

```

//WordCount.java
import java.io.IOException;
import java.util.StringTokenizer;
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.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable>{

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context
            ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }

    public static class IntSumReducer
        extends Reducer<Text,IntWritable,Text,IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable> values,
            Context context
            ) throws IOException, InterruptedException {

            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }

    public static void main(String[] args) throws Exception {

```

```

Configuration conf = new Configuration();
Job job = Job.getInstance(conf, "word count");
job.setJarByClass(WordCount.class);
job.setMapperClass(TokenizerMapper.class);
job.setCombinerClass(IntSumReducer.class);
job.setReducerClass(IntSumReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

//input.txt

Sumit
Kshitij
Samarth
Sham
Samarth
Kshitij
Sumit
Kshitij
Anurag

//Output

```

student@student:~/Desktop/WordCountTutorial$ hadoop fs -ls /out321
Found 2 items
-rw-r--r--  1 student supergroup      0 2023-04-26 13:05 /out321/_SUCCESS
-rw-r--r--  1 student supergroup    65 2023-04-26 13:05 /out321/part-r-
00000
student@student:~/Desktop/WordCountTutorial$ hadoop fs -cat /out321/part-r-
00000
Anurag      1
Kshitij     3
Samarth     2
Sham        1
Shubham     1
Sumit       2
student@student:~/Desktop/WordCountTutorial$

```

```
student@student: ~/Desktop/WordCountTutorial
Map input records=12
Map output records=11
Map output bytes=125
Map output materialized bytes=99
Input split bytes=102
Combine input records=11
Combine output records=7
Reduce input groups=7
Reduce shuffle bytes=99
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=74
CPU time spent (ms)=920
Physical memory (bytes) snapshot=398172160
Virtual memory (bytes) snapshot=3837599744
Total committed heap usage (bytes)=268959744

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=82
File Output Format Counters
  Bytes Written=65
student@student:~/Desktop/WordCountTutorial$ hadoop fs -ls /out321
Found 2 items
-rw-r--r-- 1 student supergroup 0 2023-04-26 13:05 /out321/_SUCCESS
-rw-r--r-- 1 student supergroup 65 2023-04-26 13:05 /out321/part-r-00000
student@student:~/Desktop/WordCountTutorial$ hadoop fs -cat /out321/part-r-00000Anurag 1
Kshitij 3
Sanarth 2
Sham 1
Sumit 2
1
student@student:~/Desktop/WordCountTutorial$
```

Browsing HDFS — Mozilla Firefox

How to run Word Count | WordCount.java | MapReduce Word Count | Browsing HDFS | Central Repository: org | +

localhost:50070/explorer.html#/? 110% Search

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/ Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	student	supergroup	0 B	26/4/2023, 12:51:02 pm	0	0 B	WordCountTutorial
drwxr-xr-x	student	supergroup	0 B	26/4/2023, 1:02:08 pm	0	0 B	input
drwxr-xr-x	student	supergroup	0 B	26/4/2023, 1:05:16 pm	0	0 B	out321
drwxrwx---	student	supergroup	0 B	26/4/2023, 12:02:01 pm	0	0 B	tmp

Hadoop, 2016.