```
root@07c2f64b580a:/# rm CustomWordCount.java
root@07c2f64b580a:/# cat > CustomWordCount.java <<EOF
> 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;
>
> import java.io.IOException;
>
> public class CustomWordCount {
>     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 {
>             String[] tokens = value.toString().split("\\\\s+");
>             for (String token : tokens) {
```

```
>             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(CustomWordCount.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);
>     }
```

```
root@07c2f64b580a:/# ls -l CustomWordCount.java
-rw-r--r-- 1 root root 2123 Nov 18 11:18 CustomWordCount.java
root@07c2f64b580a:/# javac -cp $(hadoop classpath) -d . CustomWordCount.java
root@07c2f64b580a:/# jar cf CustomWordCount.jar CustomWordCount/*.class
CustomWordCount/*.class : no such file or directory
root@07c2f64b580a:/# hadoop jar CustomWordCount.jar CustomWordCount <input_path> <output_path>
bash: syntax error near unexpected token `<'
root@07c2f64b580a:/# hdfs dfs -mkdir /user/hadoop/custom_input
mkdir: `/user/hadoop/custom_input': File exists
root@07c2f64b580a:/# hdfs dfs -put /path/to/input_data.txt /user/hadoop/custom_input
put: `/path/to/input_data.txt': No such file or directory
root@07c2f64b580a:/# hadoop jar CustomWordCount.jar CustomWordCount /user/hadoop/custom_input /user/hadoop/custom_output
2024-11-18 11:20:53,219 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 11:20:53,332 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 11:20:53,333 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 11:20:53,538 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
2024-11-18 11:20:53,544 WARN mapreduce.JobResourceUploader: No job jar file set.  User classes may not be found. See Job or Job#set
Jar(String)
```

```
        File System Counters
                FILE: Number of bytes read=186
                FILE: Number of bytes written=522088
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=40
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=1
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=6
                Map output bytes=64
                Map output materialized bytes=82
                Input split bytes=129
                Combine input records=6
                Combine output records=6
                Spilled Records=6
                Failed Shuffles=0
```

```
                Reduce input groups=6
                Reduce shuffle bytes=82
                Reduce input records=6
                Reduce output records=6
                Spilled Records=12
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=21
                Total committed heap usage (bytes)=657981440
        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=40
        File Output Format Counters
                Bytes Written=52
```