

Big Data Hadoop 'Assignment 5'

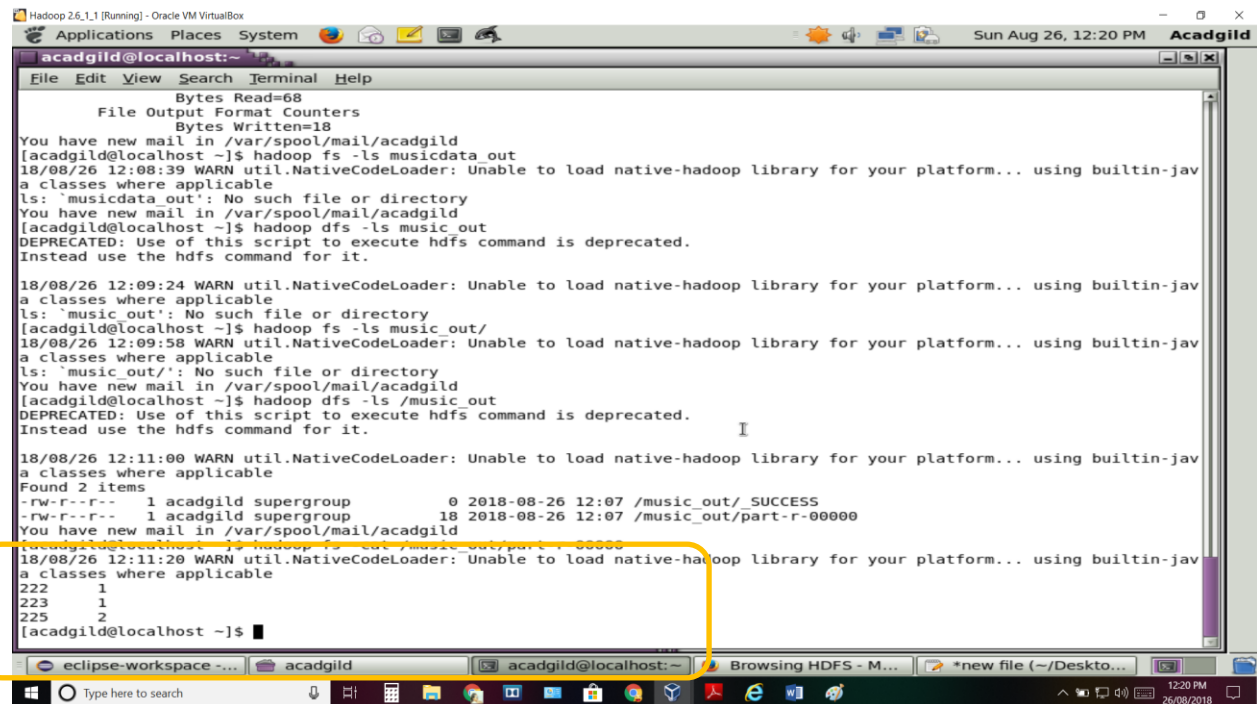
DATASET

```
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0
```

1st Column - UserId
2nd Column - TrackId
3rd Column - Songs Share status (1 for shared, 0 for not shared)
4th Column - Listening Platform (Radio or Web - 0 for radio, 1 for web)
5th Column - Song Listening Status (0 for skipped, 1 for fully heard)

TASK 1: Find the number of unique listeners in the data set.

