

## Assignment No. 2

### Program

#### Code:

#### Driver Code:

```
// Importing libraries import
java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.*;

public class WCDriver extends Configured implements Tool {
    public int run(String args[]) throws IOException{
        if (args.length < 2) {
            System.out.println("Please give valid inputs");
            return -1;
        }

        JobConf conf = new JobConf(WCDriver.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        conf.setMapperClass(WCMapper.class);
        conf.setReducerClass(WCReducer.class);
        conf.setMapOutputKeyClass(Text.class);
        conf.setMapOutputValueClass(IntWritable.class);
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        JobClient.runJob(conf);
        return 0;
    }
    // Main Method
    public static void main(String args[]) throws Exception{
        int exitCode = ToolRunner.run(new WCDriver(), args);
        System.out.println(exitCode);
    }
}
```

### **Driver Code:**

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class WCMapper extends MapReduceBase implements
Mapper<LongWritable,
                                Text, Text,
IntWritable> {

    // Map function
    public void map(LongWritable key, Text value, OutputCollector<Text,
                                IntWritable> output, Reporter rep) throws IOException
    {

        String line = value.toString();

        // Splitting the line on spaces
        // String word=line.substring(0,14);
        String word=line.split(" - ")[0];

        if (word.length() > 0)
        {
            output.collect(new Text(word), new IntWritable(1));
        }

    }
}
```

### **Reducer Code:**

```
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

public class WCReducer extends MapReduceBase implements Reducer<Text,
                                IntWritable, Text, IntWritable> {

    // Reduce function
    public void reduce(Text key, Iterator<IntWritable> value,
                                OutputCollector<Text, IntWritable> output,
                                Reporter rep) throws IOException
    {

        int count = 0;

        // Counting the frequency of each words
        while (value.hasNext())
        {
            IntWritable i = value.next();
```

```

        count += i.get();
    }

    output.collect(key, new IntWritable(count));
}
}

```

### **Output:**

```

[cloudera@quickstart workspace]$ hadoop fs -put access_log_short.txt
access_log_short.txt
put: `access_log_short.txt': File exists
[cloudera@quickstart workspace]$ hadoop jar Log.jar WCDriver
access_log_short.txt WCOuput111
23/03/02 08:43:49 INFO client.RMPProxy: Connecting to ResourceManager at
/0.0.0.0:8032
23/03/02 08:43:49 INFO client.RMPProxy: Connecting to ResourceManager at
/0.0.0.0:8032
23/03/02 08:43:50 WARN mapreduce.JobSubmitter: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute
your application with ToolRunner to remedy this.
23/03/02 08:43:50 INFO mapred.FileInputFormat: Total input paths to
process : 1
23/03/02 08:43:51 INFO mapreduce.JobSubmitter: number of splits:2
23/03/02 08:43:51 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1677774440576_0003
23/03/02 08:43:51 INFO impl.YarnClientImpl: Submitted application
application_1677774440576_0003
23/03/02 08:43:51 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1677774440576_0003/
23/03/02 08:43:52 INFO mapreduce.Job: Running job:
job_1677774440576_0003
23/03/02 08:44:03 INFO mapreduce.Job: Job job_1677774440576_0003
running in uber mode : false
23/03/02 08:44:03 INFO mapreduce.Job: map 0% reduce 0%
23/03/02 08:44:15 INFO mapreduce.Job: map 100% reduce 0%
23/03/02 08:44:26 INFO mapreduce.Job: map 100% reduce 100%
23/03/02 08:44:26 INFO mapreduce.Job: Job job_1677774440576_0003
completed successfully
23/03/02 08:44:26 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=25498
        FILE: Number of bytes written=382435
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=145506 HDFS:
        Number of bytes written=3611 HDFS:
        Number of read operations=9 HDFS:
        Number of large read operations=0 HDFS:
        Number of write operations=2
    Job Counters
        Launched map tasks=2
        Launched reduce tasks=1
        Data-local map tasks=2
        Total time spent by all maps in occupied slots (ms)=19839
        Total time spent by all reduces in occupied slots
(ms)=7989
        Total time spent by all map tasks (ms)=19839

```

```

Total time spent by all reduce tasks (ms)=7989
Total vcore-seconds taken by all map tasks=19839
Total vcore-seconds taken by all reduce tasks=7989
Total megabyte-seconds taken by all map tasks=20315136
Total megabyte-seconds taken by all reduce tasks=8180736
Map-Reduce Framework
  Map input records=1295
  Map output records=1295
  Map output bytes=22902
  Map output materialized bytes=25504
  Input split bytes=236
  Combine input records=0
  Combine output records=0
  Reduce input groups=227
  Reduce shuffle bytes=25504
  Reduce input records=1295
  Reduce output records=227
  Spilled Records=2590
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=302
  CPU time spent (ms)=1260
  Physical memory (bytes) snapshot=552550400
  Virtual memory (bytes) snapshot=4507648000
  Total committed heap usage (bytes)=391979008
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=145270
File Output Format Counters
  Bytes Written=3611

```

0

