



# **CSE488: Big Data Analytics [SPRING 2023]**

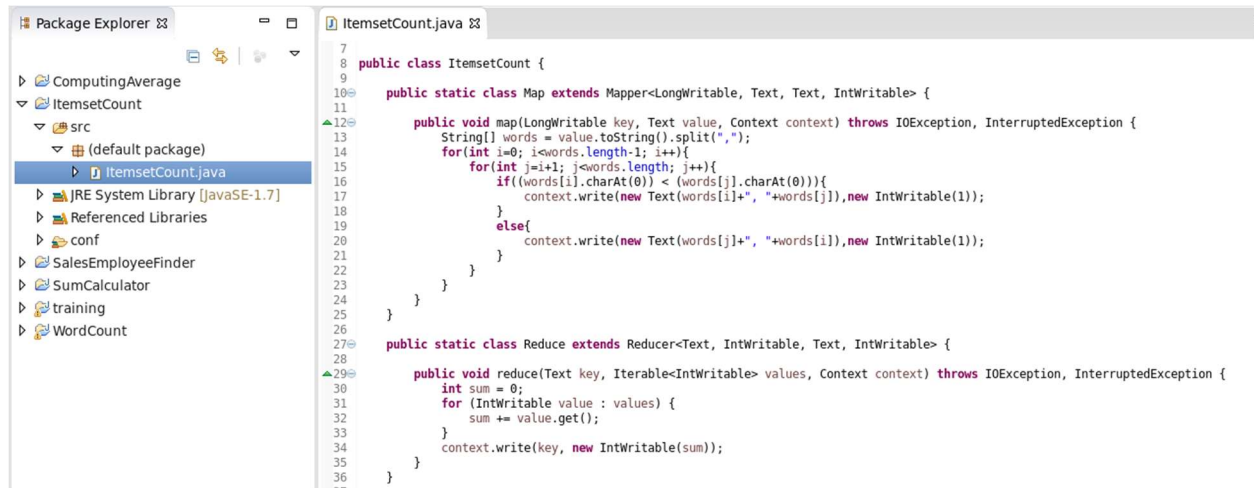
## **Lab-5 MapReduce to Itemset Count**

**Submitted by:**

Student ID: 2019-3-60-046  
Student Name: Mohsenul Kabir Mithun

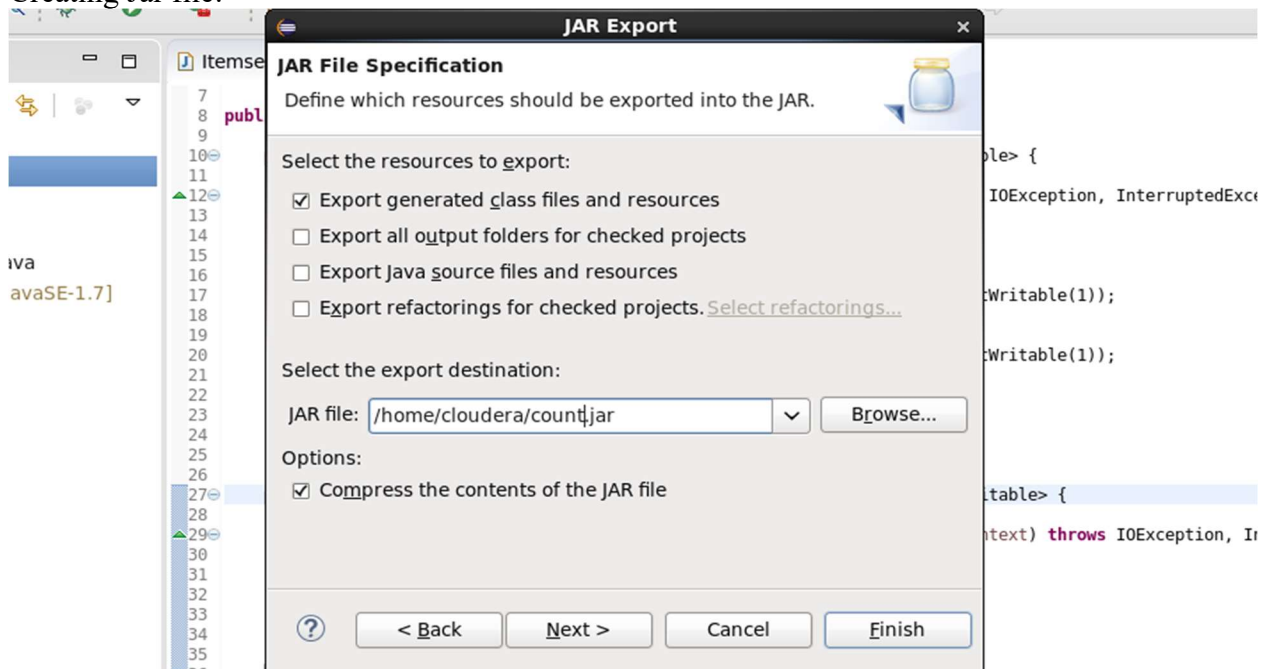
## Creating a Hadoop project in Eclipse: (MapReduce used for ItemsetCount)

