

Assignment 5:

Submitted by Jyoti Jha

DOS: 31st July'18

TASK-1:

Find the number of unique listeners in the data set.

Solution1:

Mapper code for Task1:-

```
import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.*;
public class UniqListenerMapper extends
Mapper<LongWritable,Text,Text,IntWritable>{
public void map(LongWritable key, Text values, Context context)
throws IOException,InterruptedException{
String[] line = values.toString().split("\\|");
Text uid = new Text(line[0]);
IntWritable value = new
IntWritable(1);
context.write(uid,value);
}
}
```

Reducer code for Task1:-

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
public class UniqueListenerReducer extends
Reducer<Text,IntWritable,Text,IntWritable> {
public void reduce(Text uid, Iterable<IntWritable>
count, Context context)
throws IOException, InterruptedException{
int sum = 0;
for(IntWritable values: count){
sum+= values.get();
}
context.write(uid,new IntWritable(sum));
}
}
```

Driver code for Task1:-

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
public class UniqueListener {
public static void main(String[] args) throws Exception {
// TODO Auto-generated method stub
Configuration conf = new Configuration();
Job job = new Job(conf, "Assignment_5_1");
job.setJarByClass(UniqueListener.class);
job.setMapOutputKeyClass(Text.class);
job.setMapOutputValueClass(IntWritable.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
job.setMapperClass(UniqListenerMapper.class);
job.setReducerClass(UniqueListenerReducer.class);
job.setNumReduceTasks(2);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
job.waitForCompletion(true);
}
}

```

Execution of Jar File for Task1: -

```

[acadgild@localhost ~]$ hadoop jar Desktop/Jyoti/Assignment5_Task1.jar UniqueListener /musicdata.txt /Output5_1
18/07/31 20:08:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
pplicable
18/07/31 20:08:31 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/31 20:08:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interf
xecute your application with ToolRunner to remedy this.
18/07/31 20:08:32 INFO input.FileInputFormat: Total input paths to process : 1
18/07/31 20:08:33 INFO mapreduce.JobSubmitter: number of splits:1
18/07/31 20:08:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533047119572_0001
18/07/31 20:08:34 INFO impl.YarnClientImpl: Submitted application application_1533047119572_0001
18/07/31 20:08:34 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533047119572_0001/
18/07/31 20:08:34 INFO mapreduce.Job: Running job: job_1533047119572_0001
18/07/31 20:08:47 INFO mapreduce.Job: Job job_1533047119572_0001 running in uber mode : false
18/07/31 20:08:47 INFO mapreduce.Job: map 0% reduce 0%
18/07/31 20:08:53 INFO mapreduce.Job: map 100% reduce 0%
18/07/31 20:09:05 INFO mapreduce.Job: map 100% reduce 50%
18/07/31 20:09:06 INFO mapreduce.Job: map 100% reduce 100%
18/07/31 20:09:07 INFO mapreduce.Job: Job job_1533047119572_0001 completed successfully
18/07/31 20:09:07 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=64
  FILE: Number of bytes written=324057
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=173
  HDFS: Number of bytes written=27
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Job Counters
  Launched map tasks=1
  Launched reduce tasks=2
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4457
  Total time spent by all reduces in occupied slots (ms)=17761
  Total time spent by all map tasks (ms)=4457
  Total time spent by all reduce tasks (ms)=17761
  Total vcore-milliseconds taken by all map tasks=4457
  Total vcore-milliseconds taken by all reduce tasks=17761

```

```

  Total vcore-milliseconds taken by all map tasks=4457
  Total vcore-milliseconds taken by all reduce tasks=17761
  Total megabyte-milliseconds taken by all map tasks=4563968
  Total megabyte-milliseconds taken by all reduce tasks=18187264
Map Reduce Framework

```

```

[acadgild@localhost ~]$ hadoop fs -ls /Output5_1
18/07/31 20:10:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
pplicable
Found 3 items
-rw-r--r-- 1 acadgild supergroup          0 2018-07-31 20:09 /Output5_1/ SUCCESS
-rw-r--r-- 1 acadgild supergroup          0 2018-07-31 20:09 /Output5_1/part-r-00000
-rw-r--r-- 1 acadgild supergroup        27 2018-07-31 20:09 /Output5_1/part-r-00001
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /Output5_1/part-r-00000
18/07/31 20:10:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
pplicable
[acadgild@localhost ~]$ hadoop fs -cat /Output5_1/part-r-00001
18/07/31 20:11:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
pplicable
111113 1
111115 2
111117 1

