```
Using username "maria_dev".
maria_dev@localhost's password:
Last login: Wed Sep  3 13:44:52 2025 from 172.18.0.3
```

```
[maria_dev@sandbox-hdp ~]$ echo "Hadoop is big data Hadoop is java" > example.txt
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /input
mkdir: `/input': File exists
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /Input
[maria_dev@sandbox-hdp ~]$ hdfs dfs -put example.txt /Input/
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Input
Found 1 items
-rw-r--r--   1 maria_dev hdfs         34 2025-09-03 13:58 /Input/example.txt
```

```java
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
public class WordMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();
    public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException {
        String line = value.toString();
        for (String token : line.split("\\s+")) {  // Corrected escape sequence
            word.set(token);
            context.write(word, one);
        }
    }
}
~
~
~
~
~
~
~
~
```

```
[maria_dev@sandbox-hdp ~]$ vi WordMapper.java
```

```java
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
public class WordReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOExce
ption,
InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        context.write(key, new IntWritable(sum));
    }
}
~
~
```

```
[maria_dev@sandbox-hdp ~]$ vi WordReducer.java
```

```
maria_dev@sandbox-hdp:~                                              —    □    ×

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.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class WordCountDriver {
    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Usage: WordCountDriver <input path> <output path>");
            System.exit(-1);
        }
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "word count");
        job.setJarByClass(WordCountDriver.class);
        job.setMapperClass(WordMapper.class);
        job.setReducerClass(WordReducer.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);
    }
}
~
~
~
~
~
~
~
```

```
[maria_dev@sandbox-hdp ~]$ vi WordCountDriver.java
[maria_dev@sandbox-hdp ~]$
```

```
[maria_dev@sandbox-hdp ~]$ javac -cp `hadoop classpath` -d wordcount_classes WordMapper.java
WordReducer.java WordCountDriver.java
[maria_dev@sandbox-hdp ~]$ ls wordcount_classes
WordCountDriver.class  WordMapper.class  WordReducer.class
[maria_dev@sandbox-hdp ~]$
```

```
[maria_dev@sandbox-hdp ~]$ jar -cvf wordcount.jar -C wordcount_classes/ .
added manifest
adding: WordMapper.class(in = 1867) (out= 777)(deflated 58%)
adding: WordReducer.class(in = 1592) (out= 664)(deflated 58%)
adding: WordCountDriver.class(in = 1535) (out= 853)(deflated 44%)
[maria_dev@sandbox-hdp ~]$ ls
example.txt  u.data        wordcount_classes    wordcount.jar      WordReducer.java
sample.txt   u_local.data  WordCountDriver.java  WordMapper.java
[maria_dev@sandbox-hdp ~]$
```

```
[maria_dev@sandbox-hdp ~]$ hadoop jar wordcount.jar WordCountDriver /Input /Output
25/09/03 14:11:07 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp.hortonwor
ks.com/172.18.0.2:8032
25/09/03 14:11:08 INFO client.AHSProxy: Connecting to Application History server at sandbox-h
dp.hortonworks.com/172.18.0.2:10200
25/09/03 14:11:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to remed
y this.
25/09/03 14:11:09 INFO input.FileInputFormat: Total input paths to process : 1
25/09/03 14:11:09 INFO mapreduce.JobSubmitter: number of splits:1
25/09/03 14:11:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1756907870141_0
001
25/09/03 14:11:11 INFO impl.YarnClientImpl: Submitted application application_1756907870141_0
001
25/09/03 14:11:12 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.hortonwork
s.com:8088/proxy/application_1756907870141_0001/
25/09/03 14:11:12 INFO mapreduce.Job: Running job: job_1756907870141_0001
25/09/03 14:11:36 INFO mapreduce.Job: Job job_1756907870141_0001 running in uber mode : false
25/09/03 14:11:36 INFO mapreduce.Job:  map 0% reduce 0%
25/09/03 14:12:00 INFO mapreduce.Job:  map 100% reduce 0%
25/09/03 14:12:25 INFO mapreduce.Job:  map 100% reduce 100%
25/09/03 14:12:27 INFO mapreduce.Job: Job job_1756907870141_0001 completed successfully
25/09/03 14:12:28 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=82
                FILE: Number of bytes written=305979
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=156
                HDFS: Number of bytes written=34
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        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)=21470
                Total time spent by all reduces in occupied slots (ms)=22039
                Total time spent by all map tasks (ms)=21470
                Total time spent by all reduce tasks (ms)=22039
                Total vcore-milliseconds taken by all map tasks=21470
                Total vcore-milliseconds taken by all reduce tasks=22039
                Total megabyte-milliseconds taken by all map tasks=5367500
                Total megabyte-milliseconds taken by all reduce tasks=5509750
        Map-Reduce Framework
                Map input records=1
                Map output records=7
                Map output materialized bytes=82
                Input split bytes=122
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=82
                Reduce input records=7
                Reduce output records=5
                Spilled Records=14
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=382
                CPU time spent (ms)=4430
                Physical memory (bytes) snapshot=342204416
                Virtual memory (bytes) snapshot=3906027520
                Total committed heap usage (bytes)=152567808
        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=34
        File Output Format Counters
                Bytes Written=34
maria_dev@sandbox-hdp ~]$
```

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Output
Found 2 items
-rw-r--r--   1 maria_dev hdfs          0 2025-09-03 14:12 /Output/_SUCCESS
-rw-r--r--   1 maria_dev hdfs         34 2025-09-03 14:12 /Output/part-r-00000
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat /Output/part-r-00000
Hadoop  2
big     1
data    1
is      2
java    1
[maria_dev@sandbox-hdp ~]$
```