

Line 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 lines of a text corpus in 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

Lines 3

Solution

Listing 1: LineCount 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 LineCount {
14
15     public static class LineMapper
16         extends Mapper<Object, Text, Text, IntWritable> {
17
18         private final static IntWritable one = new IntWritable(1);
19         private final static Text word = new Text("Lines");
20
21         public void map(Object key, Text value, Context context
22             ) throws IOException, InterruptedException {
23             context.write(word, one);
24         }
25     }
26
27     public static class IntSumReducer
28         extends Reducer<Text, IntWritable, Text, IntWritable> {
29         private IntWritable result = new IntWritable();
30
31         public void reduce(Text key, Iterable<IntWritable> values,
32             Context context
33             ) throws IOException, InterruptedException {
34             int sum = 0;
35             for (IntWritable val : values) {
```

```

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

```

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