

Assignment 5 Advance MapReduce

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

Command

```
hadoop jar assign5.jar mrTask1.Task1 /mapreduce/musicdata.txt  
/mapreduce/assign5/task1output
```

Code Snippet

```
[acadgild@localhost mapreduce]$ hadoop jar assign5.jar mrTask1.Task1 /mapreduce/musicdata.txt /mapreduce/assign5/task1output  
19/01/11 18:29:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/11 18:29:12 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
19/01/11 18:29:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with  
ToolRunner to remedy this.  
19/01/11 18:29:15 INFO input.FileInputFormat: Total input paths to process : 1  
19/01/11 18:29:15 INFO mapreduce.JobSubmitter: number of splits:1  
19/01/11 18:29:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547166974239_0005  
19/01/11 18:29:17 INFO impl.YarnClientImpl: Submitted application application_1547166974239_0005  
19/01/11 18:29:17 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547166974239_0005/  
19/01/11 18:29:17 INFO mapreduce.Job: Running job: job_1547166974239_0005  
19/01/11 18:29:34 INFO mapreduce.Job: Job job_1547166974239_0005 running in uber mode : false  
19/01/11 18:29:34 INFO mapreduce.Job: map 0% reduce 0%  
19/01/11 18:29:49 INFO mapreduce.Job: map 100% reduce 0%  
19/01/11 18:30:04 INFO mapreduce.Job: map 100% reduce 100%  
19/01/11 18:30:05 INFO mapreduce.Job: Job job_1547166974239_0005 completed successfully  
19/01/11 18:30:06 INFO mapreduce.Job: Counters: 49  
File System Counters  
FILE: Number of bytes read=46  
FILE: Number of bytes written=215459  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=178  
HDFS: Number of bytes written=18  
HDFS: Number of read operations=0  
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)=11671  
Total time spent by all reduces in occupied slots (ms)=12863  
Total time spent by all map tasks (ms)=11671  
Total time spent by all reduce tasks (ms)=12863  
Total vcore-milliseconds taken by all map tasks=11671  
Total vcore-milliseconds taken by all reduce tasks=12863  
Total megabyte-milliseconds taken by all map tasks=11951104  
Total megabyte-milliseconds taken by all reduce tasks=13171712  
Map-Reduce Framework
```

```
Map-Reduce Framework  
Map input records=4  
Map output records=4  
Map output bytes=32  
Map output materialized bytes=46  
Input split bytes=110  
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)=322  
CPU time spent (ms)=3710  
Physical memory (bytes) snapshot=322064384  
Virtual memory (bytes) snapshot=4118233088  
Total committed heap usage (bytes)=222429184  
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 mapreduce]$
```

Output screen shot

```
[acadgild@localhost ~]$ hdfs dfs -cat /mapreduce/assign5/task1output/part-r-00000
19/01/11 18:38:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library f
222      1
223      1
225      2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

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

Command

```
hadoop jar assign5.jar mrTask2.Task2 /mapreduce/musicdata.txt
/mapreduce/assign5/task2output
```

Code Snippet

```
[acadgild@localhost mapreduce]$ hadoop jar assign5.jar mrTask2.Task2 /mapreduce/musicdata.txt /mapreduce/assign5/task2output
19/01/11 18:33:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/11 18:33:14 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/11 18:33:17 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
19/01/11 18:33:18 INFO input.FileInputFormat: Total input paths to process : 1
19/01/11 18:33:18 INFO mapreduce.JobSubmitter: number of splits:1
19/01/11 18:33:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547166974239_0006
19/01/11 18:33:19 INFO impl.YarnClientImpl: Submitted application application_1547166974239_0006
19/01/11 18:33:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547166974239_0006/
19/01/11 18:33:19 INFO mapreduce.Job: Running job: job_1547166974239_0006
19/01/11 18:33:40 INFO mapreduce.Job: Job job_1547166974239_0006 running in uber mode : false
19/01/11 18:33:40 INFO mapreduce.Job: map 0% reduce 0%
19/01/11 18:33:52 INFO mapreduce.Job: map 100% reduce 0%
19/01/11 18:34:06 INFO mapreduce.Job: map 100% reduce 100%
19/01/11 18:34:07 INFO mapreduce.Job: Job job_1547166974239_0006 completed successfully
19/01/11 18:34:08 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215491
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=178
    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)=10181
    Total time spent by all reduces in occupied slots (ms)=11590
    Total time spent by all map tasks (ms)=10181
    Total time spent by all reduce tasks (ms)=11590
    Total vcore-milliseconds taken by all map tasks=10181
    Total vcore-milliseconds taken by all reduce tasks=11590
    Total megabyte-milliseconds taken by all map tasks=10425344
    Total megabyte-milliseconds taken by all reduce tasks=11868160
  Map-Reduce Framework
```