```

TASK-2:

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

Solution2:

Mapper code for Task2 :-

```

package task2;
import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.*;
public class FullSongMapper extends Mapper<LongWritable,Text,Text,IntWritable> {
public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException{
String[] line = value.toString().split("\\|");
Text uid = new Text(line[0]);
if(line[4].equals("1")){
context.write(uid,new IntWritable(1));
}
}
}

```

Reducer code for Task2:-

```

package task2;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
public class FullSongReducer extends Reducer<Text,IntWritable,Text,IntWritable>{

```

```
public void reduce(Text uid, Iterable<IntWritable> count, Context context) throws
IOException, InterruptedException{
int sum = 0;
for(IntWritable
value: count){
sum+= value.get();
}
context.write(uid,new
IntWritable(sum));
}
}
```

Driver code for Task2:-

```
package task2;
```

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
public class FullSong {
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = new Job(conf, "Assignment_5_2");
        job.setJarByClass(FullSong.class);
        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(IntWritable.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        job.setMapperClass(FullSongMapper.class);
        job.setReducerClass(FullSongReducer.class);
        job.setNumReduceTasks(2);
        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.waitForCompletion(true);
    }
}

```

Execution of Jar File for Task2:

```

[acagdild@localhost ~]$ hadoop jar Desktop/Jyoti/Assignment5_Task2.jar FullSong /musicdata.txt /Output5_2
18/07/31 20:12:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
pplicable
18/07/31 20:12:13 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/31 20:12:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
xecute your application with ToolRunner to remedy this.
18/07/31 20:12:15 INFO input.FileInputFormat: Total input paths to process : 1
18/07/31 20:12:15 INFO mapreduce.JobSubmitter: number of splits:1
18/07/31 20:12:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533047119572_0002
18/07/31 20:12:16 INFO impl.YarnClientImpl: Submitted application application_1533047119572_0002
18/07/31 20:12:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533047119572_0002/
18/07/31 20:12:16 INFO mapreduce.Job: Running job: job_1533047119572_0002
18/07/31 20:12:32 INFO mapreduce.Job: Job job_1533047119572_0002 running in uber mode : false
18/07/31 20:12:32 INFO mapreduce.Job:  map 0% reduce 0%
18/07/31 20:12:46 INFO mapreduce.Job:  map 100% reduce 0%
18/07/31 20:13:05 INFO mapreduce.Job:  map 100% reduce 50%
18/07/31 20:13:10 INFO mapreduce.Job:  map 100% reduce 100%
18/07/31 20:13:10 INFO mapreduce.Job: Job job_1533047119572_0002 completed successfully
18/07/31 20:13:11 INFO mapreduce.Job: Counters: 50
    File System Counters
        FILE: Number of bytes read=25
        FILE: Number of bytes written=323970
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=173
        HDFS: Number of bytes written=9
        HDFS: Number of read operations=9
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
    Job Counters
        Killed reduce tasks=1
        Launched map tasks=1
        Launched reduce tasks=2
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=11392
        Total time spent by all reduces in occupied slots (ms)=37501
        Total time spent by all map tasks (ms)=11392

```

```

        Total time spent by all reduces in occupied slots (ms)=37501
        Total time spent by all map tasks (ms)=11392
        Total time spent by all reduce tasks (ms)=37501
        Total vcore-milliseconds taken by all map tasks=11392

```

```
[acadgild@localhost ~]$ hadoop fs -ls /Output5_2
18/07/31 20:13:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
pplicable
Found 3 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-31 20:13 /Output5_2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 0 2018-07-31 20:13 /Output5_2/part-r-00000
-rw-r--r-- 1 acadgild supergroup 9 2018-07-31 20:13 /Output5_2/part-r-00001
-rw-r--r-- 1 acadgild supergroup 10 2018-07-31 20:13 /Output5_2/part-r-00002

[acadgild@localhost ~]$ hadoop fs -cat /Output5_2/part-r-00001
18/07/31 20:13:53 WARN util.NativeCodeLoader: Unable to load native-ha
pplicable
111117 1
```

TASK-3:

What are the number of times a song was shared?

Solution3:

Mapper code for task3:-

