Lab 2

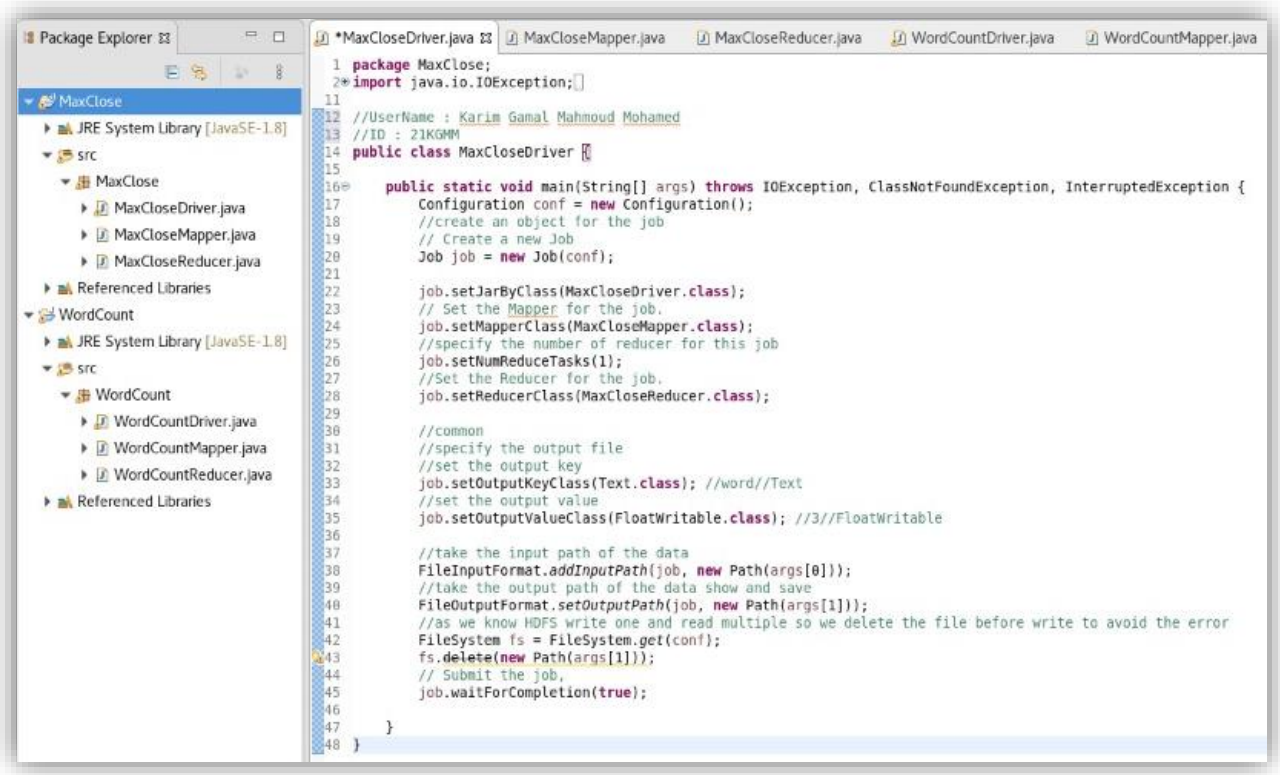
Name: Karim Gamal Mahmoud Mohamed

ID: 21KGMM

1. The jar file

The Jar file is uploaded with the Word file.

