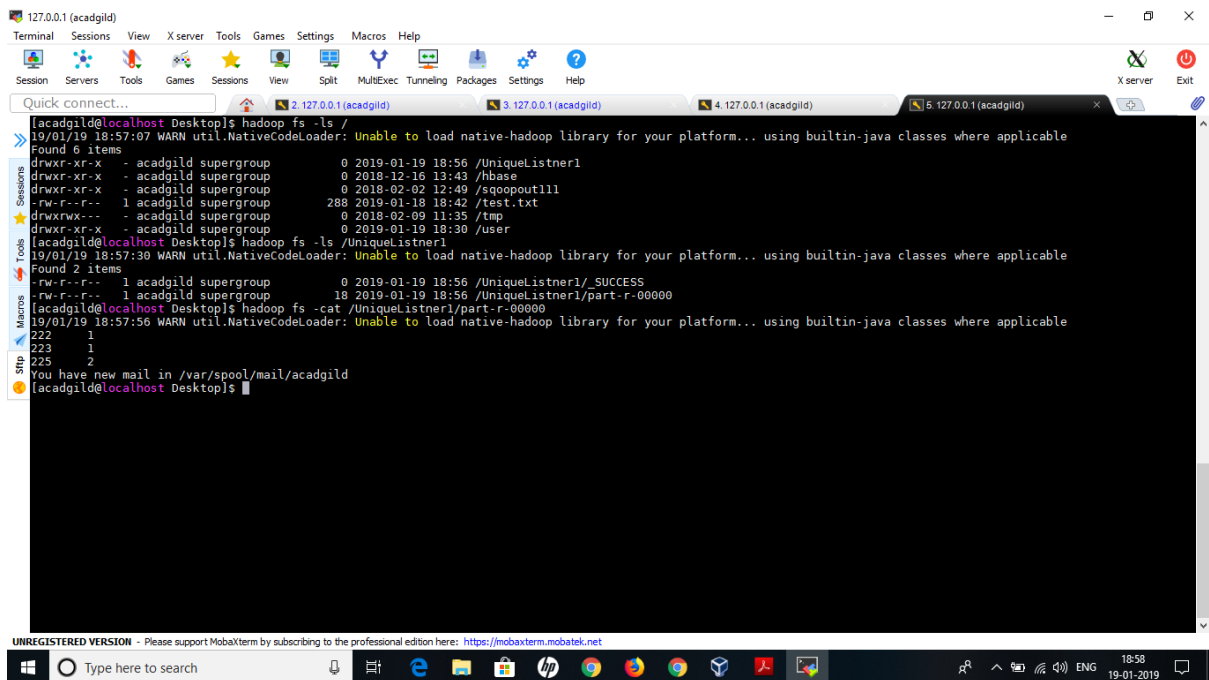
```
Total vcore-milliseconds taken by all map tasks=18066
Total vcore-milliseconds taken by all reduce tasks=19525
Total megabyte-milliseconds taken by all map tasks=18499584
Total megabyte-milliseconds taken by all reduce tasks=19993600

Map-Reduce Framework
  Map input records=4
  Map output records=4
  Map output bytes=32
  Map output materialized bytes=46
  Input split bytes=105
  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)=803
  CPU time spent (ms)=4440
  Physical memory (bytes) snapshot=345202688
  Virtual memory (bytes) snapshot=4150476800
  Total committed heap usage (bytes)=206569472

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=18
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>



127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

2 127.0.0.1 (acadgild) 3 127.0.0.1 (acadgild) 4 127.0.0.1 (acadgild) 5 127.0.0.1 (acadgild)

```
[acadgild@localhost Desktop]$ hadoop fs -ls /
19/01/19 18:57:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 6 items
drwxr-xr-x - acadgild supergroup          0 2019-01-19 18:56 /UniqueListner1
drwxr-xr-x - acadgild supergroup          0 2018-12-16 13:43 /hbase
drwxr-xr-x - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
-rw-r--r-- 1 acadgild supergroup        288 2019-01-18 18:42 /test.txt
drwxrwx--- - acadgild supergroup          0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup          0 2019-01-19 18:30 /user
[acadgild@localhost Desktop]$ hadoop fs -ls /UniqueListner1
19/01/19 18:57:39 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 2019-01-19 18:56 /UniqueListner1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        18 2019-01-19 18:56 /UniqueListner1/part-r-000000
[acadgild@localhost Desktop]$ hadoop fs -cat /UniqueListner1/part-r-000000
19/01/19 18:57:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
222
1
223
1
225
2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

