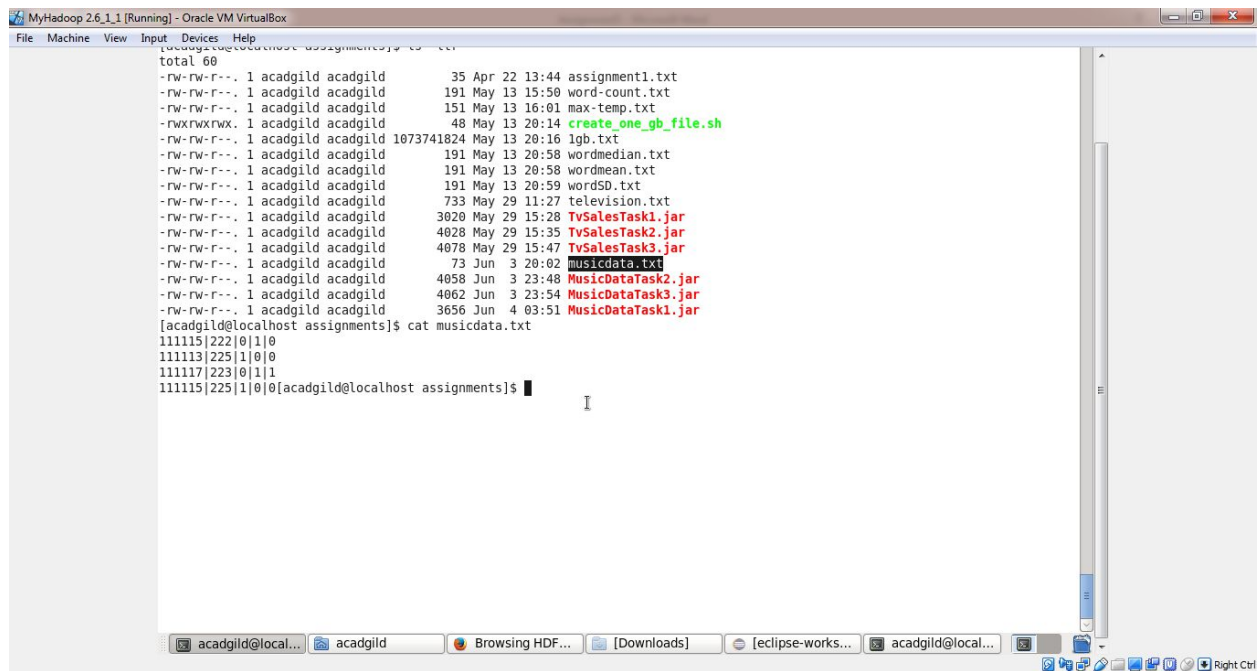


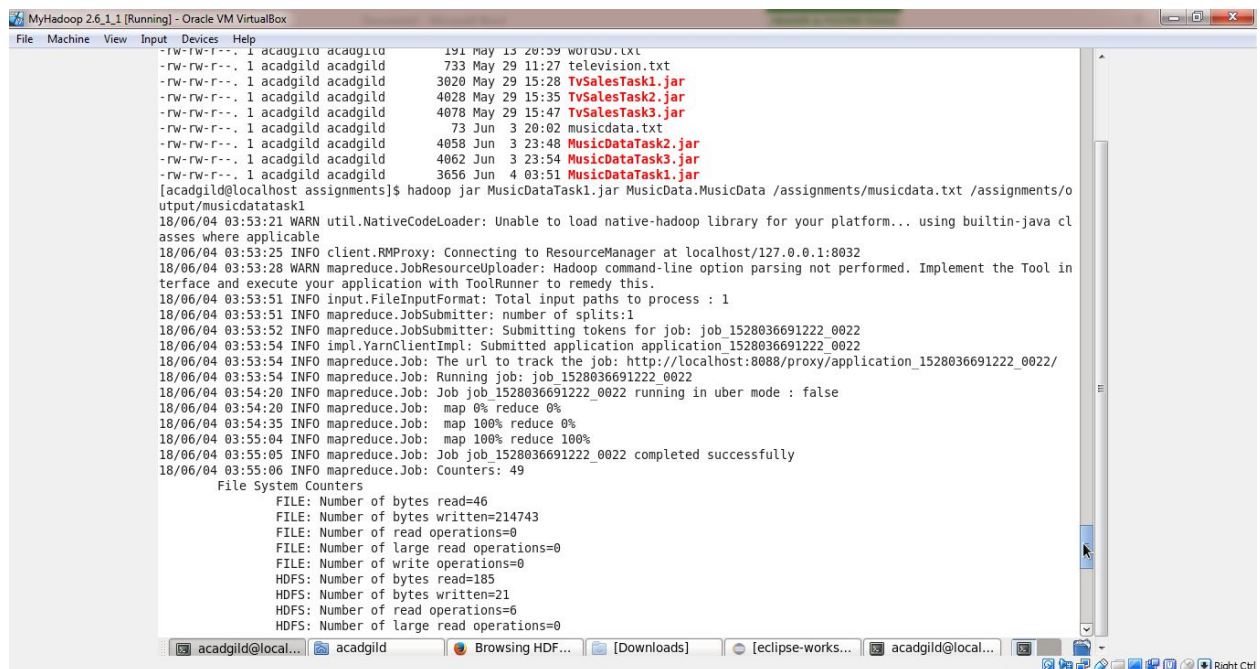
Dataset



```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
[acacgild@localhost ~]$ cat /home/assignments/assignments.txt
total 60
-rw-rw-r--. 1 acacgild acacgild      35 Apr 22 13:44 assignment1.txt
-rw-rw-r--. 1 acacgild acacgild     191 May 13 15:50 word-count.txt
-rw-rw-r--. 1 acacgild acacgild     151 May 13 16:01 max-temp.txt
-rwxrwxrwx. 1 acacgild acacgild      48 May 13 20:14 create_one_gb_file.sh
-rw-rw-r--. 1 acacgild acacgild 1073741824 May 13 20:16 lgb.txt
-rw-rw-r--. 1 acacgild acacgild     191 May 13 20:58 wordmedian.txt
-rw-rw-r--. 1 acacgild acacgild     191 May 13 20:58 wordmean.txt
-rw-rw-r--. 1 acacgild acacgild     191 May 13 20:59 word50.txt
-rw-rw-r--. 1 acacgild acacgild     733 May 29 11:27 television.txt
-rw-rw-r--. 1 acacgild acacgild    3020 May 29 15:28 TvSalesTask1.jar
-rw-rw-r--. 1 acacgild acacgild    4028 May 29 15:35 TvSalesTask2.jar
-rw-rw-r--. 1 acacgild acacgild    4078 May 29 15:47 TvSalesTask3.jar
-rw-rw-r--. 1 acacgild acacgild      73 Jun 3 20:02 musicdata.txt
-rw-rw-r--. 1 acacgild acacgild    4058 Jun 3 23:48 MusicDataTask2.jar
-rw-rw-r--. 1 acacgild acacgild    4062 Jun 3 23:54 MusicDataTask3.jar
-rw-rw-r--. 1 acacgild acacgild    3656 Jun 4 03:51 MusicDataTask1.jar
[acacgild@localhost assignments]$ cat musicdata.txt
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0[acacgild@localhost assignments]$
```