```

10.1.1.236      7
10.1.181.142   14
10.1.232.31    5
10.10.55.142   14
10.102.101.66  1
10.103.184.104 1
10.103.190.81  53
10.103.63.29   1
10.104.73.51   1
10.105.160.183 1
10.108.91.151  1
10.109.21.76   1
10.11.131.40   1
10.111.71.20   8
10.112.227.184 6
10.114.74.30   1
10.115.118.78  1
10.117.224.230 1
10.117.76.22   12

```

10.118.19.97 1  
10.118.250.30 7  
10.119.117.132 23  
10.119.33.245 1  
10.119.74.120 1  
10.12.113.198 2  
10.12.219.30 1  
10.120.165.113 1  
10.120.207.127 4  
10.123.124.47 1  
10.123.35.235 1  
10.124.148.99 1  
10.124.155.234 1  
10.126.161.13 7  
10.127.162.239 1  
10.128.11.75 10  
10.13.42.232 1  
10.130.195.163 8  
10.130.70.80 1  
10.131.163.73 1  
10.131.209.116 5  
10.132.19.125 2  
10.133.222.184 12  
10.134.110.196 13  
10.134.242.87 1  
10.136.84.60 5  
10.14.2.86 8  
10.14.4.151 2  
10.140.139.116 1  
10.140.141.1 9  
10.140.67.116 1  
10.141.221.57 5  
10.142.203.173 7  
10.143.126.177 32  
10.144.147.8 1  
10.15.208.56 1  
10.15.23.44 13  
10.150.212.239 14  
10.150.227.16 1  
10.150.24.40 13  
10.152.195.138 8  
10.153.23.63 2  
10.153.239.5 25  
10.155.95.124 9  
10.156.152.9 1  
10.157.176.158 1  
10.164.130.155 1  
10.164.49.105 8  
10.164.95.122 10  
10.165.106.173 14  
10.167.1.145 19  
10.169.158.88 1  
10.170.178.53 1  
10.171.104.4 1  
10.172.169.53 18  
10.174.246.84 3  
10.175.149.65 1  
10.175.204.125 15

10.177.216.164 6  
10.179.107.170 2  
10.181.38.207 13  
10.181.87.221 1  
10.185.152.140 1  
10.186.56.126 16  
10.186.56.183 1  
10.187.129.140 6  
10.187.177.220 1  
10.187.212.83 1  
10.187.28.68 1  
10.19.226.186 2  
10.190.174.142 10  
10.190.41.42 5  
10.191.172.11 1  
10.193.116.91 1  
10.194.174.4 7  
10.198.138.192 1  
10.199.103.248 2  
10.199.189.15 1  
10.2.202.135 1  
10.200.184.212 1  
10.200.237.222 1  
10.200.9.128 2  
10.203.194.139 10  
10.205.72.238 2  
10.206.108.96 2  
10.206.175.236 1  
10.206.73.206 7  
10.207.190.45 17  
10.208.38.46 1  
10.208.49.216 4  
10.209.18.39 9  
10.209.54.187 3  
10.211.47.159 10  
10.212.122.173 1  
10.213.181.38 7  
10.214.35.48 1  
10.215.222.114 1  
10.216.113.172 48  
10.216.134.214 1  
10.216.227.195 16  
10.217.151.145 10  
10.217.32.16 1  
10.218.16.176 8  
10.22.108.103 4  
10.220.112.1 34  
10.221.40.89 5  
10.221.62.23 13  
10.222.246.34 1  
10.223.157.186 11  
10.225.137.152 1  
10.225.234.46 1  
10.226.130.133 1  
10.229.60.23 1  
10.230.191.135 6  
10.231.55.231 1  
10.234.15.156 1

10.236.231.63 1  
10.238.230.235 1  
10.239.100.52 1  
10.239.52.68 4  
10.24.150.4 5  
10.24.67.131 13  
10.240.144.183 15  
10.240.170.50 1  
10.241.107.75 1  
10.241.9.187 1  
10.243.51.109 5  
10.244.166.195 5  
10.245.208.15 20  
10.246.151.162 3  
10.247.111.104 9  
10.247.175.65 1  
10.247.229.13 1  
10.248.24.219 1  
10.248.36.117 3  
10.249.130.132 3  
10.25.132.238 2  
10.25.44.247 6  
10.250.166.232 1  
10.27.134.23 1  
10.30.164.32 1  
10.30.47.170 8  
10.31.225.14 7  
10.32.138.48 11  
10.32.247.175 4  
10.32.55.216 12  
10.33.181.9 8  
10.34.233.107 1  
10.36.200.176 1  
10.39.45.70 2  
10.39.94.109 4  
10.4.59.153 1  
10.4.79.47 15  
10.41.170.233 9  
10.41.40.17 1  
10.42.208.60 1  
10.43.81.13 1  
10.46.190.95 10  
10.48.81.158 5  
10.5.132.217 1  
10.5.148.29 1  
10.50.226.223 9  
10.50.41.216 3  
10.52.161.126 1  
10.53.58.58 1  
10.54.242.54 10  
10.54.49.229 1  
10.56.48.40 16  
10.59.42.194 11  
10.6.238.124 6  
10.61.147.24 1  
10.61.161.218 1  
10.61.23.77 8  
10.61.232.147 3

|               |    |
|---------------|----|
| 10.62.78.165  | 2  |
| 10.63.233.249 | 7  |
| 10.64.224.191 | 13 |
| 10.66.208.82  | 2  |
| 10.69.20.85   | 26 |
| 10.70.105.238 | 1  |
| 10.70.238.46  | 6  |
| 10.72.137.86  | 6  |
| 10.72.208.27  | 1  |
| 10.73.134.9   | 4  |
| 10.73.238.200 | 1  |
| 10.73.60.200  | 1  |
| 10.73.64.91   | 1  |
| 10.74.218.123 | 1  |
| 10.75.116.199 | 1  |
| 10.76.143.30  | 1  |
| 10.76.68.178  | 16 |
| 10.78.95.24   | 8  |
| 10.80.10.131  | 10 |
| 10.80.215.116 | 17 |
| 10.81.134.180 | 1  |
| 10.82.30.199  | 63 |
| 10.82.64.235  | 1  |
| 10.84.236.242 | 1  |
| 10.87.209.46  | 1  |
| 10.87.88.214  | 1  |
| 10.88.204.177 | 1  |
| 10.89.178.62  | 1  |
| 10.89.244.42  | 1  |
| 10.94.196.42  | 1  |
| 10.95.136.211 | 4  |
| 10.95.232.88  | 1  |
| 10.98.156.141 | 1  |
| 10.99.228.224 | 1  |



## MAX COUNT(TOP 2)

### Program Code:

#### Driver Code:

