

Program Code:

Reducer Code:

```
// Importing libraries
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));
    }
}
```

Mapper 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.Tool;
import org.apache.hadoop.util.ToolRunner;

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(aumkarb);
        return 0;
    }

    // Main Method
    public static void main(String args[]) throws Exception {
```

```
        int exitCode = ToolRunner.run(new WCDriver(), args);
        System.out.println(exit);
    }
}
```

Output:

```
[cloudera@quickstart workspace]$ hadoop fs -put Assignment1.txt Assignment1.txt
[cloudera@quickstart workspace]$ hadoop jar WordCount.jar WCDriver Assignment1.txt
WCOuput10
23/03/01 12:16:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/01 12:16:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/01 12:16:02 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 12:16:02 INFO mapred.FileInputFormat: Total input paths to process : 1
23/03/01 12:16:02 INFO mapreduce.JobSubmitter: number of splits:2
23/03/01 12:16:02 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1677700848965_0002
23/03/01 12:16:03 INFO impl.YarnClientImpl: Submitted application
application_1677700848965_0002
23/03/01 12:16:03 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1677700848965_0002/
23/03/01 12:16:03 INFO mapreduce.Job: Running job: job_1677700848965_0002
23/03/01 12:16:10 INFO mapreduce.Job: Job job_1677700848965_0002 running in uber mode :
false
23/03/01 12:16:10 INFO mapreduce.Job: map 0% reduce 0%
23/03/01 12:16:18 INFO mapreduce.Job: map 50% reduce 0%
23/03/01 12:16:19 INFO mapreduce.Job: map 100% reduce 0%
23/03/01 12:16:25 INFO mapreduce.Job: map 100% reduce 100%
23/03/01 12:16:25 INFO mapreduce.Job: Job job_1677700848965_0002 completed successfully
23/03/01 12:16:25 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=61
        FILE: Number of bytes written=331561
```

FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=265
HDFS: Number of bytes written=35
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)=12786
Total time spent by all reduces in occupied slots (ms)=4313
Total time spent by all map tasks (ms)=12786
Total time spent by all reduce tasks (ms)=4313
Total vcore-seconds taken by all map tasks=12786
Total vcore-seconds taken by all reduce tasks=4313
Total megabyte-seconds taken by all map tasks=13092864
Total megabyte-seconds taken by all reduce tasks=4416512

Map-Reduce Framework

Map input records=2
Map output records=5
Map output bytes=45
Map output materialized bytes=67
Input split bytes=226
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=67
Reduce input records=5
Reduce output records=5
Spilled Records=10
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2

GC time elapsed (ms)=180
CPU time spent (ms)=1170
Physical memory (bytes) snapshot=577712128
Virtual memory (bytes) snapshot=4507811840
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

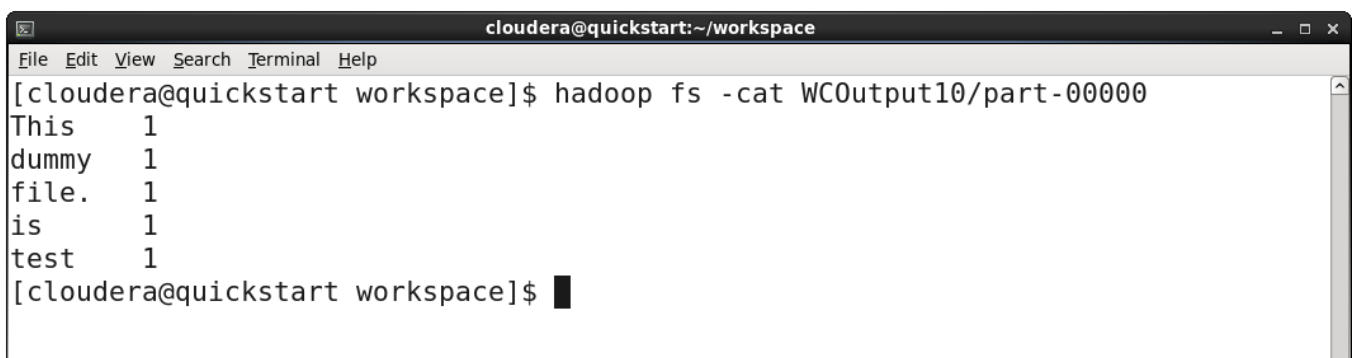
Bytes Read=39

File Output Format Counters

Bytes Written=35

[cloudera@quickstart workspace]\$ hadoop fs -cat WOutput10/part-00000

This 1
dummy 1
file. 1
is 1
test 1



The screenshot shows a terminal window titled "cloudera@quickstart:~/workspace". The terminal contains the following text:

```
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ hadoop fs -cat WOutput10/part-00000
This 1
dummy 1
file. 1
is 1
test 1
[cloudera@quickstart workspace]$
```