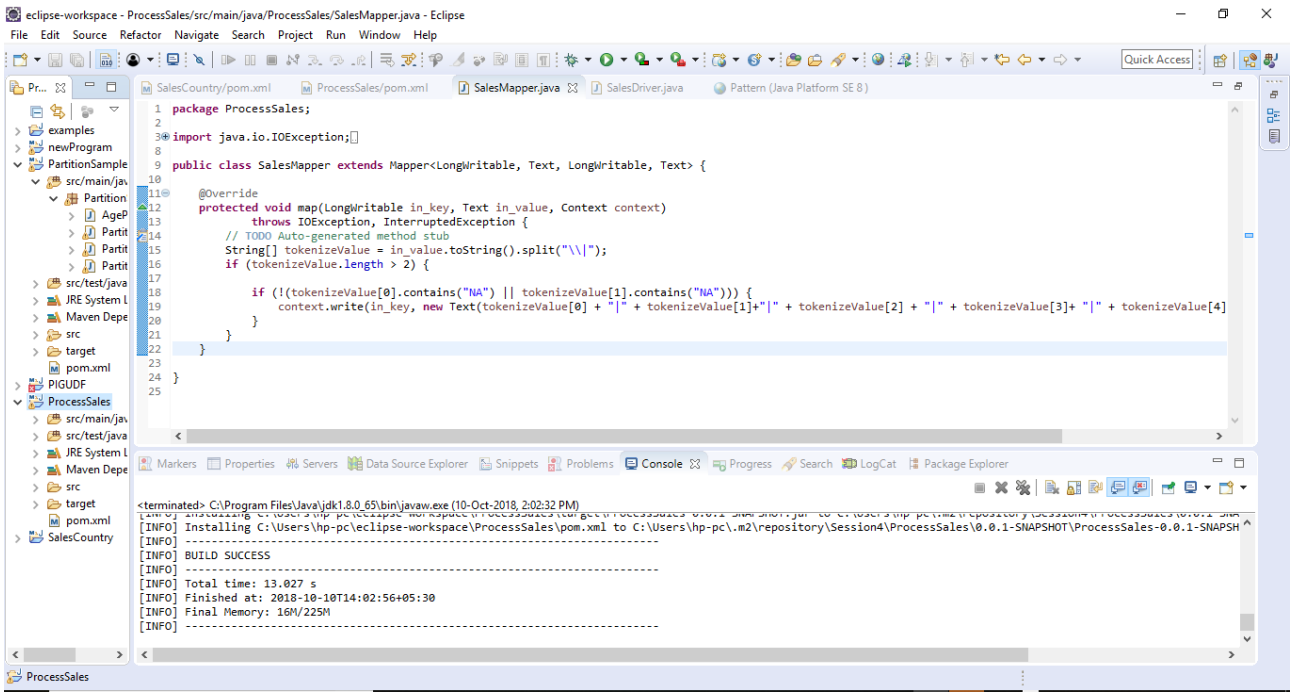
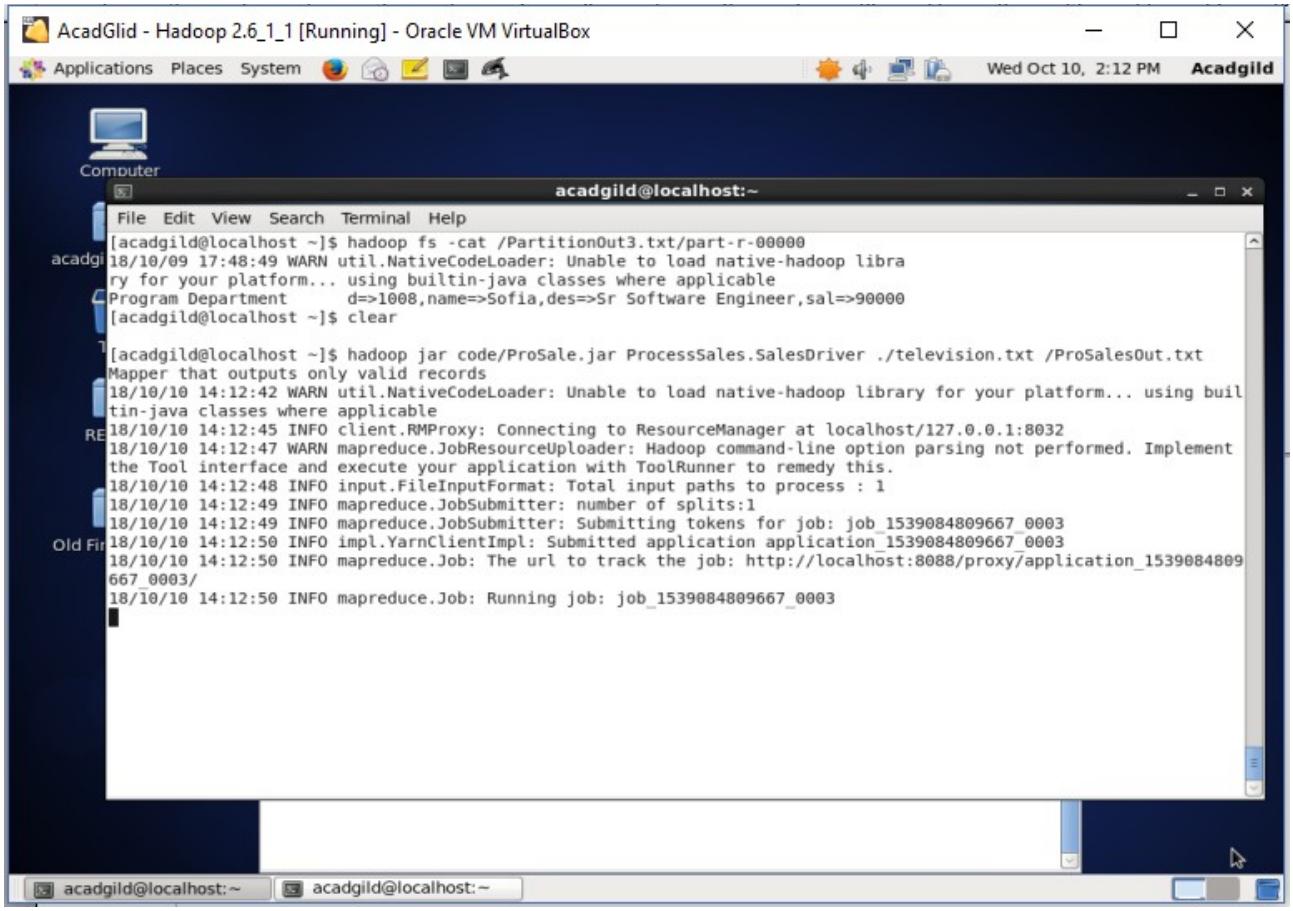
Task 1

