

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)

NOTE: All the Map reduce programs related to TASKS 1,2,3 are written and exported the programs as respective jar files. Programs are attached as Zipped files in Github.

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

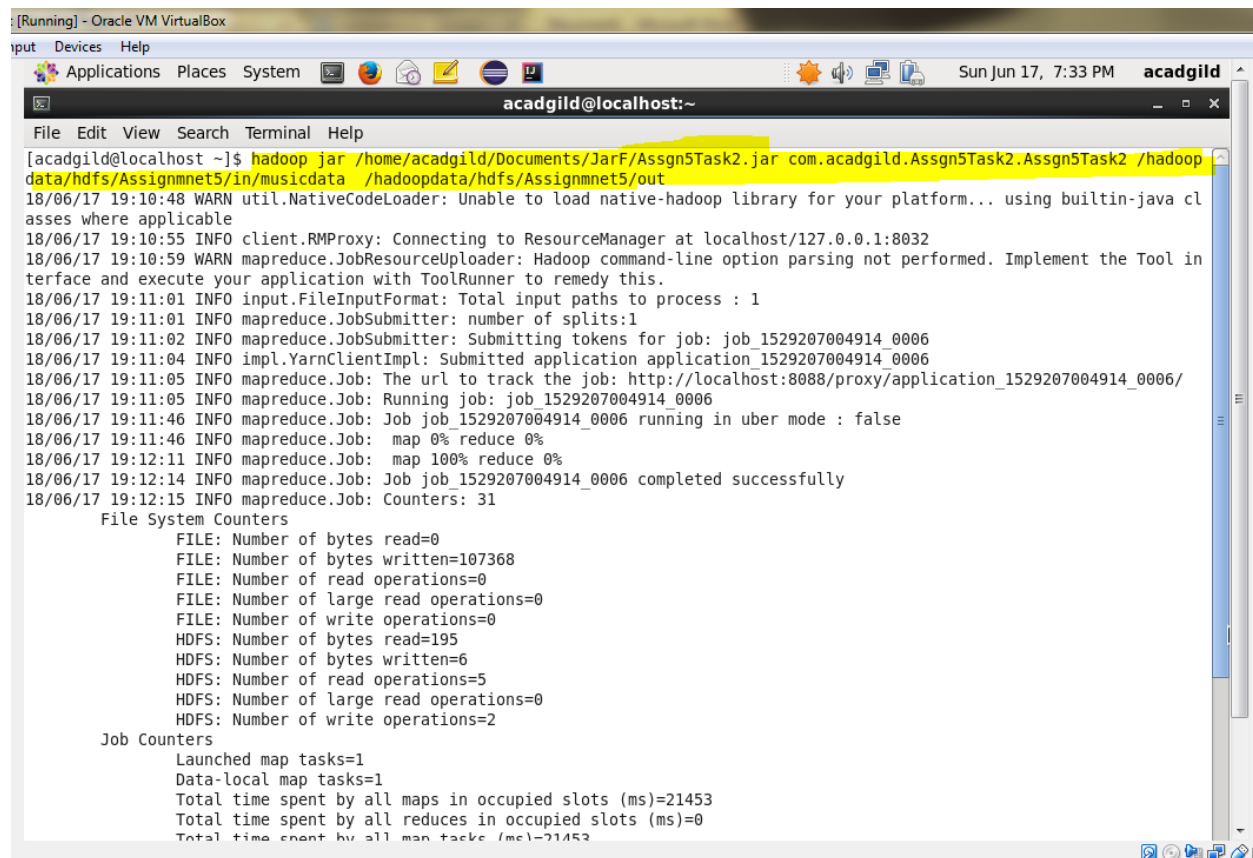
```
[acadgild@localhost ~]$ hadoop fs -mkdir /hadoopdata/hdfs/Assignmnet5/in
18/06/17 18:28:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$ hadoop fs -put /home/acadgild/Downloads/musicdata.txt.docx /hadoopdata/hdfs/Assignmnet5/in
18/06/17 18:30:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignmnet5/in
18/06/17 18:31:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup 6436 2018-06-17 18:30 /hadoopdata/hdfs/Assignmnet5/in/musicdata.txt.docx
```

```
[Running] - Oracle VM VirtualBox
put Devices Help
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Assign5Task1.jar com.acadgild.Assign5Task1.Assign5Task1 /hadoop
data/hdfs/Assignmnet5/in/musicdata /hadoopdata/hdfs/Assignmnet5/out
18/06/17 18:55:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/17 18:55:27 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/17 18:55:33 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/17 18:55:35 INFO input.FileInputFormat: Total input paths to process : 1
18/06/17 18:55:36 INFO mapreduce.JobSubmitter: number of splits:1
18/06/17 18:55:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0005
18/06/17 18:55:40 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0005
18/06/17 18:55:41 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0005/
18/06/17 18:55:41 INFO mapreduce.Job: Running job: job_1529207004914_0005
18/06/17 18:56:29 INFO mapreduce.Job: Job job_1529207004914_0005 running in uber mode : false
18/06/17 18:56:29 INFO mapreduce.Job: map 0% reduce 0%
18/06/17 18:57:16 INFO mapreduce.Job: map 100% reduce 0%
18/06/17 18:57:54 INFO mapreduce.Job: map 100% reduce 100%
18/06/17 18:57:56 INFO mapreduce.Job: Job job_1529207004914_0005 completed successfully
18/06/17 18:57:57 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=58
  FILE: Number of bytes written=215463
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=195
  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)=40627
  Total time spent by all reduces in occupied slots (ms)=33687
  Total time spent by all map tasks (ms)=40627
```

