

Task 1:

Find the number of unique listeners in the data set.

- Dataset is having 4 Records out of which 3 are unique user ids.

Solution Approach

- As we have to just fetch unique listeners i.e. nothing but a keys which can produced through Mappers we can have NullWritable Value and Text as Key
- NullWritable is a special type of Writable, as it has a zero-length serialization. No bytes are written to or read from the stream. It is used as a placeholder; for example, in MapReduce, a key or a value can be declared as a NullWritable when you don't need to use that position, effectively storing a constant empty value.

```
[acadgild@localhost ~]$ hadoop fs -mkdir /Assignment5_Input
18/03/30 19:17:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hadoop fs -put /home/acadgild/Desktop/Prachi/Assignment5/MusicData.txt /Assignment5_Input/MusicData.txt
18/03/30 19:17:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Input/MusicData.txt
18/03/30 19:18:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Input/MusicData.txt
18/03/31 03:21:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0[acadgild@localhost ~]$
```

Mapper

```
1 SongShared.java 2 SongSharedMapper.java 3 SongSharedReducer.java 4 SongHeardFullyMapper.java 5 SongHeardFullyReducer.java 6 UniqueListenersMapper.java 7 UniqueListenersReducer.java
1 *import java.io.IOException;
2
3
4 public class UniqueListenersMapper extends Mapper<LongWritable, Text, Text, NullWritable> {
5     private Text listenerID ;
6     @Override
7     public void map(LongWritable key, Text value, Context context) throws IOException , InterruptedException
8     {
9         String rowDetails = value.toString();
10        String[] parts = rowDetails.split("\\|");
11        listenerID = new Text(parts[0]);
12        context.write(listenerID, NullWritable.get());
13    }
14 }
15
16 }
```

Reducer

```

1 import java.io.IOException;
9
10 public class UniqueListernsReducer extends Reducer<Text, NullWritable, Text, NullWritable> {
11
12     @Override
13     public void reduce(Text key, Iterable<NullWritable> values, Context context) throws IOException, Interrupte
14     {
15         context.write(key, NullWritable.get());
16     }
17
18 }
19

```

Execution of Jar File

```

[acdgil@localhost ~]$ hadoop jar /home/acadgild/Desktop/Prachi/Assignment5.jar UniqueListers /Assignment5_Input /Assignment5_Output
18/03/30 19:20:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/30 19:20:12 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/30 19:20:17 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
h ToolRunner to remedy this.
18/03/30 19:20:20 INFO input.FileInputFormat: Total input paths to process : 1
18/03/30 19:20:20 INFO mapreduce.JobSubmitter: number of splits:1
18/03/30 19:20:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522393135360_0005
18/03/30 19:20:27 INFO impl.YarnClientImpl: Submitted application application_1522393135360_0005
18/03/30 19:20:30 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522393135360_0005/
18/03/30 19:20:30 INFO mapreduce.Job: Running job: job_1522393135360_0005
18/03/30 19:21:34 INFO mapreduce.Job: Job job_1522393135360_0005 running in uber mode : false
18/03/30 19:21:34 INFO mapreduce.Job: map 0% reduce 0%
18/03/30 19:22:22 INFO mapreduce.Job: map 100% reduce 0%
18/03/30 19:22:48 INFO mapreduce.Job: map 100% reduce 100%
18/03/30 19:22:49 INFO mapreduce.Job: Job job_1522393135360_0005 completed successfully
18/03/30 19:22:50 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=42
    FILE: Number of bytes written=215679
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=188
    HDFS: Number of bytes written=21
    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)=44546
    Total time spent by all reduces in occupied slots (ms)=19411
    Total time spent by all map tasks (ms)=44546
    Total time spent by all reduce tasks (ms)=19411
    Total vcore-milliseconds taken by all map tasks=44546
    Total vcore-milliseconds taken by all reduce tasks=19411
    Total megabyte-milliseconds taken by all map tasks=45615104
    Total megabyte-milliseconds taken by all reduce tasks=19876864
  Map-Reduce Framework
    Map input records=4
    Map output records=4
    Map output bytes=28
    Map output materialized bytes=42

