

An Introduction to MapReduce with Word and Line Counting

This project demonstrates the use of **MapReduce**, a powerful programming model that simplifies data processing across large datasets. Using MapReduce, I created a solution to analyze a collection of text data (a set of poems) by performing two tasks:

1. **Counting Lines in Text Files:** The first task counts the number of lines in a collection of poems. By splitting up the text and processing each line separately, MapReduce quickly provides the total line count, even when working with large datasets.
2. **Counting Words by Length:** The second task takes it a step further, counting the number of words of each length in the poems. For example, it counts how many words have 4 letters, how many have 5 letters, and so on. This requires modifying the MapReduce program to count words based on their lengths, showing how flexible MapReduce can be.

How It Works:

MapReduce divides data processing into two main stages:

- **Mapping:** In the Mapper function, each line or word is examined individually. For the line-count task, each line produces a single count. For the word-length task, each word is mapped with its length (like “Length 5”) as a key.
- **Reducing:** In the Reducer function, the results from all Mappers are combined to produce the final output. Here, all the word-length counts are summed, giving us a breakdown of word lengths across the text.

This project demonstrates how MapReduce is used to quickly and efficiently analyze large datasets by breaking down data processing into manageable steps.

Why MapReduce?

MapReduce is widely used in big data and data science because it can handle enormous datasets across many computers, making it ideal for tasks like log analysis, word counts, or any application requiring data aggregation.

2.) Change the code to count how many words of different lengths are in the poems data.

The Mapper implementation

```
// -----  
// start of the map class  
  
import java.util.StringTokenizer; // you need to add this line  
import java.io.IOException;  
import org.apache.hadoop.io.*;  
import org.apache.hadoop.mapreduce.Mapper;  
  
public class MapIt extends Mapper<Object, Text, Text, IntWritable>{  
    private final IntWritable one = new IntWritable(1);  
    private Text wordLengthText = new Text();  
  
    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {  
        // -----  
        String mapInput = value.toString();  
  
        StringTokenizer tokenIterator = new StringTokenizer( mapInput, " \\t\\n\\r\\f\\.\\-\" );  
  
        while (tokenIterator.hasMoreTokens()) {  
            String theWord = tokenIterator.nextToken();  
  
            int length = theWord.length();  
  
            wordLengthText.set("Length " + length);  
  
            context.write(wordLengthText, one);  
        }  
  
        // --- your code should end here  
  
        // -----  
    }  
}  
  
// end of the map class  
// -----
```

The Reducer implementation

```
// start of the reduce class
import java.io.IOException;

import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.Reducer;

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

    public void reduce(Text key, Iterable<IntWritable> values, Context context
                        ) throws IOException, InterruptedException {
        // -----
        // --- your code should start here

        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }

        IntWritable result = new IntWritable(1);
        result.set(sum);

        context.write(key, result);
    }
    // --- your code should end here
    // -----
}

// end of the reduce class
// -----
```

The Combiner implementation

```
// -----  
// start of the combiner class  
  
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 MapReduceIt {  
    public static void main(String[] args) throws Exception {  
        Configuration conf = new Configuration();  
        // -----  
        // start of the area that may need change  
  
        Job job = Job.getInstance(conf, "test"); // name your job  
        job.setJarByClass(MapReduceIt.class); // define the class with main method  
        job.setMapperClass(MapIt.class); // define mapper class  
        job.setCombinerClass(ReduceIt.class); // define combiner class  
        job.setReducerClass(ReduceIt.class); // define reducer class  
        job.setOutputKeyClass(Text.class); // define type of the key  
        job.setOutputValueClass(IntWritable.class); // define type of the value  
        FileInputFormat.addInputPath(job, new Path(args[0]));  
        FileOutputFormat.setOutputPath(job, new Path(args[1]));  
  
        // end of the area that may need change  
        // -----  
        System.exit(job.waitForCompletion(true) ? 0 : 1);  
    }  
}  
  
// end of the combiner class  
// -----
```

Output for Question 2.)