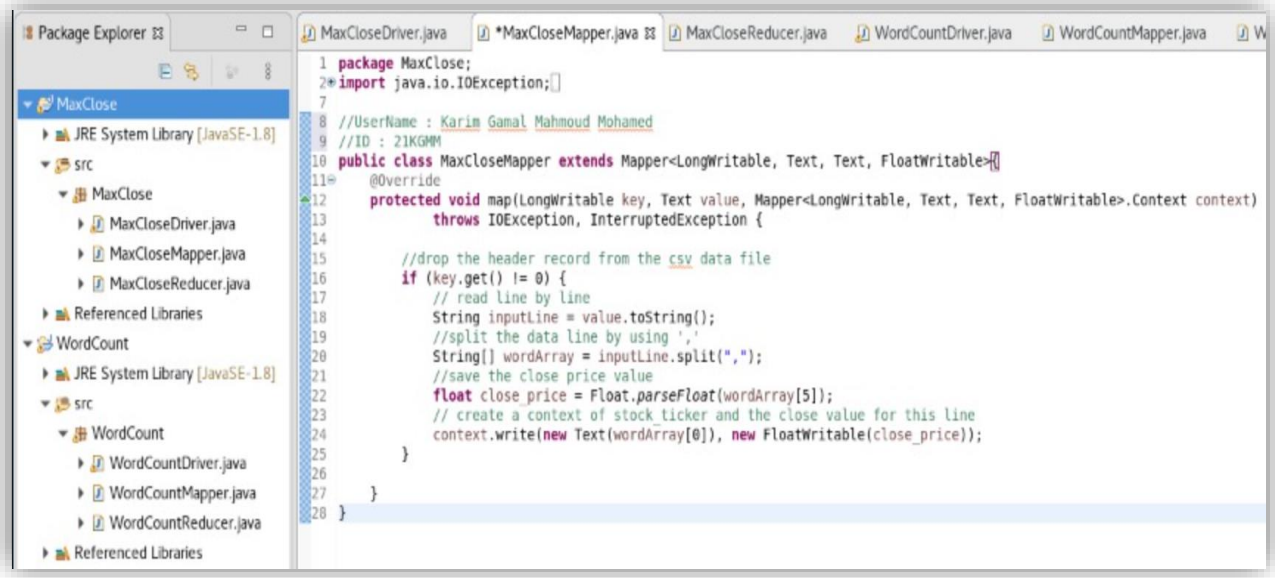
2-a) java source code for Driver :



```
package MaxClose;

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.FileSystem;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.FloatWritable;

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;


//UserName  :  Karim Gamal Mahmoud Mohamed

//ID  :  21KGMM

public class MaxCloseDriver {


public static void main(String[] args) throws IOException, ClassNotFoundException, InterruptedException
{

Configuration conf = new Configuration();

//create an object for the job

// Create a new Job

Job job = new Job(conf);

job.setJarByClass(MaxCloseDriver.class);

// Set the Mapper for the job.

job.setMapperClass(MaxCloseMapper.class);

//specify the number of reducer for this job

job.setNumReduceTasks(1);

//Set the Reducer for the job.

job.setReducerClass(MaxCloseReducer.class);
```

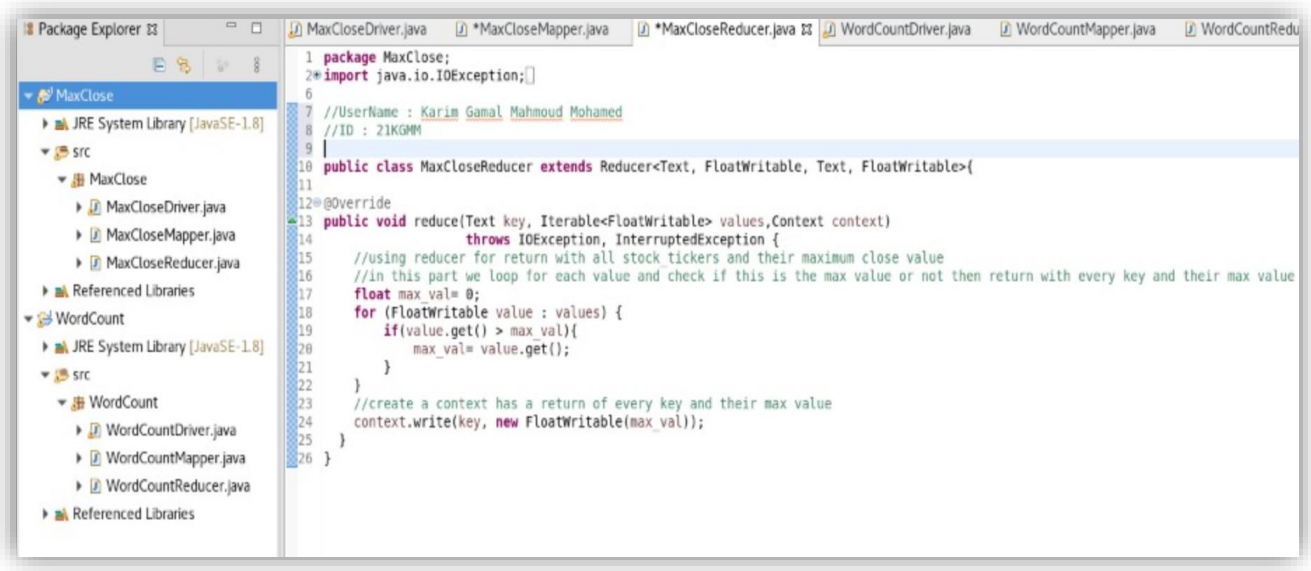
```
//common
//specify the output file
//set the output key
job.setOutputKeyClass(Text.class); //word//Text
//set the output value
job.setOutputValueClass(FloatWritable.class); //3//FloatWritable

//take the input path of the data
FileInputFormat.addInputPath(job, new Path(args[0]));
//take the output path of the data show and save
FileOutputFormat.setOutputPath(job, new Path(args[1]));
//as we know HDFS write one and read multiple so we delete the file before write to avoid the error
FileSystem fs = FileSystem.get(conf);
fs.delete(new Path(args[1]));
// Submit the job,
job.waitForCompletion(true);

}
}
```