Output: Listeners are calculated unique per track record



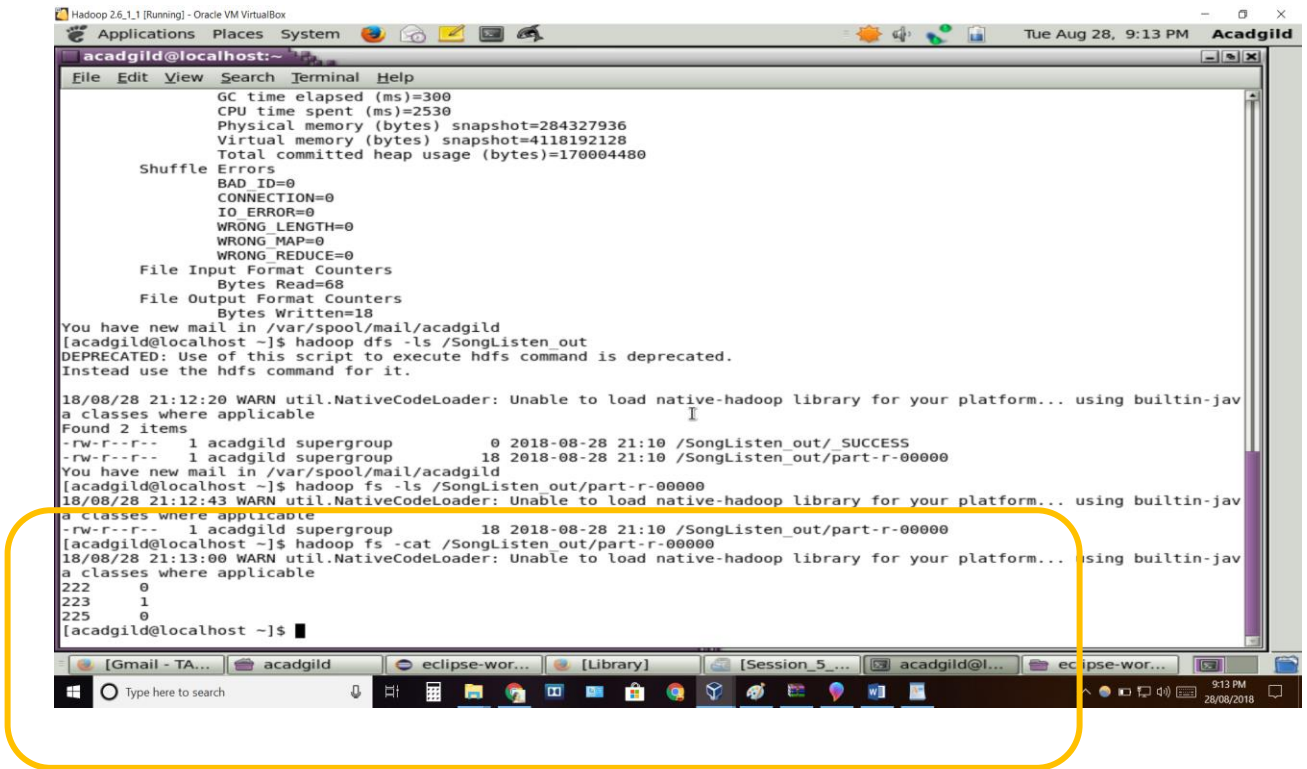
```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost: ~
File Edit View Search Terminal Help
Bytes Read=68
File Output Format Counters
Bytes Written=18
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls musicdata out
18/08/26 12:08:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
ls: 'musicdata out': No such file or directory
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop dfs -ls music_out
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/26 12:09:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
ls: 'music_out': No such file or directory
[acadgild@localhost ~]$ hadoop fs -ls music_out/
18/08/26 12:09:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
ls: 'music_out/': No such file or directory
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop dfs -ls /music_out
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/26 12:11:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-08-26 12:07 /music_out/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2018-08-26 12:07 /music_out/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /music_out/part-r-00000
18/08/26 12:11:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
222 1
223 1
225 2
[acadgild@localhost ~]$
```

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

Output: No. of times song fully heard per track.



```
acadmild@localhost:~$ hadoop dfs -ls /SongListen_out
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
18/08/28 21:12:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Found 2 items
-rw-r--r-- 1 acadmild supergroup 0 2018-08-28 21:10 /SongListen_out/_SUCCESS
-rw-r--r-- 1 acadmild supergroup 18 2018-08-28 21:10 /SongListen_out/part-r-00000
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -ls /SongListen_out/part-r-00000
18/08/28 21:12:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
-rw-r--r-- 1 acadmild supergroup 18 2018-08-28 21:10 /SongListen_out/part-r-00000
[acadmild@localhost ~]$ hadoop fs -cat /SongListen_out/part-r-00000
18/08/28 21:13:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
222 0
223 1
225 0
[acadmild@localhost ~]$
```

Task 3: What are the number of times a song was shared.

Output: No. of time song is being shared by user.

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Tue Aug 28, 9:25 PM Acadgild

acadgild@localhost:~
File Edit View Search Terminal Help

Spilled Records=12
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=242
CPU time spent (ms)=2640
Physical memory (bytes) snapshot=203287552
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=68
File Output Format Counters
Bytes Written=18

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop dfs -ls /SongShared out
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/28 21:24:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-08-28 21:23 /SongShared out/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2018-08-28 21:23 /SongShared out/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /SongShared out/part-r-00000
18/08/28 21:24:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
222 0
223 0
225 2
[acadgild@localhost ~]$
```