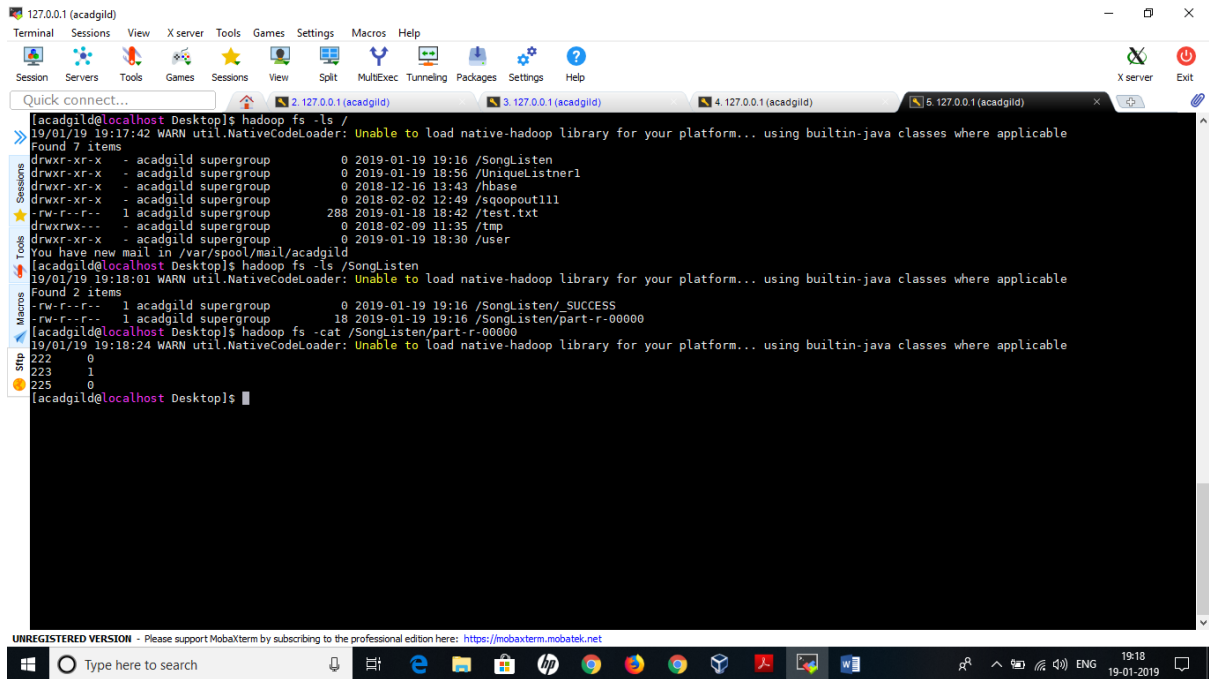
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Task2: What are the number of times a song was heard fully.

Solution:

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost Desktop]$ hadoop jar SongListen.jar /user/musicdata.txt /SongListen
19/01/19 19:14:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/19 19:14:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/19 19:14:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
h ToolRunner to remedy this.
19/01/19 19:14:51 INFO input.FileInputFormat: Total input paths to process : 1
19/01/19 19:14:51 INFO mapreduce.JobSubmitter: number of splits=1
19/01/19 19:14:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547882254218_0004
19/01/19 19:14:53 INFO impl.YarnClientImpl: Submitted application application_1547882254218_0004
19/01/19 19:14:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547882254218_0004/
19/01/19 19:14:55 INFO mapreduce.Job: Running job: job_1547882254218_0004
19/01/19 19:15:58 INFO mapreduce.Job: Job job_1547882254218_0004 running in uber mode : false
19/01/19 19:16:24 INFO mapreduce.Job: map 100% reduce 0%
19/01/19 19:16:52 INFO mapreduce.Job: map 100% reduce 100%
19/01/19 19:16:53 INFO mapreduce.Job: Job job_1547882254218_0004 completed successfully
19/01/19 19:16:54 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215459
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=176
  HDFS: Number of bytes written=18
  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)=19994
  Total time spent by all reduces in occupied slots (ms)=24901
  Total time spent by all map tasks (ms)=19994
  Total time spent by all reduce tasks (ms)=24901
  Total vcore-milliseconds taken by all map tasks=19994
  Total vcore-milliseconds taken by all reduce tasks=24901
  Total megabyte-milliseconds taken by all map tasks=20473856
  Total megabyte-milliseconds taken by all reduce tasks=25498624
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
Total vcore-milliseconds taken by all map tasks=19994
Total vcore-milliseconds taken by all reduce tasks=24901
Total megabyte-milliseconds taken by all map tasks=20473856
Total megabyte-milliseconds taken by all reduce tasks=25498624
Map-Reduce Framework
  Map input records=4
  Map output records=5
  Map output bytes=40
  Map output materialized bytes=56
  Input split bytes=105
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=56
  Reduce input records=5
  Reduce output records=3
  Spilled Records=10
  Shuffled Maps=1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=1009
  CPU time spent (ms)=4750
  Physical memory (bytes) snapshot=355786752
  Virtual memory (bytes) snapshot=4150517760
  Total committed heap usage (bytes)=206569472
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=18
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```