2-b) java source code for Mapper :



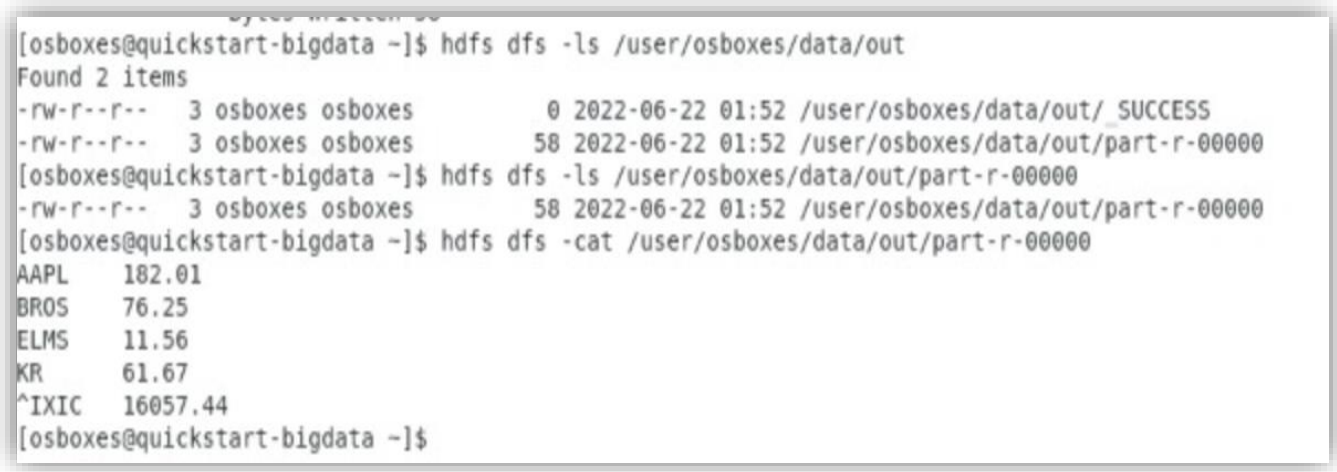
```
package MaxClose;

import java.io.IOException;

import org.apache.hadoop.io.FloatWritable;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Mapper;


//UserName  :  Karim Gamal Mahmoud Mohamed

//ID  :  21KGMM

public class MaxCloseMapper extends Mapper<LongWritable, Text, Text, FloatWritable>{

@Override

protected void map(LongWritable key, Text value, Mapper<LongWritable, Text, Text,
FloatWritable>.Context context)

throws IOException, InterruptedException {


//drop the headr record from the csv data file

if (key.get() != 0) {

// read line by line

String inputLine = value.toString();

//split the data line by using ','

String[] wordArray = inputLine.split(",");

//save the close price value

float close_price = Float.parseFloat(wordArray[5]);

// create a context of stock_ticker and the close value for this line

context.write(new Text(wordArray[0]), new FloatWritable(close_price));

}}}
```

2-c) java source code for Reducer:



```
package MaxClose;

import java.io.IOException;

import org.apache.hadoop.io.FloatWritable;

import org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.io.Text;


//UserName   :   Karim Gamal Mahmoud Mohamed

//ID   :   21KGMM


public class MaxCloseReducer extends Reducer<Text, FloatWritable, Text, FloatWritable>{


@Override

public void reduce(Text key, Iterable<FloatWritable> values,Context context)

                throws IOException, InterruptedException {

//using reducer for return with all stock_tickers and their maximum close value

//in this part we loop for each value and check if this is the max value or not then return with every key and their max value

   float max_val= 0;

   for (FloatWritable value : values) {

      if(value.get() > max_val){

           max_val= value.get();

      }}

   //create a context has a return of every key and their max value

   context.write(key, new FloatWritable(max_val));

 }}
```

