

Creacion de archivo ejecutable

Mi programa es un contador de palabras totales. Aqui la creacion del archivo .java y del .jar

```
root@namenode:/home/hadoop# nano TotalWordCount.java
root@namenode:/home/hadoop# rm TotalWordCount.java
root@namenode:/home/hadoop# nano TotalWordCount.java
root@namenode:/home/hadoop# javac -cp "$(hadoop classpath)" -d build TotalWordCount.java
root@namenode:/home/hadoop# jar -cvf TotalWordCount.jar -C build/ .
manifiesto agregado
agregando: TotalWordCount$TokenizerMapper.class(entrada = 1860) (salida = 775)(desinflado 58%)
agregando: TotalWordCount.class(entrada = 1511) (salida = 818)(desinflado 45%)
agregando: TotalWordCount$SumReducer.class(entrada = 1748) (salida = 744)(desinflado 57%)
root@namenode:/home/hadoop# su hadoop
```

El codigo es el siguiente:

```
hadoop@namenode:~$ cat TotalWordCount.java
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
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;
import java.util.StringTokenizer;

public class TotalWordCount {

    public static class TokenizerMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private final static Text totalKey = new Text("Total Words");

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

    public static class SumReducer 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, "total word count");
        job.setJarByClass(TotalWordCount.class);
        job.setMapperClass(TokenizerMapper.class);
        job.setCombinerClass(SumReducer.class);
        job.setReducerClass(SumReducer.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);
    }
}
```

Ejecución del contador total de palabras:

```

hadoop@namenode:~$ hadoop jar ./totalWordCount.jar totalWordCount /text.txt /TWC
2025-02-11 21:16:44,988 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at namenode/192.168.1.1:8032
2025-02-11 21:16:45,288 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-02-11 21:16:45,322 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1739302010926_0003
2025-02-11 21:16:45,563 INFO input.FileInputFormat: Total input files to process : 1
2025-02-11 21:16:46,564 INFO mapreduce.JobSubmitter: number of splits:1
2025-02-11 21:16:46,734 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1739302010926_0003
2025-02-11 21:16:46,734 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-02-11 21:16:46,882 INFO conf.Configuration: resource-types.xml not found
2025-02-11 21:16:46,883 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-02-11 21:16:46,988 INFO impl.YarnClientImpl: Submitted application application_1739302010926_0003
2025-02-11 21:16:47,034 INFO mapreduce.Job: The url to track the job: http://namenode:8088/proxy/application_1739302010926_0003/
2025-02-11 21:16:47,035 INFO mapreduce.Job: Running job: job_1739302010926_0003
2025-02-11 21:16:52,174 INFO mapreduce.Job: Job job_1739302010926_0003 running in uber mode : false
2025-02-11 21:16:52,176 INFO mapreduce.Job: map 0% reduce 0%
2025-02-11 21:16:56,281 INFO mapreduce.Job: map 100% reduce 0%
2025-02-11 21:17:00,351 INFO mapreduce.Job: map 100% reduce 100%
2025-02-11 21:17:01,416 INFO mapreduce.Job: Job job_1739302010926_0003 completed successfully
2025-02-11 21:17:01,502 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=24
    FILE: Number of bytes written=616733
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=2141614
    HDFS: Number of bytes written=19
    HDFS: Number of read operations=0
    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=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=2120
    Total time spent by all reduces in occupied slots (ms)=1753
    Total time spent by all map tasks (ms)=2120
    Total time spent by all reduce tasks (ms)=1753
    Total vcore-milliseconds taken by all map tasks=2120
    Total vcore-milliseconds taken by all reduce tasks=1753
    Total megabyte-milliseconds taken by all map tasks=2170880
    Total megabyte-milliseconds taken by all reduce tasks=1795872

```

```

Map-Reduce Framework
  Map input records=37454
  Map output records=381217
  Map output bytes=6099472
  Map output materialized bytes=24
  Input split bytes=94
  Combine input records=381217
  Combine output records=1
  Reduce input groups=1
  Reduce shuffle bytes=24
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=245
  CPU time spent (ms)=1510
  Physical memory (bytes) snapshot=720871424
  Virtual memory (bytes) snapshot=5140574208
  Total committed heap usage (bytes)=645398528
  Peak Map Physical memory (bytes)=479678464
  Peak Map Virtual memory (bytes)=2570821632
  Peak Reduce Physical memory (bytes)=241192960
  Peak Reduce Virtual memory (bytes)=2569752576
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=2141520
File Output Format Counters
  Bytes Written=19

```

Comprobación:

```

hadoop@namenode:~$ hdfs dfs -cat /TWC/part-r-00000
Total Words      381217

```