```

```

Map output bytes=28
Map output materialized bytes=42
Input split bytes=118
Combine input records=0
Combine output records=0
Reduce input groups=3
Reduce shuffle bytes=42
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)=960
CPU time spent (ms)=3930
Physical memory (bytes) snapshot=285585408
Virtual memory (bytes) snapshot=4118200320
Total committed heap usage (bytes)=170004480
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=21
You have new mail in /var/spool/mail/acadgild

```

Output Return As – Three unique Records

```

[acadgild@localhost ~]$ hadoop fs -ls /
18/03/30 19:25:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 19 items
drwxr-xr-x - acadgild supergroup 0 2018-03-30 19:17 /Assignment5_Input
drwxr-xr-x - acadgild supergroup 0 2018-03-30 19:22 /Assignment5_Output1
drwxr-xr-x - acadgild supergroup 0 2018-03-04 22:22 /Prachi
drwxr-xr-x - acadgild supergroup 0 2018-03-24 13:09 /assignment4_input
drwxr-xr-x - acadgild supergroup 0 2018-03-24 14:34 /assignment4_output1
drwxr-xr-x - acadgild supergroup 0 2018-03-24 14:29 /assignment4_output1[A
drwxr-xr-x - acadgild supergroup 0 2018-03-30 12:50 /assignment4_output2
drwxr-xr-x - acadgild supergroup 0 2018-03-30 13:12 /assignment4_output3
drwxr-xr-x - acadgild supergroup 0 2018-03-24 09:38 /hadoopdata
drwxr-xr-x - acadgild supergroup 0 2018-03-30 15:19 /hbase
drwxr-xr-x - acadgild supergroup 0 2018-03-04 19:15 /input
drwxr-xr-x - acadgild supergroup 0 2018-03-05 09:06 /output
drwxr-xr-x - acadgild supergroup 0 2018-03-24 12:22 /output1
drwxr-xr-x - acadgild supergroup 0 2018-03-24 12:32 /output2
drwxr-xr-x - acadgild supergroup 0 2018-03-24 12:37 /output3
drwxr-xr-x - acadgild supergroup 0 2018-03-17 09:59 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup 0 2018-03-05 08:40 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-03-18 07:10 /user
[acadgild@localhost ~]$ hadoop fs -ls /Assignment5_Output1
18/03/30 19:25:48 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 2018-03-30 19:22 /Assignment5_Output1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 21 2018-03-30 19:22 /Assignment5_Output1/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Output1/part-r-00000
18/03/30 19:26:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113
111115
111117
[acadgild@localhost ~]$

```

Task 2

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

- Dataset is having only one song which is fully heard

Solution Approach

- Here irrespective of Key value (User ID / Song ID), we should be getting number of songs fully heard.
- To achieve the same we can have below approaches
 - o Having a constant key
 - o Having NullWritable key (which is constant Empty key)
 - o Having a Mapper where cleanup method is written to get count of fully heard songs. As if we have more Constant / NullWritable keys it can make Reducer very slow. So executing / having only one output of a mapper with constant / NullWritable key can reduce load on reducer
 - Having less keys can be achieved through having combiners for every mapper and then passing output of mini-mapper (combiner) to Reducer

For this task we are following the approach of

- o Having a constant key as "One"

Mapper

```
1 import java.io.IOException;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.LongWritable;
5 import org.apache.hadoop.io.NullWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Mapper;
8 import org.apache.hadoop.mapreduce.Mapper.Context;
9
10 public class SongHeardFullyMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
11     private IntWritable HeardFully;
12     private Text keyValue = new Text("One");
13     @Override
14     public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException
15     {
16         String rowDetails = value.toString();
17         String[] parts = rowDetails.split("\\|");
18         if (parts[4].equalsIgnoreCase("1"))
19         {
20             HeardFully = new IntWritable(new Integer(parts[4]));
21             context.write(keyValue, HeardFully);
22         }
23     }
24 }
25 }
26 }
```

Reducer

```

1*import java.io.IOException;
7
8 public class SongHeardFullyReducer extends Reducer<Text, IntWritable, Text, Text> {
9     private int heardcount = 0;
10    private Text OutputofHeard ;
11    @Override
12    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException , InterruptedException
13    {
14
15        for (IntWritable value : values) {
16            heardcount+=value.get();
17        }
18        OutputofHeard = new Text("fully Heard Songs Count " + Integer.toString(heardcount));
19        context.write(key, OutputofHeard);
20    }
21 }
22 }
23