3. The stocks.txt data file



-Stocks.txt data file :

AAPL 182.01

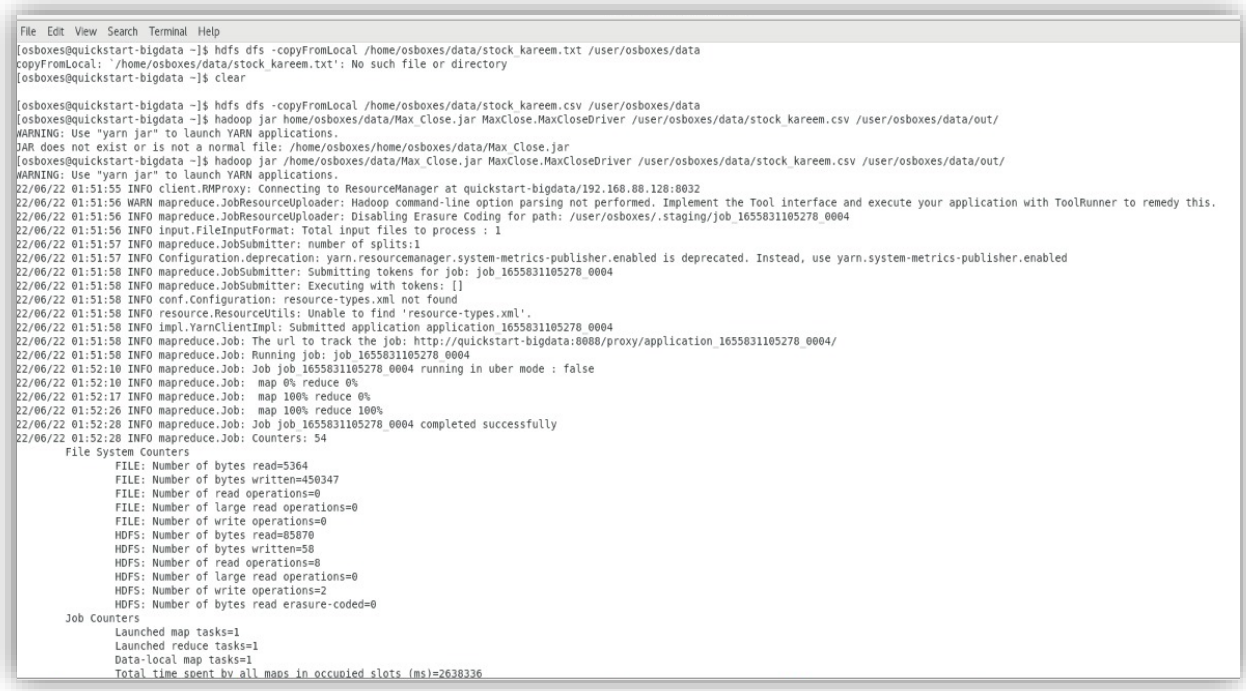
BROS 76.25

ELMS 11.56

KR 61.67

^IXIC 16057.44

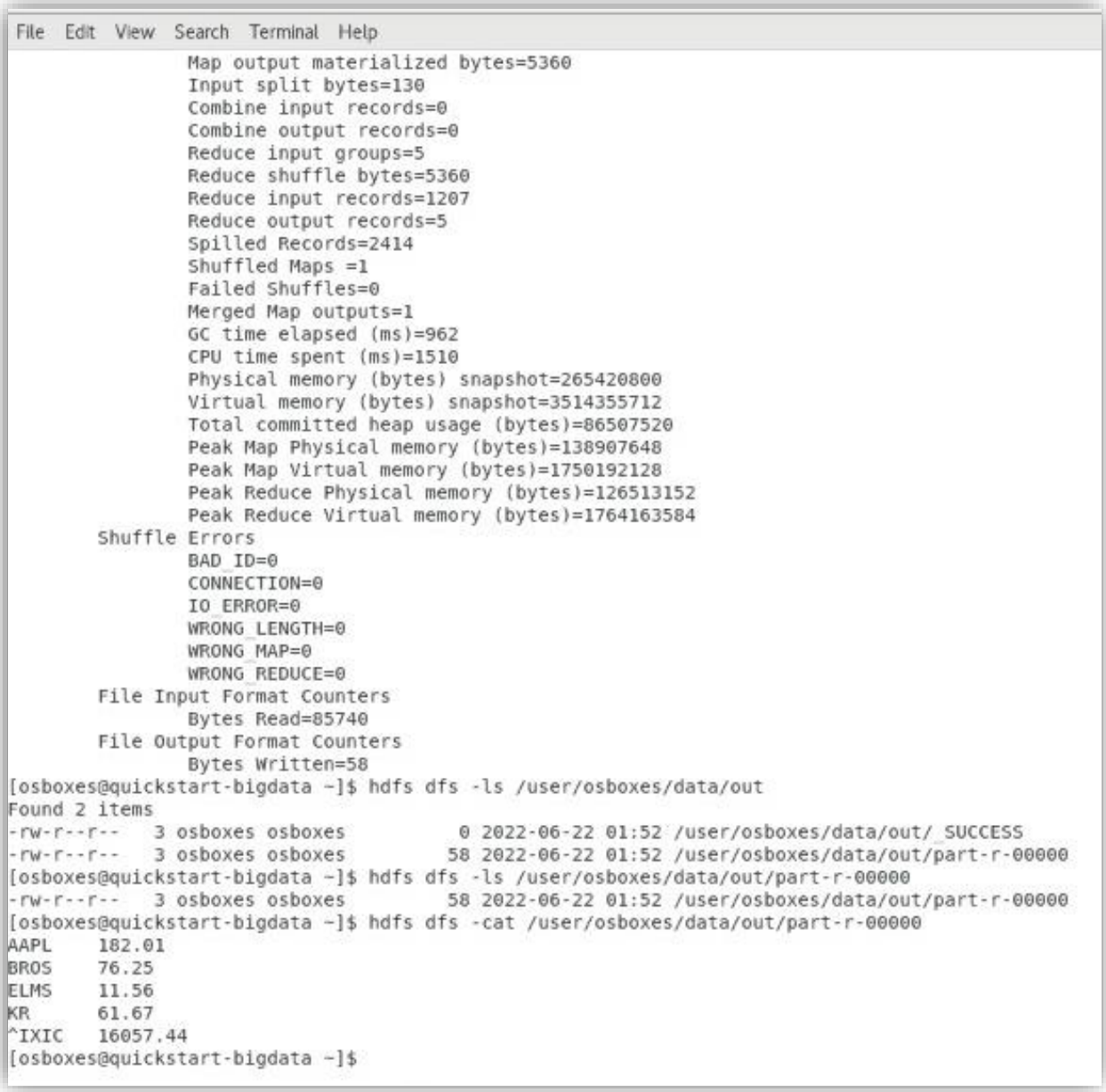
This is the final results after the Mapreduce program is completed :



```
File  Edit  View  Search  Terminal  Help
[osboxes@quickstart-bigdata ~]$ hdfs dfs -copyFromLocal /home/osboxes/data/stock_kareem.txt /user/osboxes/data
copyFromLocal: `/home/osboxes/data/stock_kareem.txt': No such file or directory
[osboxes@quickstart-bigdata ~]$ clear

[osboxes@quickstart-bigdata ~]$ hdfs dfs -copyFromLocal /home/osboxes/data/stock_kareem.csv /user/osboxes/data
[osboxes@quickstart-bigdata ~]$ hadoop jar home/osboxes/data/Max_Close.jar MaxClose.MaxCloseDriver /user/osboxes/data/stock_kareem.csv /user/osboxes/data/out/
WARNING: Use "yarn jar" to launch YARN applications.
JAR does not exist or is not a normal file: /home/osboxes/home/osboxes/data/Max_Close.jar
[osboxes@quickstart-bigdata ~]$ hadoop jar /home/osboxes/data/Max_Close.jar MaxClose.MaxCloseDriver /user/osboxes/data/stock_kareem.csv /user/osboxes/data/out/
WARNING: Use "yarn jar" to launch YARN applications.
22/06/22 01:51:55 INFO client.RMProxy: Connecting to ResourceManager at quickstart-bigdata/192.168.88.128:8032
22/06/22 01:51:56 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/06/22 01:51:56 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/osboxes/.staging/job_1655831105278_0004
22/06/22 01:51:56 INFO input.FileInputFormat: Total input files to process : 1