Note that for this problem, the map function should output a <key, value> pair with the format <Text, IntWritable> and with examples <"Length 8", 1> or <"Length 7", 1>, or similar.

```
root@2b831f15fa94:~# mkdir /root/week10java
root@2b831f15fa94:~# cd /root/week10java/
root@2b831f15fa94:~/week10java# nano MapIt.java

root@2b831f15fa94:~/week10java# nano ReduceIt.java
root@2b831f15fa94:~/week10java# nano MapReduceIt.java
root@2b831f15fa94:~/week10java# cd /root/week10java
root@2b831f15fa94:~/week10java# hadoop com.sun.tools.javac.Main *.java
root@2b831f15fa94:~/week10java# jar cf mri.jar *.class
root@2b831f15fa94:~/week10java# ls /root/week10java
MapIt.class MapIt.java MapReduceIt.class MapReduceIt.java ReduceIt.class ReduceIt.java mri.jar
root@2b831f15fa94:~/week10java# which hadoop
/usr/local/hadoop/bin/hadoop
root@2b831f15fa94:~/week10java# cd /usr/local/hadoop
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop jar /root/week10java/mri.jar MapReduceIt /user/root/poems/input /user/root/poems/output

SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-10-28 20:27:09,598 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-28 20:27:09,797 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-28 20:27:09,809 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1698523446636_0001
2023-10-28 20:27:10,446 INFO input.FileInputFormat: Total input files to process : 6
2023-10-28 20:27:10,926 INFO mapreduce.JobSubmitter: number of splits:6
2023-10-28 20:27:11,020 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698523446636_0001
2023-10-28 20:27:11,020 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-28 20:27:11,119 INFO conf.Configuration: resource-types.xml not found
2023-10-28 20:27:11,119 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-28 20:27:11,291 INFO impl.YarnClientImpl: Submitted application application_1698523446636_0001
2023-10-28 20:27:11,318 INFO mapreduce.Job: The url to track the job: http://2b831f15fa94:8088/proxy/application_1698523446636_0001/
2023-10-28 20:27:11,318 INFO mapreduce.Job: Running job: job_1698523446636_0001
2023-10-28 20:27:17,398 INFO mapreduce.Job: Job job_1698523446636_0001 running in uber mode : false
2023-10-28 20:27:17,400 INFO mapreduce.Job: map 0% reduce 0%
2023-10-28 20:27:26,465 INFO mapreduce.Job: map 100% reduce 0%
2023-10-28 20:27:30,490 INFO mapreduce.Job: map 100% reduce 100%
2023-10-28 20:27:31,511 INFO mapreduce.Job: Job job_1698523446636_0001 completed successfully
2023-10-28 20:27:31,574 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=912
    FILE: Number of bytes written=1845340
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7064
    HDFS: Number of bytes written=149
    HDFS: Number of read operations=23
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=6
    Launched reduce tasks=1
    Data-local map tasks=6
    Total time spent by all maps in occupied slots (ms)=43885
    Total time spent by all reduces in occupied slots (ms)=1646
    Total time spent by all map tasks (ms)=43885
    Total time spent by all reduce tasks (ms)=1646
```

```
Total megabyte-milliseconds taken by all map tasks=44938240
Total megabyte-milliseconds taken by all reduce tasks=1685504
Map-Reduce Framework
  Map input records=231
  Map output records=1243
  Map output bytes=16169
  Map output materialized bytes=942
  Input split bytes=708
  Combine input records=1243
  Combine output records=60
  Reduce input groups=12
  Reduce shuffle bytes=942
  Reduce input records=60
  Reduce output records=12
  Spilled Records=120
  Shuffled Maps =6
  Failed Shuffles=0
  Merged Map outputs=6
  GC time elapsed (ms)=821
  CPU time spent (ms)=2110
  Virtual memory (bytes) snapshot=17707151360
  Peak Map Physical memory (bytes)=276508672
  Peak Map Virtual memory (bytes)=2530603008
  Peak Reduce Physical memory (bytes)=187953152
  Peak Reduce Virtual memory (bytes)=2534236160
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=6356
File Output Format Counters
  Bytes Written=149
root@2b831f15fa94:/usr/local/hadoop# hadoop fs -ls /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Found 2 items
-rw-r--r--  1 root supergroup          0 2023-10-28 20:27 /user/root/poems/output/_SUCCESS

-rw-r--r--  1 root supergroup        149 2023-10-28 20:27 /user/root/poems/output/part-r-000000
root@2b831f15fa94:/usr/local/hadoop# hadoop fs -cat /user/root/poems/output/part-r-000000
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Length 1      108
Length 10     5
Length 11     4
Length 12     1
Length 2      217
Length 3      259
Length 4      264
Length 5      183
Length 6      98
Length 7      60
Length 8      32
Length 9      12
root@2b831f15fa94:/usr/local/hadoop#
```



1.) Run the example on the poems data in our Hadoop server.