Filter out records that content 'NA' in Company Name or Product Name

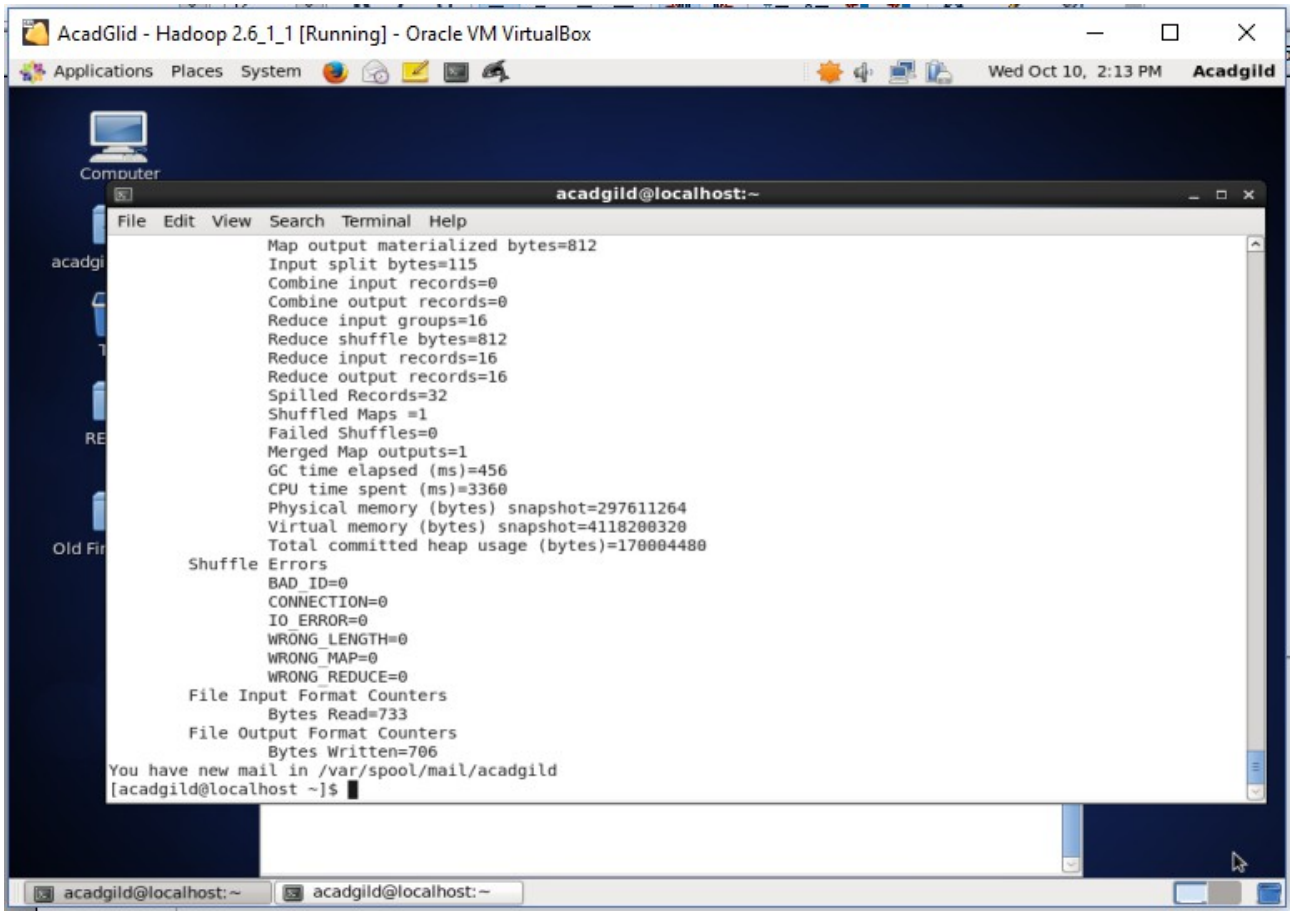


1. Uploaded television.txt data file onto hadoop HDFS using command
   hadoop fs -put television.txt input/television.txt