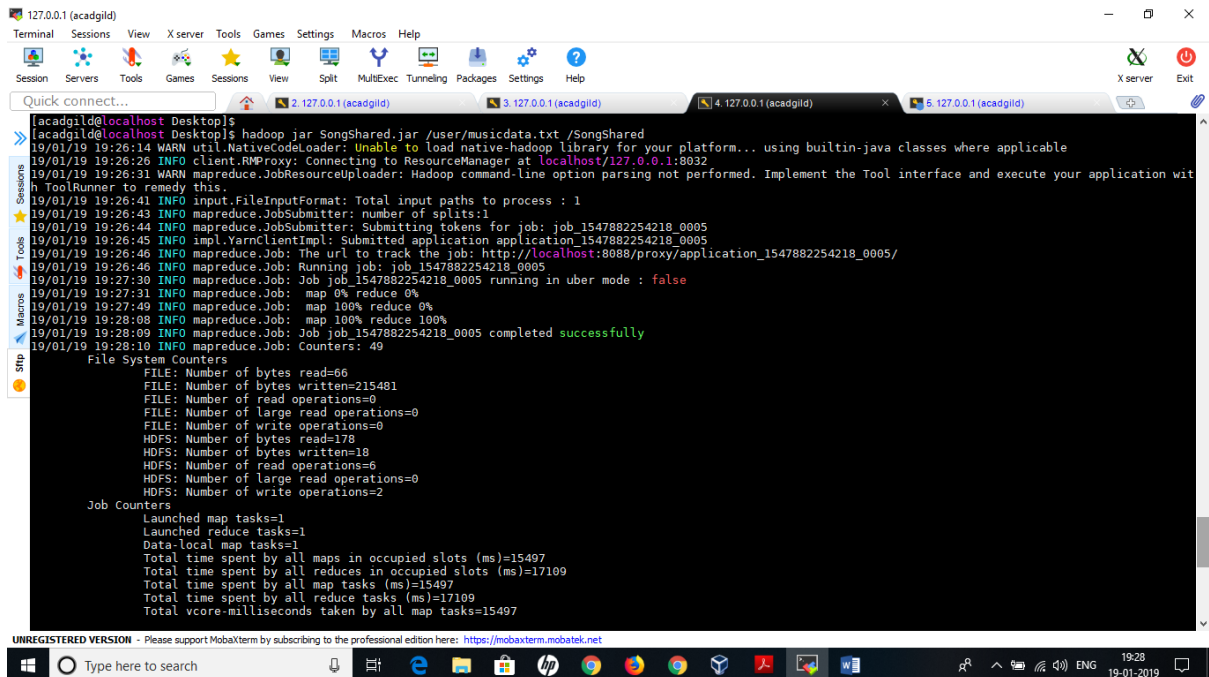
The screenshot shows a MobaXterm terminal window with a dark background. The terminal displays the following commands and output:

```
[acagild@localhost Desktop]$ hadoop fs -ls /
19/01/19 19:17:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - acagild supergroup          0 2019-01-19 19:16 /SongListen
drwxr-xr-x - acagild supergroup          0 2019-01-19 18:56 /UniqueListner1
drwxr-xr-x - acagild supergroup          0 2018-12-16 13:43 /hbase
drwxr-xr-x - acagild supergroup          0 2018-02-02 12:49 /sqoopout111
-rw-r--r-- 1 acagild supergroup        288 2019-01-18 18:42 /test.txt
drwxrwx--- - acagild supergroup          0 2018-02-09 11:35 /tmp
drwxr-xr-x - acagild supergroup          0 2019-01-19 18:30 /user
You have new mail in /var/spool/mail/acagild
[acagild@localhost Desktop]$ hadoop fs -ls /SongListen
19/01/19 19:18:01 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 acagild supergroup          0 2019-01-19 19:16 /SongListen/_SUCCESS
-rw-r--r-- 1 acagild supergroup        18 2019-01-19 19:16 /SongListen/part-r-000000
[acagild@localhost Desktop]$ hadoop fs -cat /SongListen/part-r-000000
19/01/19 19:18:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
222
0
223
1
225
0
[acagild@localhost Desktop]$
```

The terminal window has a menu bar with options: Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help. Below the menu bar is a toolbar with icons for Session, Servers, Tools, Games, Sessions, View, Split, MultiExec, Tunneling, Packages, Settings, Help. A 'Quick connect...' search bar is located below the toolbar. The terminal output shows the results of Hadoop file system commands, including listing the root directory and a specific directory, and displaying the contents of a file. The bottom of the window shows a Windows taskbar with the search bar and various application icons.