Example MapReduce: Counting the Lines in the Input Files

```
root@2b831f15fa94:~# mkdir /root/wk10clExample
root@2b831f15fa94:~# cd /root/wk10clExample
root@2b831f15fa94:~# nano MapIt.java
root@2b831f15fa94:~# nano ReduceIt.java
root@2b831f15fa94:~# nano MapReduceIt.java
root@2b831f15fa94:~# cd /root/wk10clExample
root@2b831f15fa94:~# hadoop com.sun.tools.javac.Main *.java
root@2b831f15fa94:~# jar cf mri.jar *.class
root@2b831f15fa94:~# ls /root/wk10clExample
MapIt.class MapIt.java MapReduceIt.class MapReduceIt.java ReduceIt.class ReduceIt.java mri.jar
root@2b831f15fa94:~# which hadoop
/usr/local/hadoop/bin/hadoop
root@2b831f15fa94:~# cd /usr/local/hadoop
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop fs -rm /user/root/poems/output/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Deleted /user/root/poems/output/_SUCCESS
Deleted /user/root/poems/output/part-r-000000
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop fs -rmkdir /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop jar /root/wk10clExample/mri.jar MapReduceIt /user/root/poems/input /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-10-28 21:02:10,476 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-28 21:02:10,671 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-28 21:02:10,696 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1698523446636_0002
2023-10-28 21:02:11,273 INFO input.FileInputFormat: Total input files to process : 6
2023-10-28 21:02:11,758 INFO mapreduce.JobSubmitter: number of splits:6
2023-10-28 21:02:11,860 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698523446636_0002
2023-10-28 21:02:11,860 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-28 21:02:11,956 INFO conf.Configuration: resource-types.xml not found
```

```
2023-10-28 21:02:31,282 INFO mapreduce.Job: Job job_1698523446636_0002 completed successfully
2023-10-28 21:02:31,332 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=78
  FILE: Number of bytes written=1843672
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=7064
  HDFS: Number of bytes written=10
  HDFS: Number of read operations=23
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=6
  Launched reduce tasks=1
  Data-local map tasks=6
  Total time spent by all maps in occupied slots (ms)=39449
  Total time spent by all reduces in occupied slots (ms)=1848
  Total time spent by all map tasks (ms)=39449
  Total time spent by all reduce tasks (ms)=1848
  Total vcore-milliseconds taken by all map tasks=39449
  Total vcore-milliseconds taken by all reduce tasks=1848
  Total megabyte-milliseconds taken by all map tasks=40395776
  Total megabyte-milliseconds taken by all reduce tasks=1892352
```

Map-Reduce Framework

Map input records=231
Map output records=231
Map output bytes=2310
Map output materialized bytes=108
Input split bytes=708
Combine input records=231
Combine output records=6
Reduce input groups=1
Reduce shuffle bytes=108
Reduce input records=6
Reduce output records=1
Spilled Records=12
Shuffled Maps =6
Failed Shuffles=0
Merged Map outputs=6
GC time elapsed (ms)=727
CPU time spent (ms)=1820
Physical memory (bytes) snapshot=1706758144
Virtual memory (bytes) snapshot=17704210432
Total committed heap usage (bytes)=1332740096
Peak Map Physical memory (bytes)=271503360
Peak Map Virtual memory (bytes)=2529296384
Peak Reduce Physical memory (bytes)=181460992
Peak Reduce Virtual memory (bytes)=2534502400