Task 1

Find the number of unique listeners in the data set.



```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
-rw-rw-r--. 1 acacgild acacgild     191 May 13 20:59 word50.txt
-rw-rw-r--. 1 acacgild acacgild     733 May 29 11:27 television.txt
-rw-rw-r--. 1 acacgild acacgild    3020 May 29 15:28 TvSalesTask1.jar
-rw-rw-r--. 1 acacgild acacgild    4028 May 29 15:35 TvSalesTask2.jar
-rw-rw-r--. 1 acacgild acacgild    4078 May 29 15:47 TvSalesTask3.jar
-rw-rw-r--. 1 acacgild acacgild      73 Jun 3 20:02 musicdata.txt
-rw-rw-r--. 1 acacgild acacgild    4058 Jun 3 23:48 MusicDataTask2.jar
-rw-rw-r--. 1 acacgild acacgild    4062 Jun 3 23:54 MusicDataTask3.jar
-rw-rw-r--. 1 acacgild acacgild    3656 Jun 4 03:51 MusicDataTask1.jar
[acacgild@localhost assignments]$ hadoop jar MusicDataTask1.jar MusicData.MusicData /assignments/musicdata.txt /assignments/output/musicdatatask1
18/06/04 03:53:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/04 03:53:25 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 03:53:28 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/04 03:53:51 INFO Input.FileInputFormat: Total input paths to process : 1
18/06/04 03:53:51 INFO mapreduce.JobSubmitter: number of splits:1
18/06/04 03:53:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528036691222_0022
18/06/04 03:53:54 INFO Impl.YarnClientImpl: Submitted application application_1528036691222_0022
18/06/04 03:53:54 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1528036691222_0022/
18/06/04 03:54:20 INFO mapreduce.Job: Running job: job_1528036691222_0022
18/06/04 03:54:20 INFO mapreduce.Job: Job job_1528036691222_0022 running in uber mode : false
18/06/04 03:54:35 INFO mapreduce.Job: map 0% reduce 0%
18/06/04 03:55:04 INFO mapreduce.Job: map 100% reduce 100%
18/06/04 03:55:05 INFO mapreduce.Job: Job job_1528036691222_0022 completed successfully
18/06/04 03:55:06 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=46
FILE: Number of bytes written=214743
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=185
HDFS: Number of bytes written=21
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Spilled Records=0
Shuffled Maps=1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=209
CPU time spent (ms)=3040
Physical memory (bytes) snapshot=283926528
Virtual memory (bytes) snapshot=4118192128
Total committed heap usage (bytes)=170004480

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

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$
[acadgild@localhost assignments]$ hadoop jar MusicDataTask1.jar MusicData.MusicData /assignments/musicdata.txt /assignments/output/musicdatatask1
[acadgild@localhost assignments]$ hadoop fs -ls /assignments/output/musicdatatask1
18/06/04 03:56:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-06-04 03:55 /assignments/output/musicdatatask1/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 21 2018-06-04 03:55 /assignments/output/musicdatatask1/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$ hadoop fs -cat /assignments/output/musicdatatask1/part-r-000000
18/06/04 03:56:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
111113
111115
111117
[acadgild@localhost assignments]$
```

Task 2

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

Mapreduce program provides the track id of the song which was heard fully

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
111112
111117
[acadgild@localhost assignments]$ hadoop jar MusicDataTask2.jar MusicDataTask2.MusicDataDriver /assignments/musicdata.txt /as
signments/output/musicdatatask2
18/06/04 03:57:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/04 03:57:45 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 03:57:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 03:57:47 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/04 03:57:47 INFO mapred.FileInputFormat: Total input paths to process : 1
18/06/04 03:57:48 INFO mapreduce.JobSubmitter: number of splits:2
18/06/04 03:57:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528036691222_0023
18/06/04 03:57:51 INFO impl.YarnClientImpl: Submitted application application_1528036691222_0023
18/06/04 03:57:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528036691222_0023/
18/06/04 03:57:51 INFO mapreduce.Job: Running job: job_1528036691222_0023
18/06/04 03:58:19 INFO mapreduce.Job: Job job_1528036691222_0023 running in uber mode : false
18/06/04 03:58:19 INFO mapreduce.Job: map 0% reduce 0%
18/06/04 03:58:45 INFO mapreduce.Job: map 100% reduce 0%
18/06/04 03:59:17 INFO mapreduce.Job: map 100% reduce 100%
18/06/04 03:59:19 INFO mapreduce.Job: Job job_1528036691222_0023 completed successfully
18/06/04 03:59:21 INFO mapreduce.Job: Counters: 49