Task3: What are the number of times a song was shared.

Solution:



The screenshot shows a MobaXterm terminal window with a dark background. The terminal displays the following commands and output:

```
[acagild@localhost Desktop]$ hadoop jar SongShared.jar /user/musicdata.txt /SongShared
19/01/19 19:26:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/19 19:26:26 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/19 19:26:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/01/19 19:26:41 INFO input.FileInputFormat: Total input paths to process : 1
19/01/19 19:26:43 INFO mapreduce.JobSubmitter: number of splits:1
19/01/19 19:26:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547882254218_0005
19/01/19 19:26:45 INFO impl.YarnClientImpl: Submitted application application_1547882254218_0005
19/01/19 19:26:46 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547882254218_0005/
19/01/19 19:26:46 INFO mapreduce.Job: Running job: job_1547882254218_0005
19/01/19 19:27:30 INFO mapreduce.Job: Job job_1547882254218_0005 running in uber mode : false
19/01/19 19:27:31 INFO mapreduce.Job:  map 0% reduce 0%
19/01/19 19:27:49 INFO mapreduce.Job:  map 100% reduce 0%
19/01/19 19:28:08 INFO mapreduce.Job:  map 100% reduce 100%
19/01/19 19:28:09 INFO mapreduce.Job: Job job_1547882254218_0005 completed successfully
19/01/19 19:28:10 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=86
  FILE: Number of bytes written=215481
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=176
  HDFS: Number of bytes written=18
  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)=15497
  Total time spent by all reduces in occupied slots (ms)=17109
  Total time spent by all map tasks (ms)=15497
  Total time spent by all reduce tasks (ms)=17109
  Total vcore-milliseconds taken by all map tasks=15497
```

The terminal window has a menu bar with options: Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help. Below the menu bar is a toolbar with icons for Session, Servers, Tools, Games, Sessions, View, Split, MultiExec, Tunneling, Packages, Settings, Help. A 'Quick connect...' search bar is located below the toolbar. The terminal output shows the execution of a Hadoop MapReduce job, including the submission of the job, the progress of the map and reduce tasks, and the final counters. The bottom of the window shows a Windows taskbar with the search bar and various application icons.

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 2 127.0.0.1 (acadgild) 3 127.0.0.1 (acadgild) 4 127.0.0.1 (acadgild) 5 127.0.0.1 (acadgild)
Total vcore-milliseconds taken by all map tasks=15497
Total vcore-milliseconds taken by all reduce tasks=17109
Total megabyte-milliseconds taken by all map tasks=15868928
Total megabyte-milliseconds taken by all reduce tasks=17519616
Map-Reduce Framework
  Map input records=4
  Map output records=6
  Map output bytes=49
  Map output materialized bytes=66
  Input split bytes=105
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=66
  Reduce input records=6
  Reduce output records=3
  Spilled Records=12
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=607
  CPU time spent (ms)=4510
  Physical memory (bytes) snapshot=357093264
  Virtual memory (bytes) snapshot=4150611968
  Total committed heap usage (bytes)=206569472
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=18
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 2 127.0.0.1 (acadgild) 3 127.0.0.1 (acadgild) 4 127.0.0.1 (acadgild) 5 127.0.0.1 (acadgild)
[acadgild@localhost Desktop]$ hadoop fs -ls /
19/01/19 19:28:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 8 items
drwxr-xr-x - acadgild supergroup 0 2019-01-19 19:16 /SongListen
drwxr-xr-x - acadgild supergroup 0 2019-01-19 19:28 /SongShared
drwxr-xr-x - acadgild supergroup 0 2019-01-19 18:56 /UniquelListner1
drwxr-xr-x - acadgild supergroup 0 2018-12-16 13:43 /hbase
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
-rw-r--r-- 1 acadgild supergroup 288 2019-01-18 18:42 /test.txt
drwxrwx--- - acadgild supergroup 0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup 0 2019-01-19 18:39 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ hadoop fs -ls /SongShared
19/01/19 19:29:16 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 2019-01-19 19:28 /SongShared/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2019-01-19 19:28 /SongShared/part-r-000000
[acadgild@localhost Desktop]$ hadoop fs -cat /SongShared/part-r-000000
19/01/19 19:29:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
222 0
223 0
225 2
[acadgild@localhost Desktop]$
```