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=6356

File Output Format Counters

Bytes Written=10

root@2b831f15fa94:/usr/local/hadoop# hadoop fs -ls /user/root/poems/output

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

Found 2 items

-rw-r--r-- 1 root supergroup 0 2023-10-28 21:02 /user/root/poems/output/_SUCCESS

-rw-r--r-- 1 root supergroup 10 2023-10-28 21:02 /user/root/poems/output/part-r-00000

root@2b831f15fa94:/usr/local/hadoop# hadoop fs -cat /user/root/poems/output/part-r-00000

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

Total 231

Example: Counting Words Using MapReduce

```
root@2b831f15fa94:~# mkdir /root/wk10clExample2
root@2b831f15fa94:~# cd /root/wk10clExample2
root@2b831f15fa94:~/wk10clExample2# nano MapIt.java
root@2b831f15fa94:~/wk10clExample2# nano ReduceIt.java
root@2b831f15fa94:~/wk10clExample2# nano MapReduceIt.java
root@2b831f15fa94:~/wk10clExample2# cd /root/wk10clExample2
root@2b831f15fa94:~/wk10clExample2# hadoop com.sun.tools.javac.Main *.java
root@2b831f15fa94:~/wk10clExample2# jar cf mri.jar *.class
root@2b831f15fa94:~/wk10clExample2# ls /root/wk10clExample2
MapIt.class MapIt.java MapReduceIt.class MapReduceIt.java ReduceIt.class ReduceIt.java mri.jar
root@2b831f15fa94:~/wk10clExample2# which hadoop
/usr/local/hadoop/bin/hadoop
root@2b831f15fa94:~/wk10clExample2# cd /usr/local/hadoop
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop fs -rm /user/root/poems/output/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Deleted /user/root/poems/output/_SUCCESS
Deleted /user/root/poems/output/part-r-00000
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop fs -rmrdir /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
root@2b831f15fa94:/usr/local/hadoop# bin/hadoop jar /root/wk10clExample2/mri.jar MapReduceIt /user/root/poems/input /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-10-28 21:34:51,124 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-28 21:34:51,334 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-28 21:34:51,346 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1698523446636_0003
2023-10-28 21:34:51,483 INFO input.FileInputFormat: Total input files to process : 6
2023-10-28 21:34:52,020 INFO mapreduce.JobSubmitter: number of splits:6
2023-10-28 21:34:52,104 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698523446636_0003
2023-10-28 21:34:52,105 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-28 21:34:52,198 INFO conf.Configuration: resource-types.xml not found
2023-10-28 21:34:52,199 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-28 21:34:52,239 INFO impl.YarnClientImpl: Submitted application application_1698523446636_0003
2023-10-28 21:34:52,268 INFO mapreduce.Job: The url to track the job: http://2b831f15fa94:8088/proxy/application_1698523446636_0003/
2023-10-28 21:34:52,271 INFO mapreduce.Job: Running job: job_1698523446636_0003
2023-10-28 21:34:56,338 INFO mapreduce.Job: Job job_1698523446636_0003 running in uber mode : false
2023-10-28 21:34:56,339 INFO mapreduce.Job: map 0% reduce 0%
2023-10-28 21:35:04,484 INFO mapreduce.Job: map 17% reduce 0%
2023-10-28 21:35:05,506 INFO mapreduce.Job: map 100% reduce 0%
```

```
2023-10-28 21:35:10,537 INFO mapreduce.Job: Job job_1698523446636_0003 completed successfully
2023-10-28 21:35:10,599 INFO mapreduce.Job: Counters: 54
```

File System Counters

```
FILE: Number of bytes read=7804
FILE: Number of bytes written=1859124
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=7064
HDFS: Number of bytes written=4205
HDFS: Number of read operations=23
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

Job Counters

```
Launched map tasks=6
Launched reduce tasks=1
Data-local map tasks=6
Total time spent by all maps in occupied slots (ms)=41402
Total time spent by all reduces in occupied slots (ms)=1832
Total time spent by all map tasks (ms)=41402
Total time spent by all reduce tasks (ms)=1832
Total vcore-milliseconds taken by all map tasks=41402
Total vcore-milliseconds taken by all reduce tasks=1832
Total megabyte-milliseconds taken by all map tasks=42395648
Total megabyte-milliseconds taken by all reduce tasks=1875968
```

