Roll no: 33301
Name: Ankita Adam
Batch: K11

Problem Statement: Design an application using MapReduce which processes CSV file and counts the frequency of each word

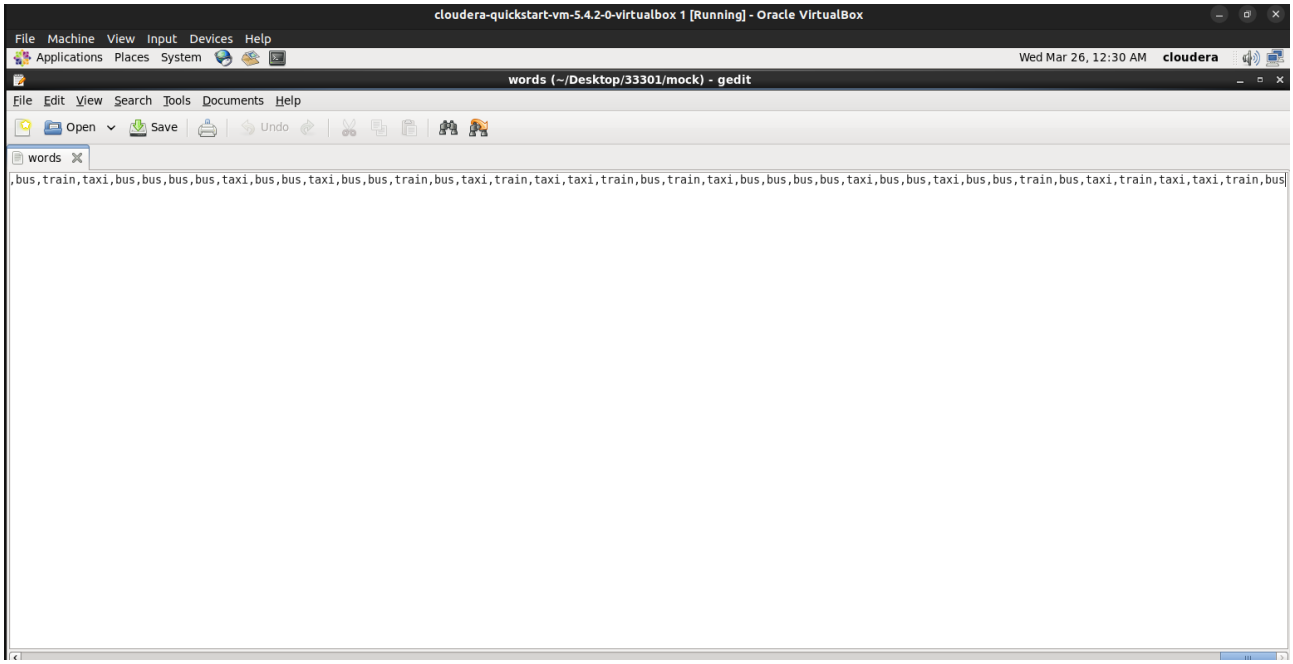
Code:

