

## Assignment Submission- Session 5

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

### Task 1

Find the number of unique listeners in the data set.

**Solution:** For this, a jar file has been created by the name “MusicDataTask\_1.jar” . It can be run as shown in the screenshot.

```
[acadgild@localhost MyMapReducePrograms]$ hadoop jar MusicDataTask_1.jar /musicdata.txt /musicDataTask1Output
18/09/07 01:05:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
18/09/07 01:06:00 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/07 01:06:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface
18/09/07 01:06:02 INFO input.FileInputFormat: Total input paths to process : 1
18/09/07 01:06:02 INFO mapreduce.JobSubmitter: number of splits:1
18/09/07 01:06:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1536260102320_0003
18/09/07 01:06:03 INFO impl.YarnClientImpl: Submitted application application_1536260102320_0003
```

Since, number of unique listener corresponds to number of unique keys for reducer OR we can say the number of times reduce method is called.

We can get the value by inbuilt counter of mapreduce(Reduce Input Groups). Also a custom counter is made(Unique Listeners) for the same as can be seen in the screenshot.

```
Total megabyte-milliseconds taken by all reduce tasks=6436864
Map-Reduce Framework
  Map input records=4
  Map output records=4
  Map output bytes=48
  Map output materialized bytes=62
  Input split bytes=100
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=62
```

```
Total committed heap usage (bytes)=206369472
MusicData_Task_1_Driver$CustCount
UniqueListeners=3
shuffle=544444
```

### Task 2

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

**Solution:** For this, a jar file has been created by the name “MusicDataTask\_2.jar” . It can be run as shown in the screenshot.

```
[acagdild@localhost MyMapReducePrograms]$ hadoop jar MusicDataTask 2.jar /musicdata.txt /musicDataTask 2_Output
18/09/09 23:18:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/09 23:18:48 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/09 23:18:51 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
18/09/09 23:18:51 INFO input.FileInputFormat: Total input paths to process : 1
18/09/09 23:18:52 INFO mapreduce.JobSubmitter: number of splits:1
18/09/09 23:18:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1536513803135_0002
18/09/09 23:18:53 INFO impl.YarnClientImpl: Submitted application application_1536513803135_0002
18/09/09 23:18:53 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1536513803135_0002/
18/09/09 23:18:53 INFO mapreduce.Job: Running job: job_1536513803135_0002
18/09/09 23:19:11 INFO mapreduce.Job: Job job_1536513803135_0002 running in uber mode : false
18/09/09 23:19:11 INFO mapreduce.Job: map 0% reduce 0%
18/09/09 23:19:25 INFO mapreduce.Job: map 100% reduce 0%
18/09/09 23:19:39 INFO mapreduce.Job: map 100% reduce 100%
18/09/09 23:19:40 INFO mapreduce.Job: Job job_1536513803135_0002 completed successfully
18/09/09 23:19:40 INFO mapreduce.Job: Counters: 49
```

The output generated is as below screenshot, number of times songs are fully heard corresponding to each song id.

```
[acagdild@localhost MyMapReducePrograms]$ hadoop fs -ls /musicDataTask 2_Output
18/09/09 23:20:49 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 acagdild supergroup 0 2018-09-09 23:19 /musicDataTask 2_Output/SUCCESS
-rw-r--r-- 1 acagdild supergroup 18 2018-09-09 23:19 /musicDataTask 2_Output/part-r-00000
You have new mail in /var/spool/mail/acagdild
[acagdild@localhost MyMapReducePrograms]$ hadoop fs -cat /musicDataTask 2_Output/part-r-00000
18/09/09 23:21:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
222 0
223 1
225 0
```

### Task 3

What are the number of times a song was shared.

**Solution:** For this, a jar file has been created by the name “MusicDataTask\_3.jar” . It can be run as shown in the screenshot.

```
[acagdild@localhost MyMapReducePrograms]$ hadoop jar MusicDataTask 3.jar /musicdata.txt /musicDataTask 3_Output
Not a valid JAR: /home/acagdild/akshat/MyMapReducePrograms/MusicDataTask 3.jar
[acagdild@localhost MyMapReducePrograms]$ hadoop jar MusicDataTask 3.jar /musicdata.txt /musicDataTask 3_Output
18/09/09 23:48:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/09 23:48:11 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/09 23:48:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
18/09/09 23:48:14 INFO input.FileInputFormat: Total input paths to process : 1
18/09/09 23:48:15 INFO mapreduce.JobSubmitter: number of splits:1
18/09/09 23:48:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1536513803135_0003
18/09/09 23:48:16 INFO impl.YarnClientImpl: Submitted application application_1536513803135_0003
18/09/09 23:48:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1536513803135_0003/
18/09/09 23:48:16 INFO mapreduce.Job: Running job: job_1536513803135_0003
18/09/09 23:48:34 INFO mapreduce.Job: Job job_1536513803135_0003 running in uber mode : false
18/09/09 23:48:34 INFO mapreduce.Job: map 0% reduce 0%
18/09/09 23:48:47 INFO mapreduce.Job: map 100% reduce 0%
```

The output generated is as below screenshot, number of times songs are shared corresponding to each song id.

```
[acagdild@localhost MyMapReducePrograms]$ hadoop fs -cat /musicDataTask 3_Output/part-r-00000
18/09/09 23:50:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
222 0
223 0
225 2
You have new mail in /var/spool/mail/acagdild
[acagdild@localhost MyMapReducePrograms]$
```