```

Map-Reduce Framework
  Map input records=231
  Map output records=1243
  Map output bytes=10983
  Map output materialized bytes=7834
  Input split bytes=708
  Combine input records=1243
  Combine output records=672
  Reduce input groups=521
  Reduce shuffle bytes=7834
  Reduce input records=672
  Reduce output records=521
  Spilled Records=1344
  Shuffled Maps =6
  Failed Shuffles=0
  Merged Map outputs=6
  GC time elapsed (ms)=756
  CPU time spent (ms)=1960
  Physical memory (bytes) snapshot=1630928896
  Virtual memory (bytes) snapshot=17702481920
  Total committed heap usage (bytes)=1262485504
  Peak Map Physical memory (bytes)=249942016
  Peak Map Virtual memory (bytes)=2529353728
  Peak Reduce Physical memory (bytes)=189112320
  Peak Reduce Virtual memory (bytes)=2535092224

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=6356

File Output Format Counters
  Bytes Written=4205

root@2b831f15fa94:/usr/local/hadoop# hadoop fs -ls /user/root/poems/output
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Found 2 items
-rw-r--r--  1 root supergroup          0 2023-10-28 21:35 /user/root/poems/output/_SUCCESS
-rw-r--r--  1 root supergroup    4205 2023-10-28 21:35 /user/root/poems/output/part-r-00000
root@2b831f15fa94:/usr/local/hadoop# hadoop fs -cat /user/root/poems/output/part-r-00000
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

```

Final output

a	46	attached	1	big	1	clasp?	1	deep,	2
about	1	auschwitz,	1	bit	2	clear	1	deepest	1
above	1	avow	1	black	6	clearly	1	despair,	1
absolute	1	away	1	blackboard,	1	cleft	1	devil	1
ach,	1	baby	1	blood	1	cold	2	die	2
achoo	1	baby,	1	blue	2	cold!	1	died	4
adores	1	back	2	bones	1	common	1	dislike	1
all	2	back,	2	boot	1	could	4	do	6
allan	1	back,	2	born	1	could've	1	do,	2
always	4	bag	1	brain	1	couldn't,	1	dogged,	1
am	1	bank	1	breathe	1	creep	1	doing,my	1
amid	1	barb	1	breeze	1	cried!	2	done	1
an	3	barely	1	bright	1	cry	1	down	3
ancestress	1	bastard,	1	brow!	1	cummings	1	downy	1
and	40	be	10	brute	3	dachau,	1	dozen	1
and,	1	bean	1	bud	2	daddy	1	drank	1
any	2	beautiful	2	buried	1	daddy,	6	dread	1
apart	1	been	4	but	11	dancing	1	dream	5
are	9	beer	1	by	10	daring	1	dream;	1
aryan	1	before	3	called	1	dark	1	dream?	1
as	2	began	1	came	2	darkest	1	drive	1
ask	1	bells	1	can	5	darling)	1	du	1
at	3	belsen	1	can't	1	day,	1	e	2
atlantic	1	between	3	carry	5	dear;	1	ear,	1
atop	1			catastrophe,	1	deem	1	easy	1
				chin	1				
				chuffing	1				

edgar	1		gobbledygoo	1	if	9	
elevator		1	god	1	imagine	1	
engine	1		god!	2	in	24	
engine,	1		god,	1	instead	1	
even	1		golden	1	instead,		1
evening	2		gone?	1	is	19	
every	2		gonna	1	it	16	
eye,	1		grains	1	it's	1	
eyes	1		grasp	1	it(anywhere		1
face	1		gray	1	jack	1	
face!	1		green	1	jaw	1	
face,	2		ground	1	jew	5	
farmhouse		1	grows	1	jump	1	
fascist,		1	guess	1	jumped	2	
fat	1		had	2	just	1	
fate(for		1	hadn't	2	keep,	1	