2. Copied ProSale.jar JAR file from windows machine to unix machine using WinSCP
   Command to run the Mapreduce program

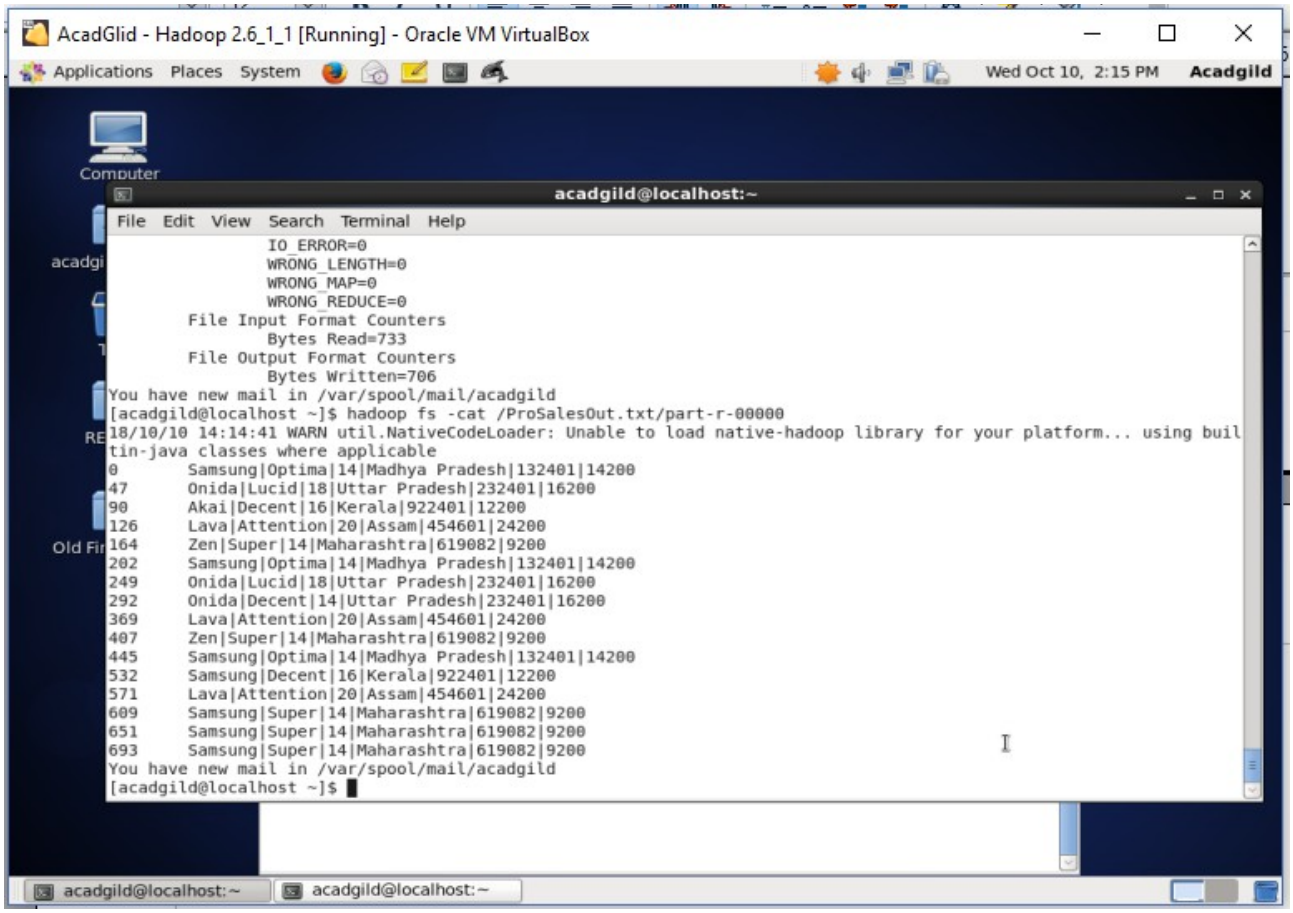
   hadoop jar ProSale.JAR ProcessSales.SalesDrive input/television.txt ProSalesOut.txt