OUTPUT:

```
total committed heap usage (bytes)=159040032
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
com.acadgild.Assgn5Task1.Assgn5Task1$COUNTERS1
  RECORD_COUNTER=2
File Input Format Counters
  Bytes Read=68
File Output Format Counters
  Bytes Written=18
number of unique listeners : 2
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignmnet5/out
18/06/17 18:59:21 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 2018-06-17 18:57 /hadoopdata/hdfs/Assignmnet5/out/_SUCCESS
-rw-r--r--  1 acadgild supergroup        18 2018-06-17 18:57 /hadoopdata/hdfs/Assignmnet5/out/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignmnet5/out/part-r-00000
18/06/17 18:59:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
111113
111117 acadgild
```

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



```
[Running] - Oracle VM VirtualBox
Input Devices Help
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Assgn5Task2.jar com.acadgild.Assgn5Task2.Assgn5Task2 /hadoop
data/hdfs/Assignmnet5/in/musicdata /hadoopdata/hdfs/Assignmnet5/out
18/06/17 19:10:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/17 19:10:55 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/17 19:10:59 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/17 19:11:01 INFO input.FileInputFormat: Total input paths to process : 1
18/06/17 19:11:01 INFO mapreduce.JobSubmitter: number of splits:1
18/06/17 19:11:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0006
18/06/17 19:11:04 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0006
18/06/17 19:11:05 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0006/
18/06/17 19:11:05 INFO mapreduce.Job: Running job: job_1529207004914_0006
18/06/17 19:11:46 INFO mapreduce.Job: Job job_1529207004914_0006 running in uber mode : false
18/06/17 19:11:46 INFO mapreduce.Job: map 0% reduce 0%
18/06/17 19:12:11 INFO mapreduce.Job: map 100% reduce 0%
18/06/17 19:12:14 INFO mapreduce.Job: Job job_1529207004914_0006 completed successfully
18/06/17 19:12:15 INFO mapreduce.Job: Counters: 31
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=107368
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=195
  HDFS: Number of bytes written=6
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=21453
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=21453
```

```

HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=21453
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=21453
  Total vcore-milliseconds taken by all map tasks=21453
  Total megabyte-milliseconds taken by all map tasks=21967872
Map-Reduce Framework
  Map input records=4
  Map output records=1
  Input split bytes=127
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=258
  CPU time spent (ms)=1360
  Physical memory (bytes) snapshot=65679360
  Virtual memory (bytes) snapshot=340971520
  Total committed heap usage (bytes)=16252928
com.acadgild.Assgn5Task2.Assgn5Task2$COUNTERS
  RECORD_COUNTER=1
File Input Format Counters
  Bytes Read=68
File Output Format Counters
  Bytes Written=6
number of times a song was heard fully :1
[acadgild@localhost ~]$

```

OUTPUT:

```

[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignmnet5/out
18/06/17 19:35:00 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 2018-06-17 19:12 /hadoopdata/hdfs/Assignmnet5/out/_SUCCESS
-rw-r--r--  1 acadgild supergroup          6 2018-06-17 19:12 /hadoopdata/hdfs/Assignmnet5/out/part-m-000000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignmnet5/out/part-m-000000
18/06/17 19:35:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
223      1
[acadgild@localhost ~]$

```

TASK3: What are the number of times a song was shared.

```
Running] - Oracle VM VirtualBox
ut Devices Help
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Assgn5Task3.jar com.acadgild.Assgn5Task3.Assgn5Task3 /hadoop
data/hdfs/Assignmnet5/in/musicdata /hadoopdata/hdfs/Assignmnet5/out
18/06/17 19:51:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/17 19:51:34 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/17 19:51:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/17 19:51:41 INFO input.FileInputFormat: Total input paths to process : 1
18/06/17 19:51:42 INFO mapreduce.JobSubmitter: number of splits:1
18/06/17 19:51:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0007
18/06/17 19:51:45 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0007
18/06/17 19:51:45 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0007/
18/06/17 19:51:45 INFO mapreduce.Job: Running job: job_1529207004914_0007
18/06/17 19:52:20 INFO mapreduce.Job: Job job_1529207004914_0007 running in uber mode : false
18/06/17 19:52:20 INFO mapreduce.Job: map 0% reduce 0%
18/06/17 19:52:46 INFO mapreduce.Job: map 100% reduce 0%
18/06/17 19:52:47 INFO mapreduce.Job: Job job_1529207004914_0007 completed successfully
18/06/17 19:52:48 INFO mapreduce.Job: Counters: 31
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=107368
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=195
  HDFS: Number of bytes written=12
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=22539
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=22539
```

OUTPUT:

```
Running] - Oracle VM VirtualBox
ut Devices Help
Launched map tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=22539
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=22539
Total vcore-milliseconds taken by all map tasks=22539
Total megabyte-milliseconds taken by all map tasks=23079936
Map-Reduce Framework
  Map input records=4
  Map output records=2
  Input split bytes=127
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=294
  CPU time spent (ms)=1380
  Physical memory (bytes) snapshot=65437696
  Virtual memory (bytes) snapshot=340971520
  Total committed heap usage (bytes)=16252928
com.acadgild.Assgn5Task3.Assgn5Task3$COUNTERS3
  RECORD_COUNTER=2
File Input Format Counters
  Bytes Read=68
File Output Format Counters
  Bytes Written=12
Number of times a song was shared : 2
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignmnet5/out
18/06/17 19:54:38 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 2018-06-17 19:52 /hadoopdata/hdfs/Assignmnet5/out/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 12 2018-06-17 19:52 /hadoopdata/hdfs/Assignmnet5/out/part-m-000000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignmnet5/out/part-m-000000
18/06/17 19:54:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
225 1
225 1
[acadgild@localhost ~]$
```