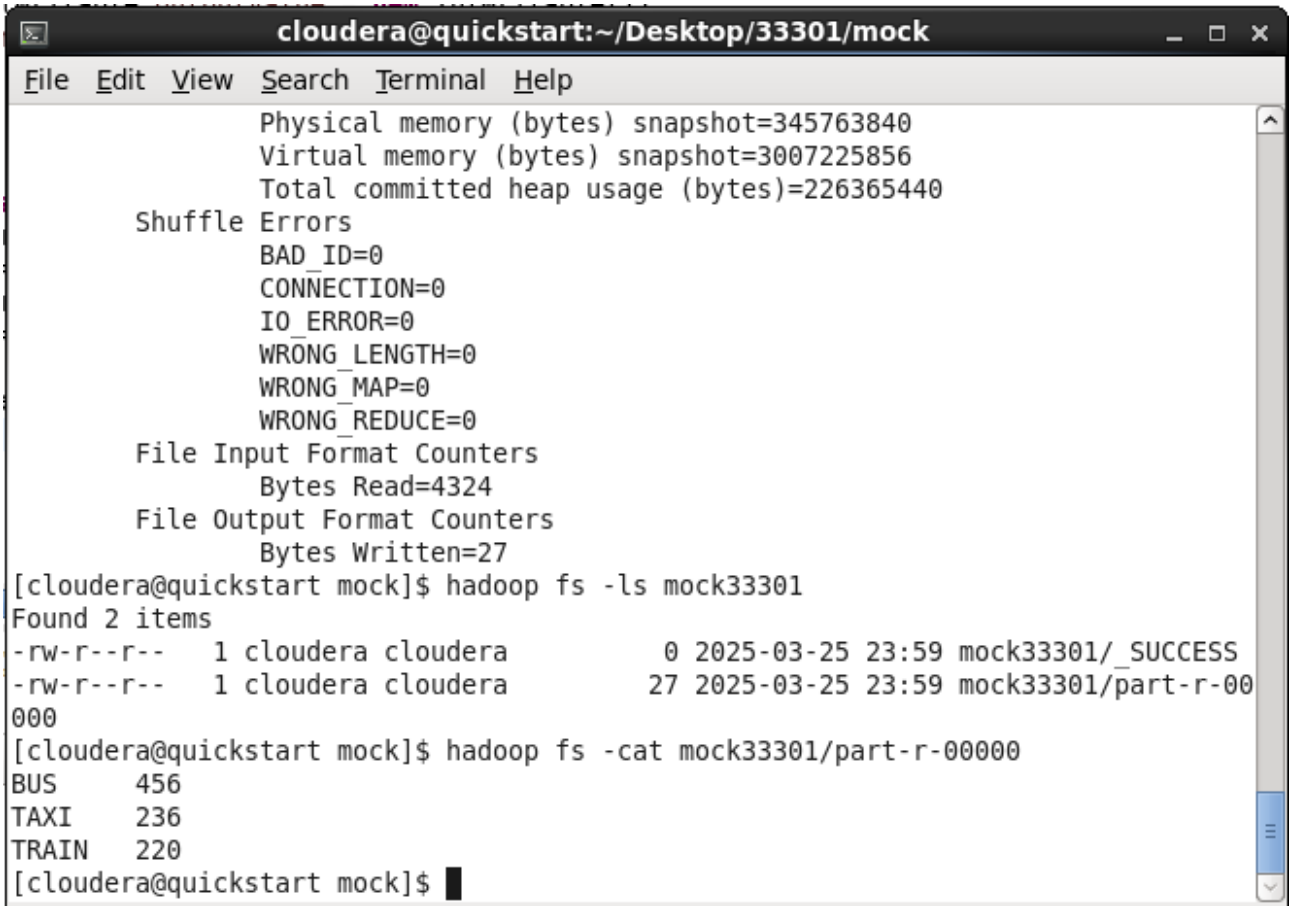
```
package demo;


import java.io.IOException;
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 org.apache.hadoop.util.GenericOptionsParser;

public class WordCount {
	public static void main(String [] args) throws Exception {
		Configuration c=new Configuration();
		String[] files=new GenericOptionsParser(c,args).getRemainingArgs();
		Path input=new Path(files[0]);
		Path output=new Path(files[1]);
		Job j=new Job(c,"wordcount");
		j.setJarByClass(WordCount.class);
		j.setMapperClass(MapForWordCount.class);
		j.setReducerClass(ReduceForWordCount.class);
		j.setOutputKeyClass(Text.class);
		j.setOutputValueClass(IntWritable.class);
		FileInputFormat.addInputPath(j, input);
		FileOutputFormat.setOutputPath(j, output);
		System.exit(j.waitForCompletion(true)?0:1);
	}

	public static class MapForWordCount extends Mapper<LongWritable, Text, Text,
IntWritable> {
		public void map(LongWritable key, Text value, Context con) throws IOException,
InterruptedException {
			String line = value.toString();
			String[] words=line.split(",");
			for(String word: words ) {
				Text outputKey = new Text(word.toUpperCase().trim());
				IntWritable outputValue = new IntWritable(1);
				con.write(outputKey, outputValue);
```

```
                }
            }
        }

        public static class ReduceForWordCount extends Reducer<Text, IntWritable, Text,
IntWritable> {
            public void reduce(Text word, Iterable<IntWritable> values, Context con) throws
IOException, InterruptedException {
                int sum = 0;
                for(IntWritable value : values){
                  sum += value.get();
                }
                con.write(word, new IntWritable(sum));
            }

        }

}
```

Output:

words: csv file




cloudera@quickstart:~/Desktop/33301/mock
File Edit View Search Terminal Help
Physical memory (bytes) snapshot=345763840
Virtual memory (bytes) snapshot=3007225856
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=4324
File Output Format Counters
Bytes Written=27
[cloudera@quickstart mock]$ hadoop fs -ls mock33301
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2025-03-25 23:59 mock33301/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 27 2025-03-25 23:59 mock33301/part-r-00000
[cloudera@quickstart mock]$ hadoop fs -cat mock33301/part-r-00000
BUS 456
TAXI 236
TRAIN 220
[cloudera@quickstart mock]$