On successful run



Checking the output



Task 2

Calculate total units sold for each company

1. Copied ProSaleCompWise.jar JAR file from windows machine to unix machine using WinSCP

Mapper code

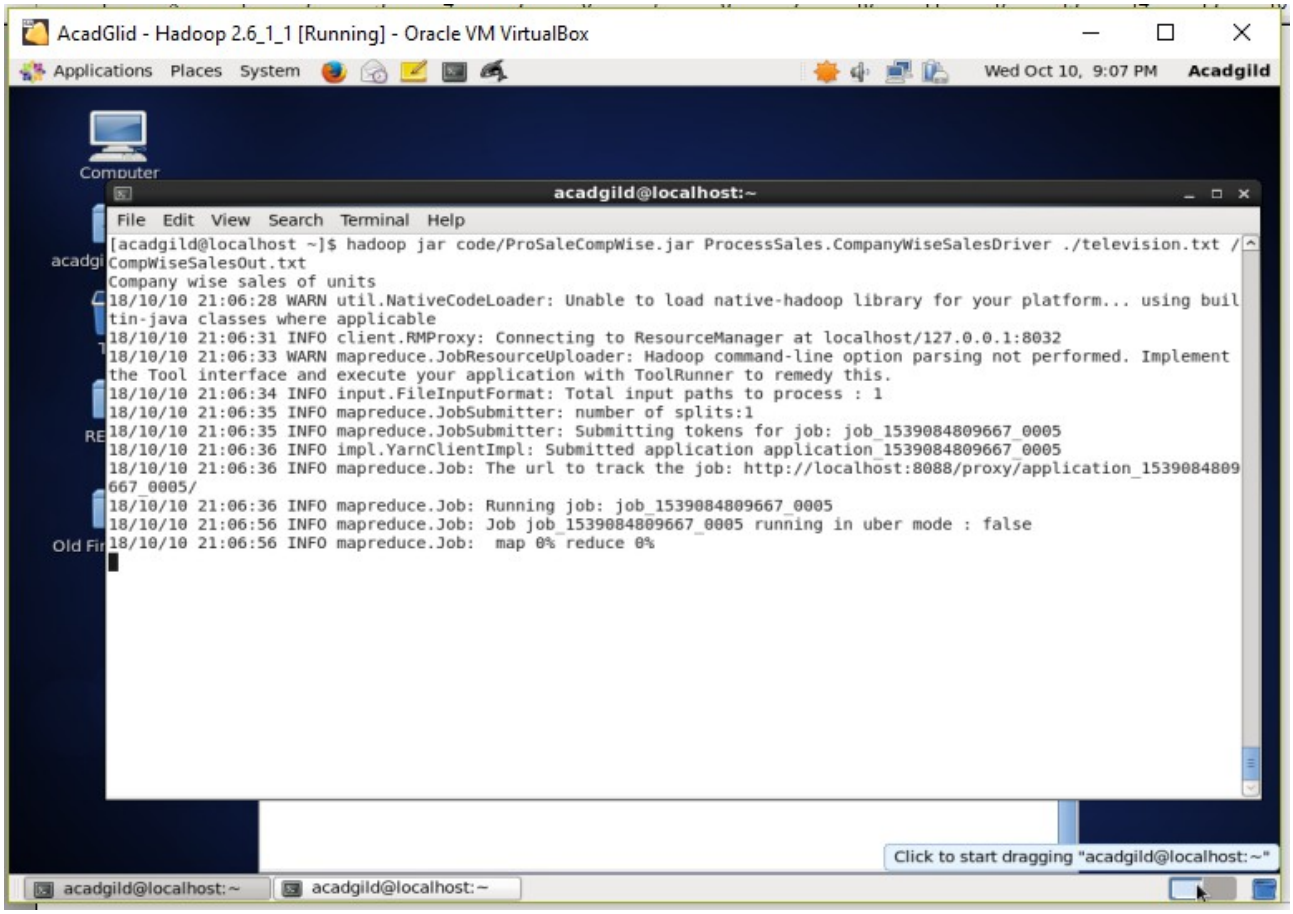
```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Mapper.Context;

public class CompanyWiseSalesMapper  extends Mapper<LongWritable, Text, Text, IntWritable> {

    @Override
    protected void map(LongWritable in_key, Text in_value, Context context)
            throws IOException, InterruptedException {
        // TODO Auto-generated method stub
        String[] tokenizeValue = in_value.toString().split("\\|");
        if (tokenizeValue.length > 2) {

            if (!(tokenizeValue[0].contains("NA") || tokenizeValue[1].contains("NA"))) {
                context.write(new Text(tokenizeValue[0]), new IntWritable(1));
            }
        }
    }

}
```