```

Total megabyte-milliseconds taken by all reduce tasks=11868160
Map-Reduce Framework
  Map input records=4
  Map output records=5
  Map output bytes=40
  Map output materialized bytes=56
  Input split bytes=110
  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)=324
  CPU time spent (ms)=3100
  Physical memory (bytes) snapshot=323792896
  Virtual memory (bytes) snapshot=4118183936
  Total committed heap usage (bytes)=222429184
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 mapreduce]$

```

Output screen shot

```

[acadgild@localhost ~]$ hdfs dfs -cat /mapreduce/assign5/task2output/part-r-000000
19/01/11 18:39:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
222      0
223      1
225      0
[acadgild@localhost ~]$

```

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

Command

```

hadoop jar assign5.jar mrTask3.Task3 /mapreduce/musicdata.txt
/mapreduce/assign5/task3output

```

Code Snippet

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop jar assign5.jar mrTask3.Task3 /mapreduce/musicdata.txt /mapreduce/assign5/task3output
19/01/11 18:35:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/11 18:35:42 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/11 18:35:44 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
19/01/11 18:35:45 INFO input.FileInputFormat: Total input paths to process : 1
19/01/11 18:35:45 INFO mapreduce.JobSubmitter: number of splits:1
19/01/11 18:35:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1547166974239_0007
19/01/11 18:35:46 INFO impl.YarnClientImpl: Submitted application application_1547166974239_0007
19/01/11 18:35:46 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1547166974239_0007/
19/01/11 18:35:46 INFO mapreduce.Job: Running job: job_1547166974239_0007
19/01/11 18:36:04 INFO mapreduce.Job: Job job_1547166974239_0007 running in uber mode : false
19/01/11 18:36:04 INFO mapreduce.Job:  map 0% reduce 0%
19/01/11 18:36:16 INFO mapreduce.Job:  map 100% reduce 0%
19/01/11 18:36:30 INFO mapreduce.Job:  map 100% reduce 100%
19/01/11 18:36:31 INFO mapreduce.Job: Job job_1547166974239_0007 completed successfully
19/01/11 18:36:32 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=66
  FILE: Number of bytes written=215513
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=178
  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)=9760
  Total time spent by all reduces in occupied slots (ms)=11748
  Total time spent by all map tasks (ms)=9760
  Total time spent by all reduce tasks (ms)=11748
  Total vcore-milliseconds taken by all map tasks=9760
  Total vcore-milliseconds taken by all reduce tasks=11748
  Total megabyte-milliseconds taken by all map tasks=9994240
  Total megabyte-milliseconds taken by all reduce tasks=12029952
Map-Reduce Framework
```

```
Total megabyte-milliseconds taken by all map tasks=9994240
Total megabyte-milliseconds taken by all reduce tasks=12029952
Map-Reduce Framework
  Map input records=4
  Map output records=6
  Map output bytes=48
  Map output materialized bytes=66
  Input split bytes=110
  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)=288
  CPU time spent (ms)=2990
  Physical memory (bytes) snapshot=320487424
  Virtual memory (bytes) snapshot=4118192128
  Total committed heap usage (bytes)=222429184
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=60
File Output Format Counters
  Bytes Written=18
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Output screen shot

```
[acadgild@localhost ~]$ hdfs dfs -cat /mapreduce/assign5/task3output/part-r-00000
19/01/11 18:39:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library
222    0
223    0
225    2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