File System Counters
FILE: Number of bytes read=16
FILE: Number of bytes written=323047
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=308
HDFS: Number of bytes written=6
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=49440
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
    reduce input groups=1
    Reduce shuffle bytes=22
    Reduce input records=1
    Reduce output records=1
    Spilled Records=2
    Shuffled Maps =2
    Failed Shuffles=0
    Merged Map outputs=2
    GC time elapsed (ms)=644
    CPU time spent (ms)=4440
    Physical memory (bytes) snapshot=467738624
    Virtual memory (bytes) snapshot=6172844032
    Total committed heap usage (bytes)=307437568

    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=110
    File Output Format Counters
    Bytes Written=6

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$
[acadgild@localhost assignments]$ hadoop fs -ls /assignments/output/musicdatatask2
18/06/04 04:00:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-06-04 03:59 /assignments/output/musicdatatask2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 6 2018-06-04 03:59 /assignments/output/musicdatatask2/part-00000
[acadgild@localhost assignments]$ hadoop fs -cat /assignments/output/musicdatatask2/part-00000
18/06/04 04:00:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
223 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$
```

Task 3

What are the number of times a song was shared.

Mapreduce program provides the track id of the song and the number of times it was shared.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

[acadgild@localhost assignments]$ hadoop jar MusicDataTask3.jar MusicDataTask3.MusicDataDriver /assignments/musicdata.txt /as
signments/output/musicdatatask3
18/06/04 04:03:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/04 04:03:27 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 04:03:28 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/04 04:03:30 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/04 04:03:31 INFO mapred.FileInputFormat: Total input paths to process : 1
18/06/04 04:03:32 INFO mapreduce.JobSubmitter: number of splits:2
18/06/04 04:03:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528036691222_0024
18/06/04 04:03:34 INFO impl.YarnClientImpl: Submitted application application_1528036691222_0024
18/06/04 04:03:35 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1528036691222_0024/
18/06/04 04:03:35 INFO mapreduce.Job: Running job: job_1528036691222_0024
18/06/04 04:03:49 INFO mapreduce.Job: Job job_1528036691222_0024 running in uber mode : false
18/06/04 04:04:29 INFO mapreduce.Job: map 0% reduce 0%
18/06/04 04:04:29 INFO mapreduce.Job: map 100% reduce 0%
18/06/04 04:05:17 INFO mapreduce.Job: map 100% reduce 100%
18/06/04 04:05:18 INFO mapreduce.Job: Job job_1528036691222_0024 completed successfully
18/06/04 04:05:20 INFO mapreduce.Job: Counters: 49
    File System Counters
    FILE: Number of bytes read=26
    FILE: Number of bytes written=323055
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=308
    HDFS: Number of bytes written=6
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2

    Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=74189
    Total time spent by all reduces in occupied slots (ms)=23132
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=32
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=1594
CPU time spent (ms)=7350
Physical memory (bytes) snapshot=472371200
Virtual memory (bytes) snapshot=6172839936
Total committed heap usage (bytes)=307437568

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=110
File Output Format Counters
  Bytes Written=6

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$ hadoop fs -ls /assignments/output/musicdatatask3
18/06/04 04:05:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-06-04 04:05 /assignments/output/musicdatatask3/ SUCCESS
-rw-r--r-- 1 acadgild supergroup          6 2018-06-04 04:05 /assignments/output/musicdatatask3/part-00000
[acadgild@localhost assignments]$ hadoop fs -cat /assignments/output/musicdatatask3/part-00000
18/06/04 04:06:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
225      2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$
```

acacgild
Browsing HDF...
[Downloads]
[eclipse-works...
acadgild@local...
Right Ctrl