fate,my	1		hair	1	keeping	1	
fear	1		hand	1	kill	1	
feet	1		hardly	1	killed	2	
few!	1		harness	1	kiss	1	
fill	1		has	2	knew	2	
finally	1		have	6	know	2	
fine	2		he	3	knows	1	
fine!	2		head	1	lake	1	
fingers	1		head,	1	langston		1
flake	1		hear	1	language		1
flat	1		heart	5	less	3	
floors	1		heart(i	1	let	1	
flown	1		heart)	1	lie	1	
foot	2		heart)i	1	life	3	
foot,	1		heavy,	1	life;which		1
for	8		here	3	like	4	
forced	1		hide)	1	liked	1	
forever	1		high	2	little	1	
freakish		1	high!	1	live	1	
friend	1		higher	1	lived	1	
frisco	1		his	3	livin'	1	
from	3		hold	1	livin',	1	
frost	1		holler,	1	look	1	
frozen	1		hollered!		lot	1	
full	1		hope	2	love	2	
german	2		horse	1	lovely,	1	
get	1		house	1	luck	1	
ghastly	1		how	2	luftwaffe,		1
gipsy	1		hughes	1	made	1	
gives	1		i	57	man	2	
glad	3		i'll	1	man,	3	
glue	1		i'm	3	marble	1	
go	3		i've	2	may	4	
go,my	1		ich,	4	me	10	

me(i	1	place,	1	sleep	1	therefore	1
meant	1	plath	1	sleep,	1	these	1
meinkampf	1	poe	1	smell	1	they	6
might	1	polack	1	snare	1	thick	1
might've	2	polish	1	sneeze,	1	thin,	1
miles	2	poor	1	snow	1	think	4
mind	1	pours	1	snows	1	thirty	1
mistake	1	pray	1	snowy	1	this	2
model	1	precious	1	so	7	though	1
moon	1	prelutsky	1	some	3	though;	1
more,	1	pretty	1	soon	1	thought	4
much	1	promises	1	soul	1	through	6
must	1	pulled	1	sound's	1	thus	1
mustache	1	pure	1	source	1	tickled	1
my	17	put	1	speak	1	tighter	1
name	1	queer	1	squeak	1	time	1
nauset	1	rack	1	stake	1	to	23
near	1	rattle	1	stamping	1	toe	1
neat	1	recover	1	stand	2	toes,	1
never	4	red	1	stars	1	together	1
night,	1	remains	1	statue	1	tongue	1
no	5	river,	1	still	1	tongue,	1
nobody	1	roar	1	stood	2	took	1
none,	1	robert	1	stop	1	topping	1
nose	7	roller	1	stopping	1	tormented	1
nose,	1	root	2	stuck	3	town	2
not	12	root,	2	sun	1	treat,	1
not,	1	sack,	1	sunk	1	tree	1
now	1	said	2	surf	1	tried	2
now,	1	sand	1	swastika	1	true	1
o	3	sandwiched	1	sweep	1	true)	1
obliged	1	sank	1	sweet	1	twenty	1
obscene	1	save	1	sweet)i	1	twice	1
of	22	says	1	sylvia	1	two	3
off	3	scared	1	take	1	tyrol,	1
on	9	scraped	1	talk	2	up	4
once	1	screw	1	taroc	2	upon	1
one	3	seal	1	telephone's	1	used	1
only	2	secret	1	tell	1	vampire	1
or	8	see	5	ten	1	very	1
other	3	seem	2	than	1	vienna	1
out	1	set	1	that	6	village,	1
over	1	seven	1	that's	1	villagers	1
pack	2	shake	1	that,	1	vision,	1
panzer	2	shoe	1	the	55	voices	1
parting	1	shore,	1	them	1	want	2
pasted	2	since	1	then	1	wars	1
picture	1	sing	1	there	4	wars,	2
pitiless	1	sixteen	1	there!	1	was	9
place	1	sky	3	there's	1	watch	1

```
water 1
water! 1
waters 1
wave? 1
we 2
weep 1
weep! 1
weird 1
well 1
went 1
were 4
what 1
whatever 3
when 2
where 3
which 1
while 2
white, 1
who 3
whose 1
will 3
wind 1
wine! 1
wire 1
with 9
within 5
without 2
woman 1
wonder 1
woods 5
world(for 1
world,my 1
worm 1
would 7
wrong, 1
year 1
year, 1
years, 2
yet 2
you 33
you'd 1
you, 3
your 31
root@2b831f15fa94:/usr/local/hadoop#
```