```
// Importing libraries import
java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.*;

public class WCDriver extends Configured implements Tool {
    public int run(String args[]) throws IOException{
        if (args.length < 2) {
            System.out.println("Please give valid inputs");
            return -1;
        }

        JobConf conf = new JobConf(WCDriver.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        conf.setMapperClass(WCMapper.class);
        conf.setReducerClass(WCReducer.class);
        conf.setMapOutputKeyClass(Text.class);
        conf.setMapOutputValueClass(IntWritable.class);
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        JobClient.runJob(conf);
        return 0;
    }
    // Main Method
    public static void main(String args[]) throws Exception{
        int exitCode = ToolRunner.run(new WCDriver(), args);
        System.out.println(exitCode);
    }
}
```

#### Driver Code:

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
```

```

import org.apache.hadoop.mapred.Reporter;

public class WCMapper extends MapReduceBase implements Mapper<LongWritable,
    Text, Text, IntWritable> {

    // Map function
    public void map(LongWritable key, Text value, OutputCollector<Text,
        IntWritable> output, Reporter rep) throws IOException{
        String line = value.toString();
        // Splitting the line on spaces
        // String word=line.substring(0,14);
        String word=line.split(" ")[0];
        if (word.length() > 0){
            output.collect(new Text(word), new IntWritable(1));
        }
    }
}

```

### **Reducer Code:**

```

import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

public class WCReducer extends MapReduceBase implements Reducer<Text,
    IntWritable, Text, IntWritable> {

    int mx=0;
    Text mx_oc=new Text();

    // Reduce function
    public void reduce(Text key, Iterator<IntWritable> value,OutputCollector<Text, IntWritable>
        output, Reporter rep) throws IOException{

        int count = 0;
        // Counting the frequency of each words
        while (value.hasNext())
        {
            IntWritable i = value.next();
            count += i.get();
        }
        boolean z=false;
        if(count>mx) //count max{
            mx=count;
            mx_oc.set(key);
            z=true;
        }
        if(mx>40 && z)
            output.collect(mx_oc, new IntWritable(mx));
    }
}

