

## ASSIGNMENT 5.1

### Input Data Set

acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -cat /user/musicdata.txt
18/02/14 03:27:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

### Task1: Find the no. of unique listners in data set.

acadgild@localhost:~

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/Uniquelistner.jar Task1.Uniquelistner Driver /user/musicdata.txt /Uniquelistner
18/02/14 03:37:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/02/14 03:37:34 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/02/14 03:37:34 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
oolRunner to remedy this.
18/02/14 03:37:35 INFO input.FileInputFormat: Total input paths to process : 1
18/02/14 03:37:35 INFO mapreduce.JobSubmitter: number of splits:1
18/02/14 03:37:35 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1518445375024_0077
18/02/14 03:37:35 INFO impl.YarnClientImpl: Submitted application application_1518445375024_0077
18/02/14 03:37:35 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1518445375024_0077/
18/02/14 03:37:35 INFO mapreduce.Job: Running job: job_1518445375024_0077
18/02/14 03:37:41 INFO mapreduce.Job: Job job_1518445375024_0077 running in uber mode : false
18/02/14 03:37:41 INFO mapreduce.Job: map 0% reduce 0%
18/02/14 03:37:48 INFO mapreduce.Job: map 100% reduce 0%
18/02/14 03:37:54 INFO mapreduce.Job: map 100% reduce 100%
18/02/14 03:37:54 INFO mapreduce.Job: Job job_1518445375024_0077 completed successfully
18/02/14 03:37:54 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=46
    FILE: Number of bytes written=215445
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=175
    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)=3310
    Total time spent by all reduces in occupied slots (ms)=3546
    Total time spent by all map tasks (ms)=3310
    Total time spent by all reduce tasks (ms)=3546
    Total vcore-milliseconds taken by all map tasks=3310
    Total vcore-milliseconds taken by all reduce tasks=3546
    Total megabyte-milliseconds taken by all map tasks=3389440
    Total megabyte-milliseconds taken by all reduce tasks=3631104
  Map-Reduce Framework
    Map input records=4
    Map output records=4
```

acadgild@localhost:~

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=3310
Total time spent by all reduces in occupied slots (ms)=3546
Total time spent by all map tasks (ms)=3310
Total time spent by all reduce tasks (ms)=3546
Total vcore-milliseconds taken by all map tasks=3310
Total vcore-milliseconds taken by all reduce tasks=3546
Total megabyte-milliseconds taken by all map tasks=3389440
Total megabyte-milliseconds taken by all reduce tasks=3631104
Map-Reduce Framework
  Map input records=4
  Map output records=4
  Map output bytes=32
  Map output materialized bytes=46
  Input split bytes=105
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=46
  Reduce input records=4
  Reduce output records=3
  Spilled Records=8
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=149
  CPU time spent (ms)=1560
  Physical memory (bytes) snapshot=416718848
  Virtual memory (bytes) snapshot=4164886528
  Total committed heap usage (bytes)=311427072
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=70
File Output Format Counters
  Bytes Written=18
```

## Output:-

acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -ls /UniqueListner/
18/02/14 03:50:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-02-14 03:37 /UniqueListner/_SUCCESS
-rw-r--r--  1 acadgild supergroup        18 2018-02-14 03:37 /UniqueListner/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /UniqueListner/part-r-00000
18/02/14 03:50:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
222      1
223      1
225      2
[acadgild@localhost ~]$
```

## Task2: Find the no. of time a song was heardfully.

```
acadgild@localhost:~$
[acadgild@localhost ~]$
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/SongListen.jar Task2.SongListen Driver /user/musicdata.txt /SongHeard
18/02/14 03:43:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
18/02/14 03:43:04 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/02/14 03:43:05 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
oolRunner to remedy this.
18/02/14 03:43:05 INFO input.FileInputFormat: Total input paths to process : 1
18/02/14 03:43:06 INFO mapreduce.JobSubmitter: number of splits:1
18/02/14 03:43:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1518445375024_0078
18/02/14 03:43:06 INFO impl.YarnClientImpl: Submitted application application_1518445375024_0078
18/02/14 03:43:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1518445375024_0078/
18/02/14 03:43:06 INFO mapreduce.Job: Running job: job_1518445375024_0078
18/02/14 03:43:13 INFO mapreduce.Job: Job job_1518445375024_0078 running in uber mode : false
18/02/14 03:43:13 INFO mapreduce.Job: map 0% reduce 0%
18/02/14 03:43:18 INFO mapreduce.Job: map 100% reduce 0%
18/02/14 03:43:24 INFO mapreduce.Job: map 100% reduce 100%
18/02/14 03:43:24 INFO mapreduce.Job: Job job_1518445375024_0078 completed successfully
18/02/14 03:43:25 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=56
      FILE: Number of bytes written=215457
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=175
      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)=3229
      Total time spent by all reduces in occupied slots (ms)=3521
      Total time spent by all map tasks (ms)=3229
      Total time spent by all reduce tasks (ms)=3521
      Total vcore-milliseconds taken by all map tasks=3229
      Total vcore-milliseconds taken by all reduce tasks=3521
      Total megabyte-milliseconds taken by all map tasks=3306496
      Total megabyte-milliseconds taken by all reduce tasks=3605504
    Map-Reduce Framework
      Map input records=4
```