```
import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.*;
public class SongShareMapper extends Mapper<LongWritable,Text,Text,IntWritable> {
public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException{
String[] line = value.toString().split("\\|");
if(line[2].equals("1")){
context.write(new Text(line[0]), new IntWritable(1));
}
}
}
```

Reducer code for task3:-

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
public class SongShareReducer extends Reducer<Text,IntWritable,Text, IntWritable> {
public void reducer(Text uid, Iterable<IntWritable> count, Context context) throws
IOException, InterruptedException{
int sum = 0;
for(IntWritable
value:count){
sum+= value.get();
}
context.write(uid, new
IntWritable(sum));
}
}
```

Driver code for task3:-

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import task2.FullSong;
import task2.FullSongMapper;
import task2.FullSongReducer;
public class SongShare {
    public static void main(String[] args) throws Exception {
        // TODO Auto-generated method stub
        Configuration conf = new Configuration();
        Job job = new Job(conf, "Assignment_5_3");
        job.setJarByClass(SongShare.class);
        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(IntWritable.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        job.setMapperClass(SongShareMapper.class);
        job.setReducerClass(SongShareReducer.class);
        job.setNumReduceTasks(2);
        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.waitForCompletion(true);
    }
}

```

Execution of Jar File for Task3:-

```
[acadgild@localhost ~]$ hadoop jar Desktop/Jyoti/Assignment5_Task3.jar SongShare /musicdata.txt /Output5_3
18/07/31 22:17:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
pplicable
18/07/31 22:17:39 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/31 22:17:41 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the To
xecute your application with ToolRunner to remedy this.
18/07/31 22:17:41 INFO input.FileInputFormat: Total input paths to process : 1
18/07/31 22:17:42 INFO mapreduce.JobSubmitter: number of splits:1
18/07/31 22:17:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533047119572_0003
18/07/31 22:17:43 INFO impl.YarnClientImpl: Submitted application application_1533047119572_0003
18/07/31 22:17:44 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533047119572_00
18/07/31 22:17:44 INFO mapreduce.Job: Running job: job_1533047119572_0003
18/07/31 22:18:01 INFO mapreduce.Job: Job job_1533047119572_0003 running in uber mode : false
18/07/31 22:18:01 INFO mapreduce.Job: map 0% reduce 0%
18/07/31 22:18:19 INFO mapreduce.Job: map 100% reduce 0%
18/07/31 22:18:41 INFO mapreduce.Job: map 100% reduce 50%
18/07/31 22:18:42 INFO mapreduce.Job: map 100% reduce 100%
18/07/31 22:18:43 INFO mapreduce.Job: Job job_1533047119572_0003 completed successfully
18/07/31 22:18:43 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=38
  FILE: Number of bytes written=324002
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=173
  HDFS: Number of bytes written=18
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Job Counters
  Killed reduce tasks=1
  Launched map tasks=1
  Launched reduce tasks=2
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=12596
  Total time spent by all reduces in occupied slots (ms)=39491
  Total time spent by all map tasks (ms)=12596
  Total time spent by all reduce tasks (ms)=39491
  Total vcore-milliseconds taken by all map tasks=12596
```

```

  Total time spent by all reduces in occupied slots (ms)=39491
  Total time spent by all map tasks (ms)=12596
  Total time spent by all reduce tasks (ms)=39491
  Total vcore-milliseconds taken by all map tasks=12596
  Total vcore-milliseconds taken by all reduce tasks=39491
  Total megabyte-milliseconds taken by all map tasks=12898304
  Total megabyte-milliseconds taken by all reduce tasks=40438784
Map-Reduce Framework
  Map input records=4
  Map output records=2
  Map output bytes=22
  Map output materialized bytes=38
  Input split bytes=100
  Combine input records=0
  Combine output records=0
  Reduce input groups=2
  Reduce shuffle bytes=38
  Reduce input records=2
  Reduce output records=2
  Spilled Records=4
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=623
  CPU time spent (ms)=5100
  Physical memory (bytes) snapshot=394375168
  Virtual memory (bytes) snapshot=6179631104
  Total committed heap usage (bytes)=202575872
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=73
File Output Format Counters
  Bytes Written=18
You have new mail in /var/spool/mail/acadgild
```


Output of Task3:-

```
[acadgild@localhost ~]$ hadoop fs -ls /Output5_3
18/07/31 22:49:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
Found 3 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-31 22:18 /Output5_3/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 0 2018-07-31 22:18 /Output5_3/part-r-00000
-rw-r--r-- 1 acadgild supergroup 18 2018-07-31 22:18 /Output5_3/part-r-00001

[acadgild@localhost ~]$ hadoop fs -cat /Output5_3/part-r-00001
18/07/31 22:19:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
lllll13 1
lllll15 1
```