

# Character Count with MapReduce\*

Montasser AKERMI

Last Update: February, 2025

---

\*The course material is hosted at <https://akermi.org/c2/>.

## Learning objectives

Write a MapReduce application that counts the number of characters in a corpus of text of one or multiple files.

## Input data

my dreams my dreams  
what has become of their sweetness  
what indeed has become of my youth

## Output data

Characters 73

## Solution

Listing 1: CharCount Class

```
1 import java.io.IOException;
2
3 import org.apache.hadoop.conf.Configuration;
4 import org.apache.hadoop.fs.Path;
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Job;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Reducer;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
12
13 public class CharCount {
14
15     public static class CharMapper
16         extends Mapper<Object, Text, Text, IntWritable> {
17
18         private final static Text word = new Text("Characters");
19         private IntWritable length = new IntWritable();
20
21         public void map(Object key, Text value, Context context
22             ) throws IOException, InterruptedException {
23             length.set(value.getLength());
24             context.write(word, length);
25         }
26     }
27
28     public static class IntSumReducer
29         extends Reducer<Text, IntWritable, Text, IntWritable> {
30         private IntWritable result = new IntWritable();
31
32         public void reduce(Text key, Iterable<IntWritable> values,
33             Context context
34             ) throws IOException, InterruptedException {
35             int sum = 0;
```

```

36     for (IntWritable val : values) {
37         sum += val.get();
38     }
39     result.set(sum);
40     context.write(key, result);
41 }
42 }
43
44 public static void main(String[] args) throws Exception {
45     Configuration conf = new Configuration();
46     Job job = Job.getInstance(conf, "Char count");
47     job.setJarByClass(CharCount.class);
48     job.setMapperClass(CharMapper.class);
49     job.setCombinerClass(IntSumReducer.class);
50     job.setReducerClass(IntSumReducer.class);
51     job.setOutputKeyClass(Text.class);
52     job.setOutputValueClass(IntWritable.class);
53     FileInputFormat.addInputPath(job, new Path(args[0]));
54     FileOutputFormat.setOutputPath(job, new Path(args[1]));
55     System.exit(job.waitForCompletion(true) ? 0 : 1);
56 }
57 }

```

*Note: It is highly recommended to manually write the code instead of copy/pasting.*