```

Running the Jar to get number of songs heard fully

```

[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/Prachi/Assignment5.jar SongHeardFully /Assignment5_Input /Assignment5_Output2
18/03/30 20:35:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/30 20:35:02 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/30 20:35:04 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
h ToolRunner to remedy this.
18/03/30 20:35:05 INFO input.FileInputFormat: Total input paths to process : 1
18/03/30 20:35:05 INFO mapreduce.JobSubmitter: number of splits:1
18/03/30 20:35:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522393135360_0011
18/03/30 20:35:06 INFO impl.YarnClientImpl: Submitted application application_1522393135360_0011
18/03/30 20:35:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522393135360_0011/
18/03/30 20:35:06 INFO mapreduce.Job: Running job: job_1522393135360_0011
18/03/30 20:35:23 INFO mapreduce.Job: Job job_1522393135360_0011 running in uber mode : false
18/03/30 20:35:23 INFO mapreduce.Job: map 0% reduce 0%
18/03/30 20:35:35 INFO mapreduce.Job: map 100% reduce 0%
18/03/30 20:35:49 INFO mapreduce.Job: map 100% reduce 100%
18/03/30 20:35:50 INFO mapreduce.Job: Job job_1522393135360_0011 completed successfully
18/03/30 20:35:50 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=16
  FILE: Number of bytes written=215627
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=188
  HDFS: Number of bytes written=30
  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)=9205
  Total time spent by all reduces in occupied slots (ms)=10927
  Total time spent by all map tasks (ms)=9205
  Total time spent by all reduce tasks (ms)=10927
  Total vcore-milliseconds taken by all map tasks=9205
  Total vcore-milliseconds taken by all reduce tasks=10927
  Total megabyte-milliseconds taken by all map tasks=9425920
  Total megabyte-milliseconds taken by all reduce tasks=11189248
Map-Reduce Framework
  Map input records=4
  Map output records=1
  Map output bytes=8
  Map output materialized bytes=16
  Input split bytes=118

```

```

  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=16
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=274
  CPU time spent (ms)=2820
  Physical memory (bytes) snapshot=295092224
  Virtual memory (bytes) snapshot=4118192128
  Total committed heap usage (bytes)=170004480
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=30
[acadgild@localhost ~]$ hadoop fs -ls /Assignment5_Output2
18/03/30 20:36:09 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 2018-03-30 20:35 /Assignment5_Output2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 30 2018-03-30 20:35 /Assignment5_Output2/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Output2/part-r-00000
18/03/30 20:36:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
One fully Heard Songs Count 1
[acadgild@localhost ~]$

```

Task 3

What are the number of times a song was shared

- This dataset has 2 songs which are shared

Solution Approach

- As discussed above this can be implemented in different ways
- Followed two Approaches

Approach A –

- Having NullWritable key (which is constant Empty key)

Mapper Class

```
1*import java.io.IOException;
7
8 public class SongHeardFullyReducer extends Reducer<Text, IntWritable, Text, Text> {
9     private int heardcount = 0;
10    private Text OutputofHeard ;
11    @Override
12    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException , Interrupte
13    {
14
15        for (IntWritable value : values) {
16            heardcount+=value.get();
17        }
18        OutputofHeard = new Text("fully Heard Songs Count " + Integer.toString(heardcount));
19        context.write(key, OutputofHeard);
20    }
21
22 }
```

Reducer Class

```

1=import java.io.IOException;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.NullWritable;
5 import org.apache.hadoop.io.Text;
6 import org.apache.hadoop.mapreduce.Reducer;
7 import org.apache.hadoop.mapreduce.Reducer.Context;
8
9 public class SongSharedReducer extends Reducer<NullWritable, IntWritable, NullWritable, Text> {
10     private int heardcount = 0;
11     private Text OutputofHeard ;
12
13     @Override
14     public void reduce(NullWritable key, Iterable<IntWritable> values, Context context) throws IOException , InterruptedException {
15
16         for (IntWritable value : values) {
17             heardcount+=value.get();
18         }
19         OutputofHeard = new Text("Song shared for " + Integer.toString(heardcount));
20         context.write(NullWritable.get(), OutputofHeard);
21     }
22 }
23 }