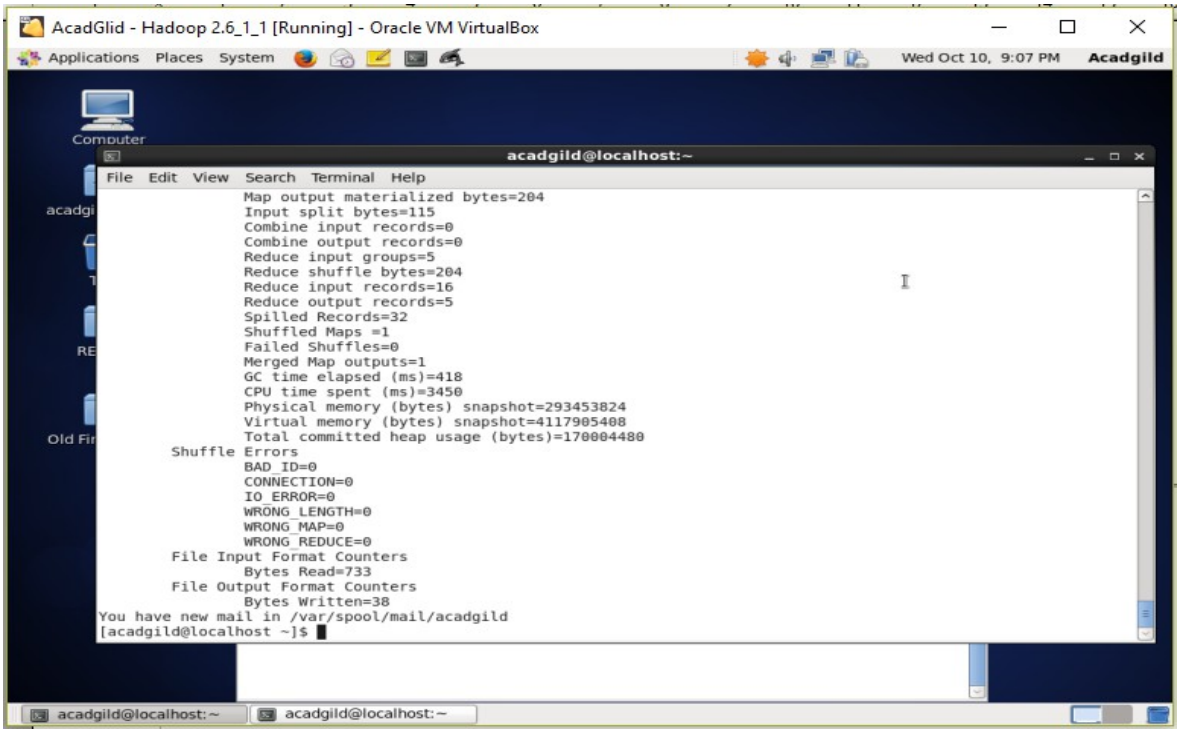
Reducer code

```
package ProcessSales;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.Reducer.Context;

public class CompanyWiseSalesReducer extends Reducer<Text, IntWritable,Text, IntWritable> {

    @Override
    protected void reduce(Text in_key, Iterable<IntWritable> in_value,
            Context context) throws IOException, InterruptedException {
        // TODO Auto-generated method stub
        int sum_value = 0;
        for(IntWritable node_value:in_value){
            sum_value += Integer.parseInt(node_value.toString());
        }
        context.write(in_key, new IntWritable(sum_value));
    }
}
```