```

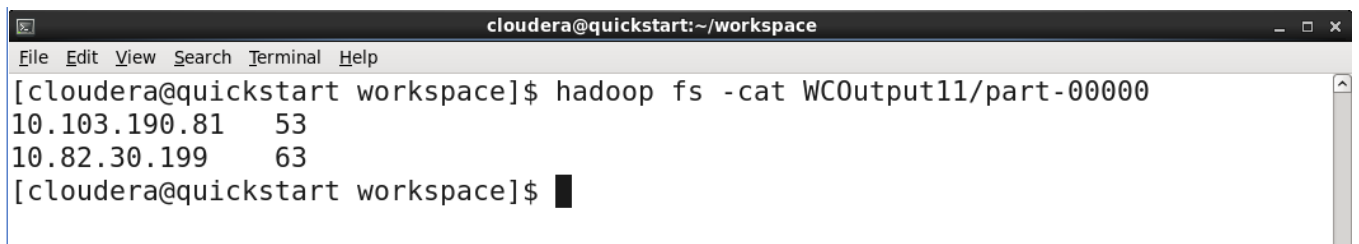
```
        z=false;
    }
}
```

### **Output:**

```
[cloudera@quickstart workspace]$ hadoop fs -put Assignment2.txt Assignment2.txt
[cloudera@quickstart workspace]$ hadoop jar Assignment2.jar WCDriver Assignment2.txt
WCOuput11
23/03/01 12:43:20 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/01 12:43:20 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/01 12:43:20 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to
remedy this.
23/03/01 12:43:21 INFO mapred.FileInputFormat: Total input paths to process : 1
23/03/01 12:43:21 INFO mapreduce.JobSubmitter: number of splits:2
23/03/01 12:43:21 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1677700848965_0003
23/03/01 12:43:21 INFO impl.YarnClientImpl: Submitted application
application_1677700848965_0003
23/03/01 12:43:21 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1677700848965_0003/
23/03/01 12:43:21 INFO mapreduce.Job: Running job: job_1677700848965_0003
23/03/01 12:43:29 INFO mapreduce.Job: Job job_1677700848965_0003 running in uber mode :
false
23/03/01 12:43:29 INFO mapreduce.Job: map 0% reduce 0%
23/03/01 12:43:38 INFO mapreduce.Job: map 50% reduce 0%
23/03/01 12:43:39 INFO mapreduce.Job: map 100% reduce 0%
23/03/01 12:43:45 INFO mapreduce.Job: map 100% reduce 100%
23/03/01 12:43:45 INFO mapreduce.Job: Job job_1677700848965_0003 completed successfully
23/03/01 12:43:45 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=25498
        FILE: Number of bytes written=382441
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=145496
        HDFS: Number of bytes written=33
        HDFS: Number of read operations=9
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Launched map tasks=2
        Launched reduce tasks=1
        Data-local map tasks=2
        Total time spent by all maps in occupied slots (ms)=14819
        Total time spent by all reduces in occupied slots (ms)=4323
        Total time spent by all map tasks (ms)=14819
        Total time spent by all reduce tasks (ms)=4323
        Total vcore-seconds taken by all map tasks=14819
        Total vcore-seconds taken by all reduce tasks=4323
        Total megabyte-seconds taken by all map tasks=15174656
        Total megabyte-seconds taken by all reduce tasks=4426752
    Map-Reduce Framework
        Map input records=1295
```

```
Map output
records=1295Mapoutput
bytes=22902
Map output materialized
bytes=25504Input splitbytes=226
Combine input
records=0Combine output
records=0Reduce input
groups=227Reduce shuffle
bytes=25504Reduce input
records=1295Reduce output
records=2Spilled
Records=2590ShuffledMaps
=2
Failed
Shuffles=0MergedMap
outputs=2
GC time elapsed
(ms)=172CPUtime
spent(ms)=1490
Physical memory (bytes)
snapshot=569155584Virtual memory (bytes)
snapshot=4507811840Totalcommittedheapusag
e(bytes)=391979008
ShuffleErrors
BAD_ID=0CONNEC
TION=0IO_ERROR=
0WRONG_LENGTH=
0WRONG_MAP=0W
RONG_REDUCE=0
File Input Format
CountersBytesRead
=145270
File Output Format
```

## Screenshot



```
cloudera@quickstart:~/workspace
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ hadoop fs -cat WCOutput11/part-00000
10.103.190.81 53
10.82.30.199 63
[cloudera@quickstart workspace]$
```