22/06/22 01:51:57 INFO mapreduce.JobSubmitter: number of splits:1
22/06/22 01:51:57 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
22/06/22 01:51:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655831105278_0004
22/06/22 01:51:58 INFO mapreduce.JobSubmitter: Executing with tokens: []
22/06/22 01:51:58 INFO conf.Configuration: resource-types.xml not found
22/06/22 01:51:58 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/06/22 01:51:58 INFO impl.YarnClientImpl: Submitted application application_1655831105278_0004
22/06/22 01:51:58 INFO mapreduce.Job: The url to track the job: http://quickstart-bigdata:8088/proxy/application_1655831105278_0004/
22/06/22 01:51:58 INFO mapreduce.Job: Running job: job_1655831105278_0004
22/06/22 01:52:10 INFO mapreduce.Job: Job job_1655831105278_0004 running in uber mode : false
22/06/22 01:52:10 INFO mapreduce.Job:  map 0% reduce 0%
22/06/22 01:52:17 INFO mapreduce.Job:  map 100% reduce 0%
22/06/22 01:52:26 INFO mapreduce.Job:  map 100% reduce 100%
22/06/22 01:52:28 INFO mapreduce.Job: Job job_1655831105278_0004 completed successfully
22/06/22 01:52:28 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=5364
                FILE: Number of bytes written=450347
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85870
                HDFS: Number of bytes written=58
                HDFS: Number of read operations=8
                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)=2638336
```