```

Running Jar file to get the output

```

[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/Prachi/Assignment5.jar SongShared /Assignment5_input /Assignment5_output
18/03/30 20:50:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/30 20:50:29 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/30 20:50:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
  h ToolRunner to remedy this.
18/03/30 20:50:33 INFO InputFileInputFormat: Total input paths to process : 1
18/03/30 20:50:33 INFO mapreduce.JobSubmitter: number of splits:1
18/03/30 20:50:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522393135360_0013
18/03/30 20:50:35 INFO impl.YarnClientImpl: Submitted application application_1522393135360_0013
18/03/30 20:50:35 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522393135360_0013/
18/03/30 20:50:35 INFO mapreduce.Job: Running job: job_1522393135360_0013
18/03/30 20:50:53 INFO mapreduce.Job: Job job_1522393135360_0013 running in uber mode : false
18/03/30 20:50:53 INFO mapreduce.Job:  map 0% reduce 0%
18/03/30 20:51:17 INFO mapreduce.Job:  map 100% reduce 0%
18/03/30 20:51:42 INFO mapreduce.Job:  map 100% reduce 100%
18/03/30 20:51:44 INFO mapreduce.Job: Job job_1522393135360_0013 completed successfully
18/03/30 20:51:45 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=18
    FILE: Number of bytes written=215631
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=188
    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)=21645
    Total time spent by all reduces in occupied slots (ms)=22236
    Total time spent by all map tasks (ms)=21645
    Total time spent by all reduce tasks (ms)=22236
    Total vcore-milliseconds taken by all map tasks=21645
    Total vcore-milliseconds taken by all reduce tasks=22236
    Total megabyte-milliseconds taken by all map tasks=22164480
    Total megabyte-milliseconds taken by all reduce tasks=22769664
  Map-Reduce Framework
    Map input records=4
    Map output records=2

```

```

Map output bytes=8
Map output materialized bytes=18
Input split bytes=118
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=18
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=286
CPU time spent (ms)=3560
Physical memory (bytes) snapshot=287334400
Virtual memory (bytes) snapshot=4118192128
Total committed heap usage (bytes)=170004480

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
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /Assignment5_Output3
18/03/30 20:52:01 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 2018-03-30 20:51 /Assignment5_Output3/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2018-03-30 20:51 /Assignment5_Output3/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Output3/part-r-000000
18/03/30 20:52:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Song shared for 2
[acadgild@localhost ~]$

```

Approach B –

Having a Mapper where cleanup method is written to get count of fully heard songs. As if we have more Constant / NullWritable keys it can make Reducer very slow. So executing / having only one output of a mapper with constant / NullWritable key can reduce load on reducer

Mapper

```

1 import java.io.IOException;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.LongWritable;
5 import org.apache.hadoop.io.NullWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Mapper;
8 import org.apache.hadoop.mapreduce.Mapper.Context;
9
10 public class SongShared_BMapper extends Mapper<LongWritable, Text, NullWritable, IntWritable> {
11     private IntWritable HeardFully = new IntWritable();
12     int heardCount = 0;
13     @Override
14     public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException
15     {
16         String rowDetails = value.toString();
17         String[] parts = rowDetails.split("\\|");
18         if (parts[3].equalsIgnoreCase("1"))
19         {
20             heardCount++;
21         }
22     }
23 }
24
25 @Override
26 public void cleanup(Context context) throws IOException, InterruptedException
27 {
28     HeardFully.set(heardCount);
29     context.write(NullWritable.get(), HeardFully);
30 }
31
32 }
33

```


Reducer

```
1 import java.io.IOException;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.NullWritable;
5 import org.apache.hadoop.io.Text;
6 import org.apache.hadoop.mapreduce.Reducer;
7 import org.apache.hadoop.mapreduce.Reducer.Context;
8
9 public class SongSharedReducer extends Reducer<NullWritable, IntWritable, NullWritable, Text> {
10     private int heardcount = 0;
11     private Text OutputofHeard ;
12     @Override
13     public void reduce(NullWritable key, Iterable<IntWritable> values, Context context) throws IOException , Ir
14     {
15
16         for (IntWritable value : values) {
17             heardcount+=value.get();
18         }
19         OutputofHeard = new Text("Song shared for " + Integer.toString(heardcount));
20         context.write(NullWritable.get(), OutputofHeard);
21     }
22 }
23 }
```