```
      Total megabyte-milliseconds taken by all reduce tasks=3605504
    Map-Reduce Framework
      Map input records=4
      Map output records=5
      Map output bytes=40
      Map output materialized bytes=56
      Input split bytes=105
      Combine input records=0
      Combine output records=0
      Reduce input groups=3
      Reduce shuffle bytes=56
      Reduce input records=5
      Reduce output records=3
      Spilled Records=10
      Shuffled Maps =1
      Failed Shuffles=0
      Merged Map outputs=1
      GC time elapsed (ms)=136
      CPU time spent (ms)=1490
      Physical memory (bytes) snapshot=418312192
      Virtual memory (bytes) snapshot=4167532544
      Total committed heap usage (bytes)=325058560
    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=70
    File Output Format Counters
      Bytes Written=18
```

## Output:-

```
acadgild@localhost:~  
[acadgild@localhost ~]$ hdfs dfs -ls /SongHeard  
18/02/14 03:51:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u  
Found 2 items  
-rw-r--r-- 1 acadgild supergroup 0 2018-02-14 03:43 /SongHeard/_SUCCESS  
-rw-r--r-- 1 acadgild supergroup 18 2018-02-14 03:43 /SongHeard/part-r-00000  
[acadgild@localhost ~]$ hdfs dfs -cat /SongHeard/part-r-00000  
18/02/14 03:51:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u  
222 0  
223 1  
225 0  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```

## Task3: Find the no. of times a song was shared.

```
acadgild@localhost:~  
[acadgild@localhost ~]$  
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/SongShare.jar Task2.SongShared_Driver /user/musicdata.txt /SongShare  
Exception in thread "main" java.lang.ClassNotFoundException: Task2.SongShared_Driver  
    at java.net.URLClassLoader.findClass(URLClassLoader.java:381)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)  
    at java.lang.Class.forName0(Native Method)  
    at java.lang.Class.forName(Class.java:348)  
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)  
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$  
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/SongShare.jar Task3.SongShared_Driver /user/musicdata.txt /SongShare  
18/02/14 03:46:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where  
18/02/14 03:46:49 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/02/14 03:46:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface an  
oolRunner to remedy this.  
18/02/14 03:46:50 INFO input.FileInputFormat: Total input paths to process : 1  
18/02/14 03:46:50 INFO mapreduce.JobSubmitter: number of splits:1  
18/02/14 03:46:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1518445375024_0079  
18/02/14 03:46:50 INFO impl.YarnClientImpl: Submitted application application_1518445375024_0079  
18/02/14 03:46:50 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1518445375024_0079/  
18/02/14 03:46:50 INFO mapreduce.Job: Running job: job_1518445375024_0079  
18/02/14 03:46:58 INFO mapreduce.Job: Job job_1518445375024_0079 running in uber mode : false  
18/02/14 03:46:58 INFO mapreduce.Job: map 0% reduce 0%  
18/02/14 03:47:04 INFO mapreduce.Job: map 100% reduce 0%  
18/02/14 03:47:10 INFO mapreduce.Job: map 100% reduce 100%  
18/02/14 03:47:10 INFO mapreduce.Job: Job job_1518445375024_0079 completed successfully  
18/02/14 03:47:10 INFO mapreduce.Job: Counters: 49  
File System Counters  
    FILE: Number of bytes read=66  
    FILE: Number of bytes written=215479  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=175  
    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
```

acadgild@localhost:~

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=3343
Total time spent by all reduces in occupied slots (ms)=3550
Total time spent by all map tasks (ms)=3343
Total time spent by all reduce tasks (ms)=3550
Total vcore-milliseconds taken by all map tasks=3343
Total vcore-milliseconds taken by all reduce tasks=3550
Total megabyte-milliseconds taken by all map tasks=3423232
Total megabyte-milliseconds taken by all reduce tasks=3635200
Map-Reduce Framework
  Map input records=4
  Map output records=6
  Map output bytes=48
  Map output materialized bytes=66
  Input split bytes=105
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=66
  Reduce input records=6
  Reduce output records=3
  Spilled Records=12
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=147
  CPU time spent (ms)=1650
  Physical memory (bytes) snapshot=416935936
  Virtual memory (bytes) snapshot=4165120000
  Total committed heap usage (bytes)=310902784
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=70
File Output Format Counters
  Bytes Written=18
```

## Output:-

acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -ls /SongShare/
18/02/14 03:49:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-02-14 03:47 /SongShare/_SUCCESS
-rw-r--r--  1 acadgild supergroup        18 2018-02-14 03:47 /SongShare/part-r-000000
[acadgild@localhost ~]$ hdfs dfs -cat /SongShare/part-r-000000
18/02/14 03:49:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
222      0
223      0
225      2
[acadgild@localhost ~]$
```