Source Code SS: (Mapper and Reducer Class)

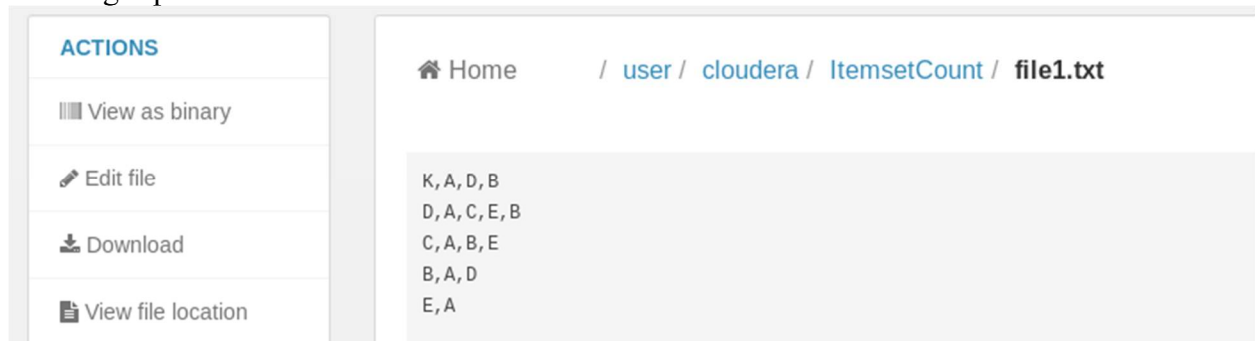


```
7 public class ItemsetCount {
8
9
10
11
12 public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
13
14     public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
15         String[] words = value.toString().split(",");
16         for(int i=0; i<words.length-1; i++){
17             for(int j=i+1; j<words.length; j++){
18                 if((words[i].charAt(0)) < (words[j].charAt(0))){
19                     context.write(new Text(words[i]+" "+words[j]), new IntWritable(1));
20                 }
21             }
22         }
23     }
24 }
25
26 public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {
27
28     public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
29         int sum = 0;
30         for (IntWritable value : values) {
31             sum += value.get();
32         }
33         context.write(key, new IntWritable(sum));
34     }
35 }
36 }
```