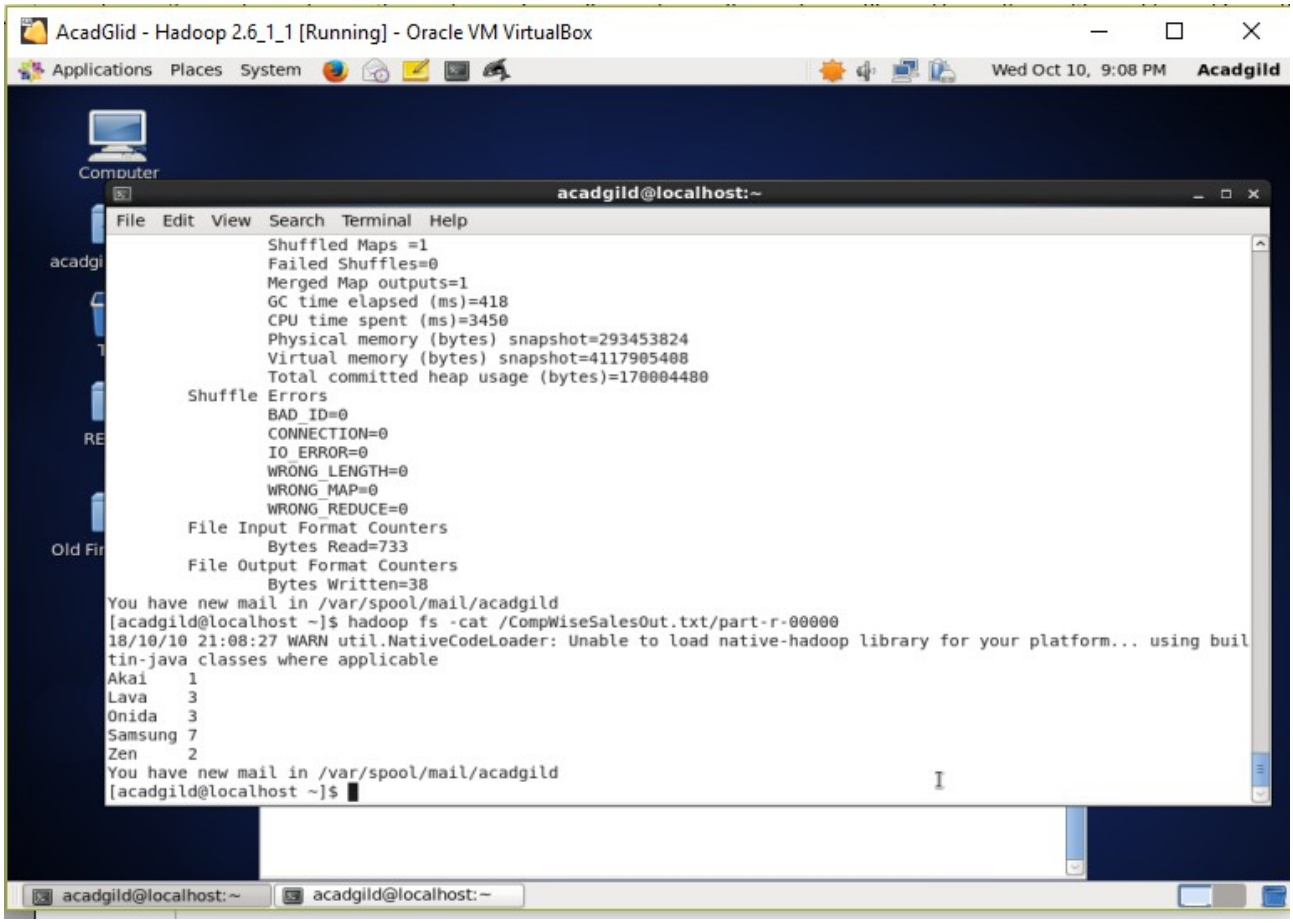
2. Command to run the Mapreduce program

   hadoop jar ProSaleCompWise.JAR ProcessSales.CompanyWiseSalesDriver input/television.txt CompWiseSalesOut.txt





Output of mapreduce program



Task 3

Write a map reduce program to calculate total units sold in each state for onida company