File Edit View Search Terminal Help

```
                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)=2638336
                Total time spent by all reduces in occupied slots (ms)=2988032
                Total time spent by all map tasks (ms)=5153
                Total time spent by all reduce tasks (ms)=5836
                Total vcore-milliseconds taken by all map tasks=5153
                Total vcore-milliseconds taken by all reduce tasks=5836
                Total megabyte-milliseconds taken by all map tasks=2638336
                Total megabyte-milliseconds taken by all reduce tasks=2988032
        Map-Reduce Framework
                Map input records=1208
                Map output records=1207
                Map output bytes=10609
                Map output materialized bytes=5360
                Input split bytes=130
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=5360
                Reduce input records=1207
                Reduce output records=5
                Spilled Records=2414
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=962
                CPU time spent (ms)=1510
                Physical memory (bytes) snapshot=265420800
                Virtual memory (bytes) snapshot=3514355712
                Total committed heap usage (bytes)=86507520
                Peak Map Physical memory (bytes)=138907648
                Peak Map Virtual memory (bytes)=1750192128
                Peak Reduce Physical memory (bytes)=126513152
                Peak Reduce Virtual memory (bytes)=1764163584
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
```


File Edit View Search Terminal Help
Map output materialized bytes=5360
Input split bytes=130
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=5360
Reduce input records=1207
Reduce output records=5
Spilled Records=2414
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=962
CPU time spent (ms)=1510
Physical memory (bytes) snapshot=265420800
Virtual memory (bytes) snapshot=3514355712
Total committed heap usage (bytes)=86507520
Peak Map Physical memory (bytes)=138907648
Peak Map Virtual memory (bytes)=1750192128
Peak Reduce Physical memory (bytes)=126513152
Peak Reduce Virtual memory (bytes)=1764163584
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=85740
File Output Format Counters
Bytes Written=58
[osboxes@quickstart-bigdata ~]$ hdfs dfs -ls /user/osboxes/data/out
Found 2 items
-rw-r--r-- 3 osboxes osboxes 0 2022-06-22 01:52 /user/osboxes/data/out/_SUCCESS
-rw-r--r-- 3 osboxes osboxes 58 2022-06-22 01:52 /user/osboxes/data/out/part-r-00000
[osboxes@quickstart-bigdata ~]$ hdfs dfs -ls /user/osboxes/data/out/part-r-00000
-rw-r--r-- 3 osboxes osboxes 58 2022-06-22 01:52 /user/osboxes/data/out/part-r-00000
[osboxes@quickstart-bigdata ~]$ hdfs dfs -cat /user/osboxes/data/out/part-r-00000
AAPL 182.01
BROS 76.25
ELMS 11.56
KR 61.67
^IXIC 16057.44
[osboxes@quickstart-bigdata ~]$