

Assignment 5.1 Advanced Mapreduce

Task 1 :

Dataset is sample data of songs heard by users on an online streaming platform. The

Description of data set attached in musicdata.txt is as follows: -

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)

111115|222|0|1|0

111113|225|1|0|0

111117|223|0|1|1

111115|225|1|0|0

Problem Statement:

Write Map Reduce program for following tasks.

Task 1:

Find the number of unique listeners in the data set.

Uploaded the Driver(MusicTask1.java), Mapper(MusicTask1Mapper.java) and reducer(MusicTask1Reducer..java) files separately along with code explanation.

Output:

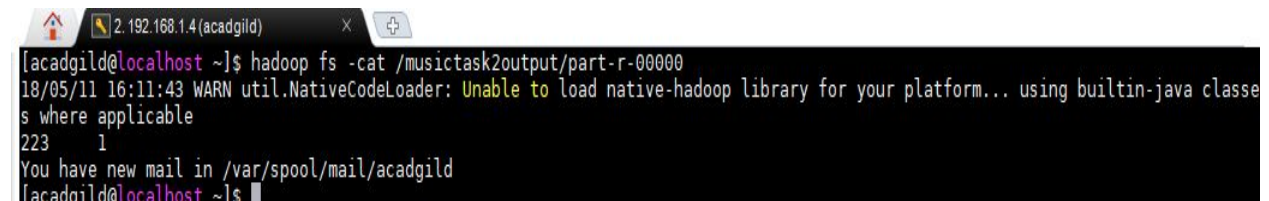
```
[acadgild@localhost ~]$ hadoop fs -ls /musictaskloutput
18/05/11 16:08:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-05-11 16:00 /musictaskloutput/_SUCCESS
-rw-r--r--  1 acadgild supergroup        22 2018-05-11 16:00 /musictaskloutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /musictaskloutput/part-r-00000
18/05/11 16:08:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113  225
111117  223
```

Task 2

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

Uploaded the Driver(MusicTask2.java), Mapper(MusicTask2Mapper.java) and reducer(MusicTask2Reducer.java) files separately along with code explanation.

Output:



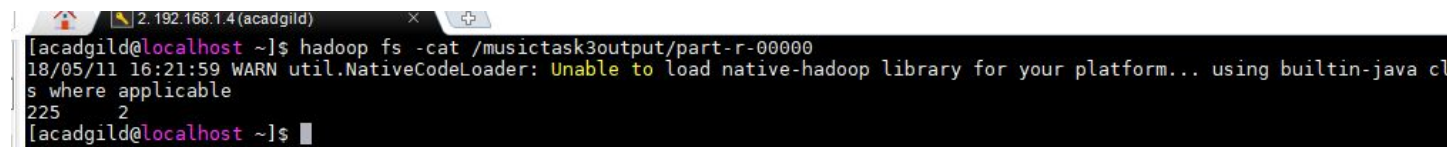
```
2.192.168.1.4 (acadgild) x
[acadgild@localhost ~]$ hadoop fs -cat /musictask2output/part-r-00000
18/05/11 16:11:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
223 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Task 3:

What are the number of times a song was shared.

Uploaded the Driver(MusicTask3.java), Mapper(MusicTask3Mapper.java) and reducer(MusicTask3Reducer.java) files separately along with code explanation.

Output:



```
2.192.168.1.4 (acadgild) x
[acadgild@localhost ~]$ hadoop fs -cat /musictask3output/part-r-00000
18/05/11 16:21:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
225 2
[acadgild@localhost ~]$
```