Mapper code

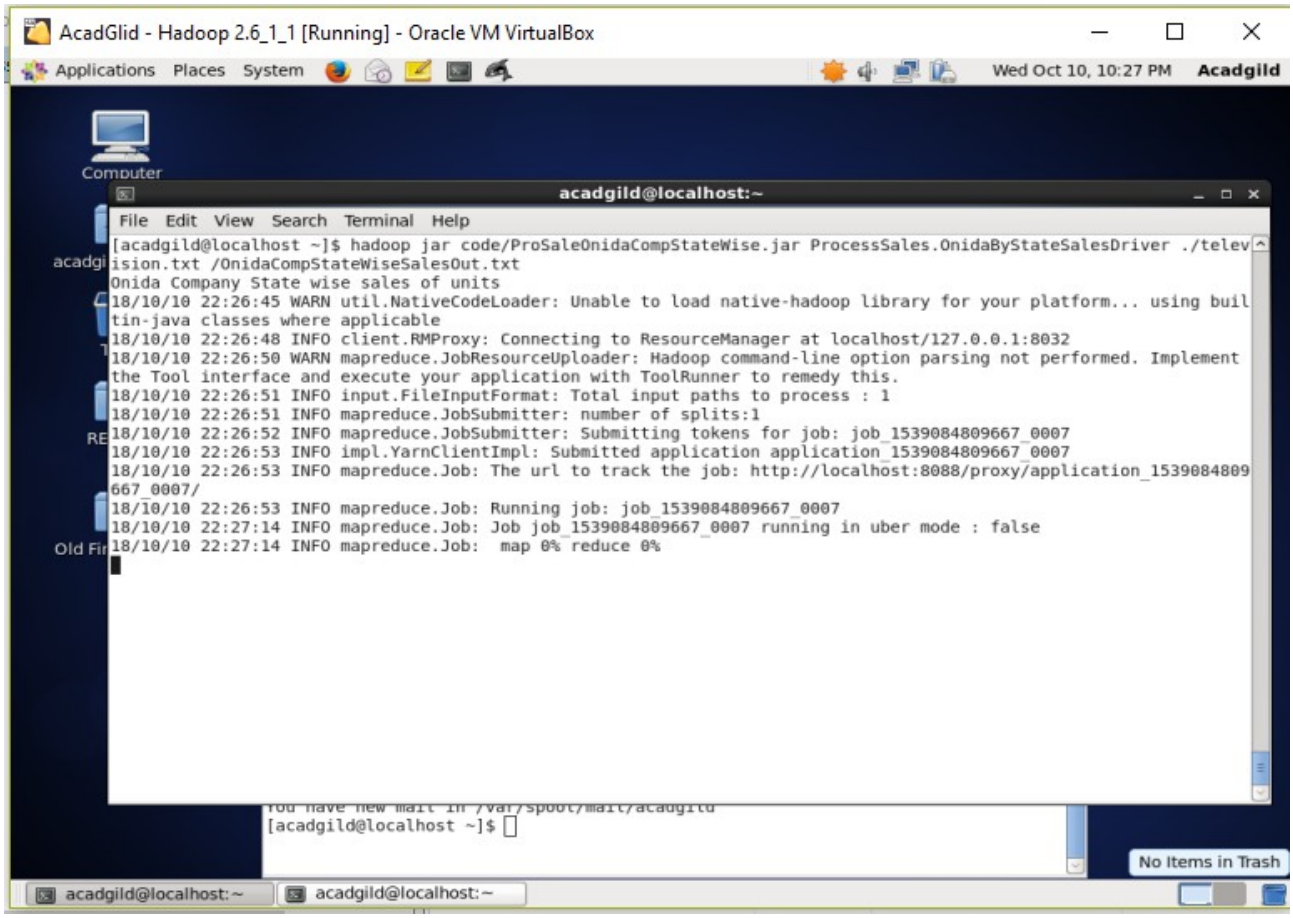
```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Mapper.Context;

public class OnidaByStateSalesMapper extends Mapper<LongWritable, Text, Text, IntWritable> {

    @Override
    protected void map(LongWritable in_key, Text in_value, Context context)
            throws IOException, InterruptedException {
        // TODO Auto-generated method stub
        String[] tokenizeValue = in_value.toString().split("\\|");
        if (tokenizeValue.length > 2) {

            if (!(tokenizeValue[0].contains("NA") || tokenizeValue[1].contains("NA"))) {
                if (tokenizeValue[0].contains("Onida")) {
                context.write(new Text(tokenizeValue[3]), new IntWritable(1));
                }
            }
        }
    }
```