Running Jar to get output

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/Prachi/Assignment5.jar SongShared_B /Assignment5_Input /Assignment5_Output4
18/03/31 04:07:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/31 04:07:51 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/31 04:07:58 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application wit
h ToolRunner to remedy this.
18/03/31 04:08:06 INFO input.FileInputFormat: Total input paths to process : 1
18/03/31 04:08:07 INFO mapreduce.JobSubmitter: number of splits:1
18/03/31 04:08:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522393135360_0014
18/03/31 04:08:14 INFO impl.YarnClientImpl: Submitted application application_1522393135360_0014
18/03/31 04:08:14 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522393135360_0014/
18/03/31 04:08:14 INFO mapreduce.Job: Running job: job_1522393135360_0014
18/03/31 04:09:20 INFO mapreduce.Job: Job job_1522393135360_0014 running in uber mode : false
18/03/31 04:09:28 INFO mapreduce.Job: map 0% reduce 0%
18/03/31 04:09:59 INFO mapreduce.Job: map 100% reduce 0%
18/03/31 04:10:24 INFO mapreduce.Job: map 100% reduce 100%
18/03/31 04:10:26 INFO mapreduce.Job: Job job_1522393135360_0014 completed successfully
18/03/31 04:10:26 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=12
    FILE: Number of bytes written=215623
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=188
    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)=36356
    Total time spent by all reduces in occupied slots (ms)=21022
    Total time spent by all map tasks (ms)=36356
    Total time spent by all reduce tasks (ms)=21022
    Total vcore-milliseconds taken by all map tasks=36356
    Total vcore-milliseconds taken by all reduce tasks=21022
    Total megabyte-milliseconds taken by all map tasks=37228544
    Total megabyte-milliseconds taken by all reduce tasks=21526528
  Map-Reduce Framework
    Map input records=4
    Map output records=1
    Map output bytes=4
    Map output materialized bytes=12
    Input split bytes=118
    Combine input records=0
```

```

    Map output records=1
    Map output bytes=4
    Map output materialized bytes=12
    Input split bytes=116
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=12
    Reduce input records=1
    Reduce output records=1
    Spilled Records=2
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=451
    CPU time spent (ms)=3660
    Physical memory (bytes) snapshot=287158272
    Virtual memory (bytes) snapshot=4118192128
    Total committed heap usage (bytes)=170004480
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
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /
18/03/31 04:11:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 22 items
drwxr-xr-x - acadgild supergroup          0 2018-03-30 19:17 /Assignment5_Input
drwxr-xr-x - acadgild supergroup          0 2018-03-30 19:22 /Assignment5_Output1
drwxr-xr-x - acadgild supergroup          0 2018-03-30 20:35 /Assignment5_Output2
drwxr-xr-x - acadgild supergroup          0 2018-03-30 20:51 /Assignment5_Output3
drwxr-xr-x - acadgild supergroup          0 2018-03-31 04:10 /Assignment5_Output4
drwxr-xr-x - acadgild supergroup          0 2018-03-04 22:22 /Prachi
drwxr-xr-x - acadgild supergroup          0 2018-03-24 13:09 /assignment4_input
drwxr-xr-x - acadgild supergroup          0 2018-03-24 14:34 /assignment4_output1
drwxr-xr-x - acadgild supergroup          0 2018-03-24 14:29 /assignment4_output1IA
drwxr-xr-x - acadgild supergroup          0 2018-03-30 12:50 /assignment4_output2
drwxr-xr-x - acadgild supergroup          0 2018-03-30 13:12 /assignment4_output3
drwxr-xr-x - acadgild supergroup          0 2018-03-24 09:38 /hadoopdata
drwxr-xr-x - acadgild supergroup          0 2018-03-30 15:19 /hbase

```

Output

```

[acadgild@localhost ~]$ hadoop fs -cat /Assignment5_Output4/part-r-00000
18/03/31 04:16:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Song shared for 2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

```