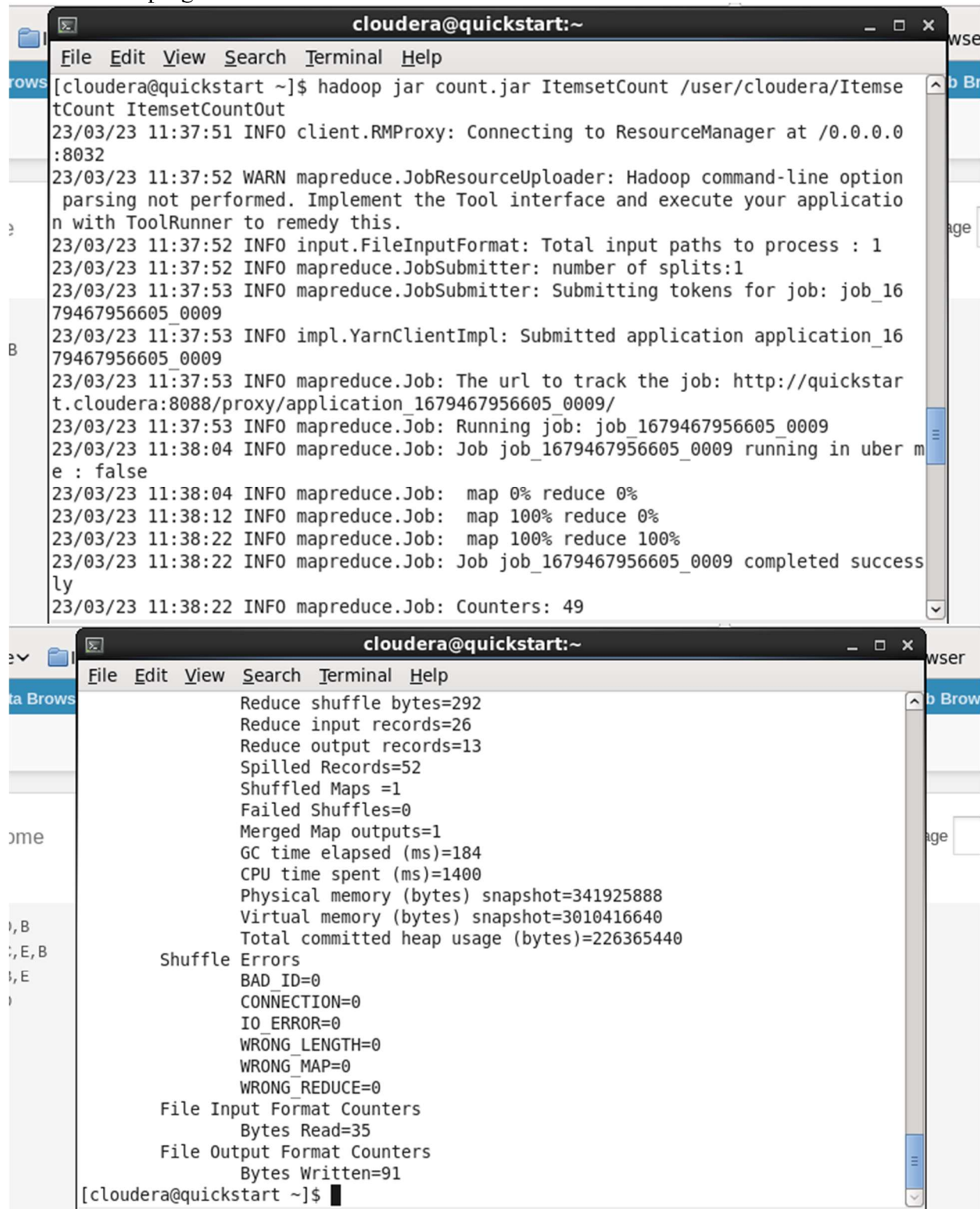
Creating Jar file:



Creating Input file:



CMD run the program:



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar count.jar ItemsetCount /user/cloudera/ItemsetCount ItemsetCountOut  
23/03/23 11:37:51 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032  
23/03/23 11:37:52 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
23/03/23 11:37:52 INFO input.FileInputFormat: Total input paths to process : 1  
23/03/23 11:37:52 INFO mapreduce.JobSubmitter: number of splits:1  
23/03/23 11:37:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1679467956605_0009  
23/03/23 11:37:53 INFO impl.YarnClientImpl: Submitted application application_1679467956605_0009  
23/03/23 11:37:53 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1679467956605_0009/  
23/03/23 11:37:53 INFO mapreduce.Job: Running job: job_1679467956605_0009  
23/03/23 11:38:04 INFO mapreduce.Job: Job job_1679467956605_0009 running in uber mode : false  
23/03/23 11:38:04 INFO mapreduce.Job: map 0% reduce 0%  
23/03/23 11:38:12 INFO mapreduce.Job: map 100% reduce 0%  
23/03/23 11:38:22 INFO mapreduce.Job: map 100% reduce 100%  
23/03/23 11:38:22 INFO mapreduce.Job: Job job_1679467956605_0009 completed successfully  
23/03/23 11:38:22 INFO mapreduce.Job: Counters: 49  
  
Reduce shuffle bytes=292  
Reduce input records=26  
Reduce output records=13  
Spilled Records=52  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=184  
CPU time spent (ms)=1400  
Physical memory (bytes) snapshot=341925888  
Virtual memory (bytes) snapshot=3010416640  
Total committed heap usage (bytes)=226365440  
  
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=35  
File Output Format Counters  
Bytes Written=91  
[cloudera@quickstart ~]$
```

OutPut:

ACTIONS

View as binary

Edit file

Download

View file location

Refresh

INFO

Last modified

March 22, 2023 10:38 p.m.

Home / user / cloudera / ItemsetCountOut / part-r-00000

A, B	4
A, C	2
A, D	3
A, E	3
A, K	1
B, C	2
B, D	3
B, E	2
B, K	1
C, D	1
C, E	2
D, E	1
D, K	1