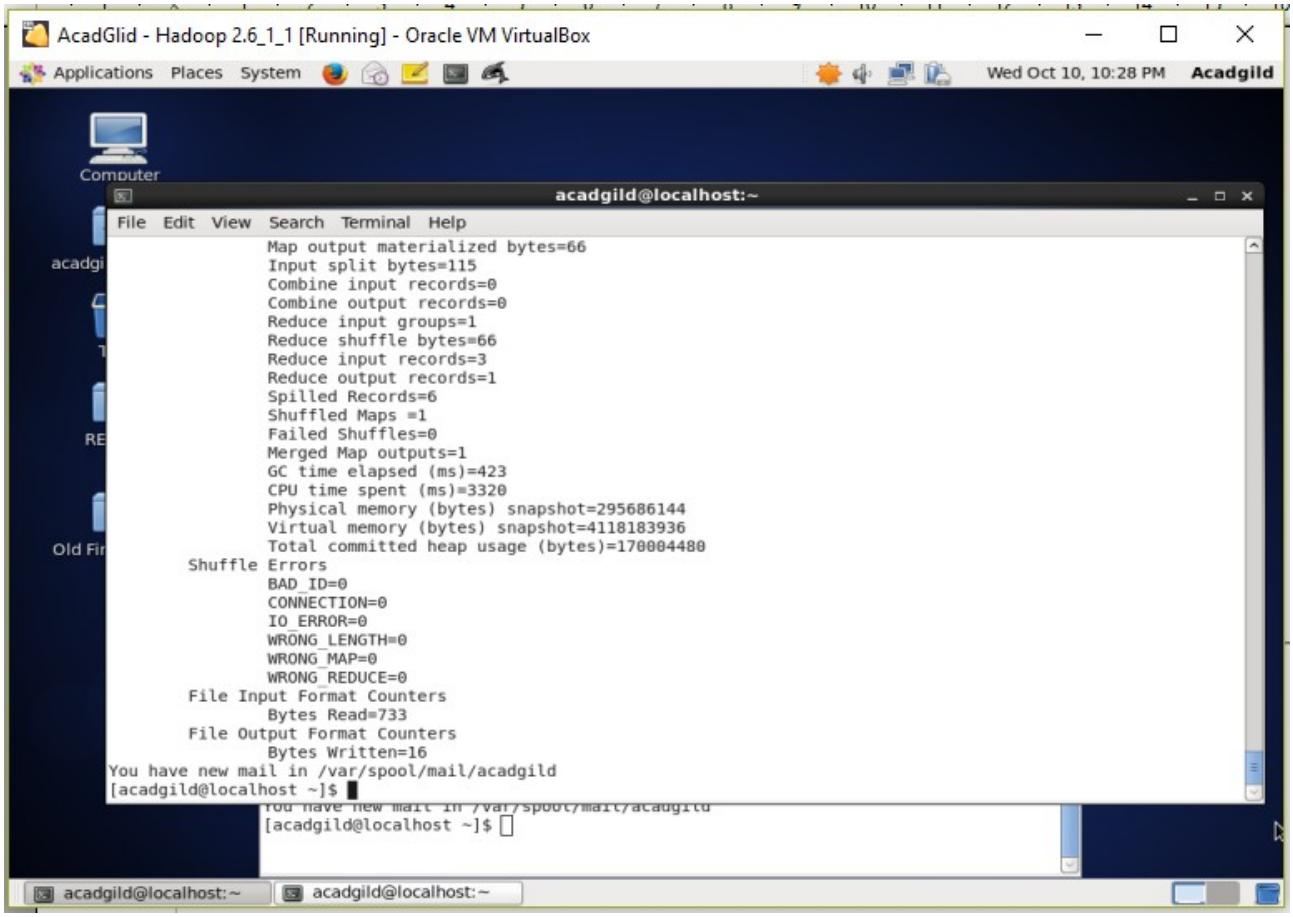
Reducer Code

```
package ProcessSales;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.Reducer.Context;

public class OnidaByStateSalesReducer extends Reducer<Text, IntWritable,Text, IntWritable> {

    @Override
    protected void reduce(Text in_key, Iterable<IntWritable> in_value,
            Context context) throws IOException, InterruptedException {
        // TODO Auto-generated method stub
        int sum_value = 0;
        for(IntWritable node_value:in_value){
            sum_value += Integer.parseInt(node_value.toString());
        }
        context.write(in_key, new IntWritable(sum_value));
    }
}
```

1. Copied ProSaleCompWise.jar JAR file from windows machine to unix machine using WinSCP
2. Command to run the Mapreduce program

   hadoop jar ProSaleOnidaCompStateWise.JAR ProcessSales.OnidaByStateSalesDriver input/television.txt OnidaCompStateWiseSalesOut.txt


AcadGlid - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Wed Oct 10, 10:27 PM Acadgild
Computer
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar code/ProSaleOnidaCompStateWise.jar ProcessSales.OnidaByStateSalesDriver ./television.txt /OnidaCompStateWiseSalesOut.txt
Onida Company State wise sales of units
18/10/10 22:26:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/10/10 22:26:48 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/10 22:26:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/10/10 22:26:51 INFO input.FileInputFormat: Total input paths to process : 1
18/10/10 22:26:51 INFO mapreduce.JobSubmitter: number of splits:1
18/10/10 22:26:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1539084809667_0007
18/10/10 22:26:53 INFO impl.YarnClientImpl: Submitted application application_1539084809667_0007
18/10/10 22:26:53 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1539084809667_0007/
18/10/10 22:26:53 INFO mapreduce.Job: Running job: job_1539084809667_0007
18/10/10 22:27:14 INFO mapreduce.Job: Job job_1539084809667_0007 running in uber mode : false
18/10/10 22:27:14 INFO mapreduce.Job: map 0% reduce 0%
[acadgild@localhost ~]$
No Items in Trash
acadgild@localhost:~
acadgild@localhost:~


AcadGlid - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Wed Oct 10, 10:28 PM Acadgild
Computer
acadgild@localhost:~
File Edit View Search Terminal Help
Map output materialized bytes=66
Input split bytes=115
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=66
Reduce input records=3
Reduce output records=1
Spilled Records=6
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=423
CPU time spent (ms)=3320
Physical memory (bytes) snapshot=295686144
Virtual memory (bytes) snapshot=4118183936
Total committed heap usage (bytes)=170004480
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=733
File Output Format Counters
Bytes Written=16
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
[acadgild@localhost ~]$
acadgild@localhost:~
acadgild@localhost:~

Output

Records which are invalid “NA” value in product name have been filtered out.. hence kerala state not appearing