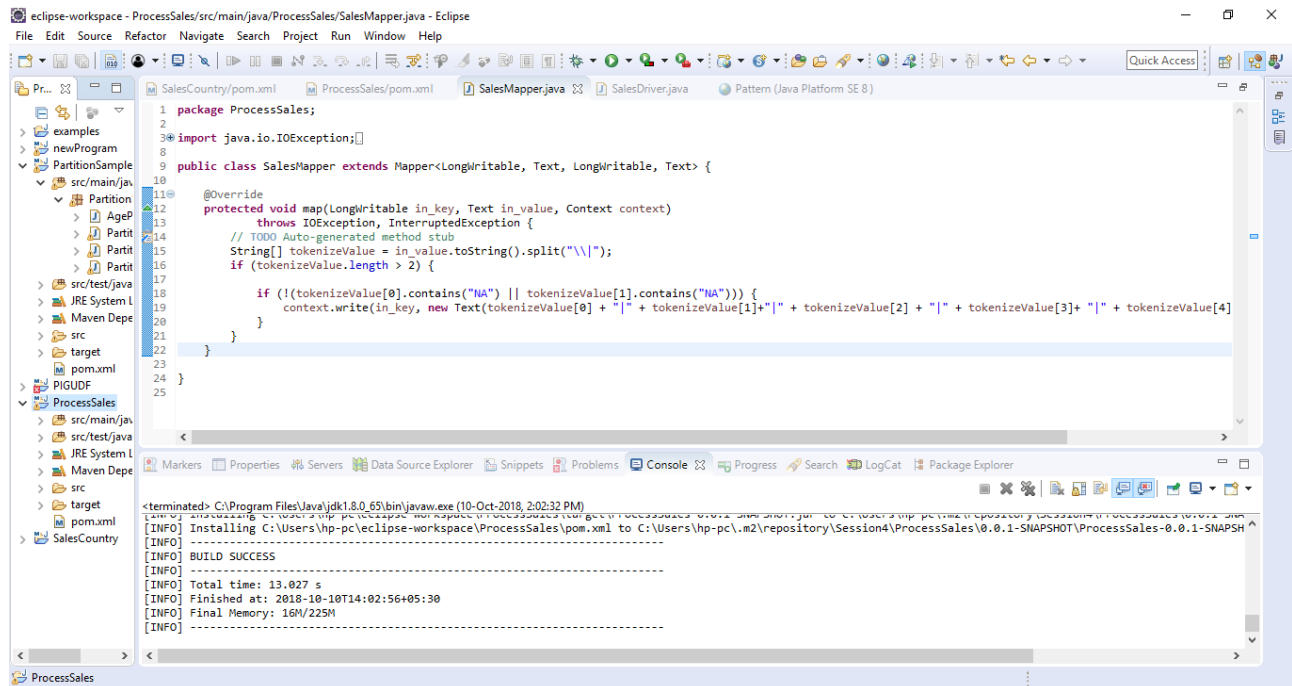


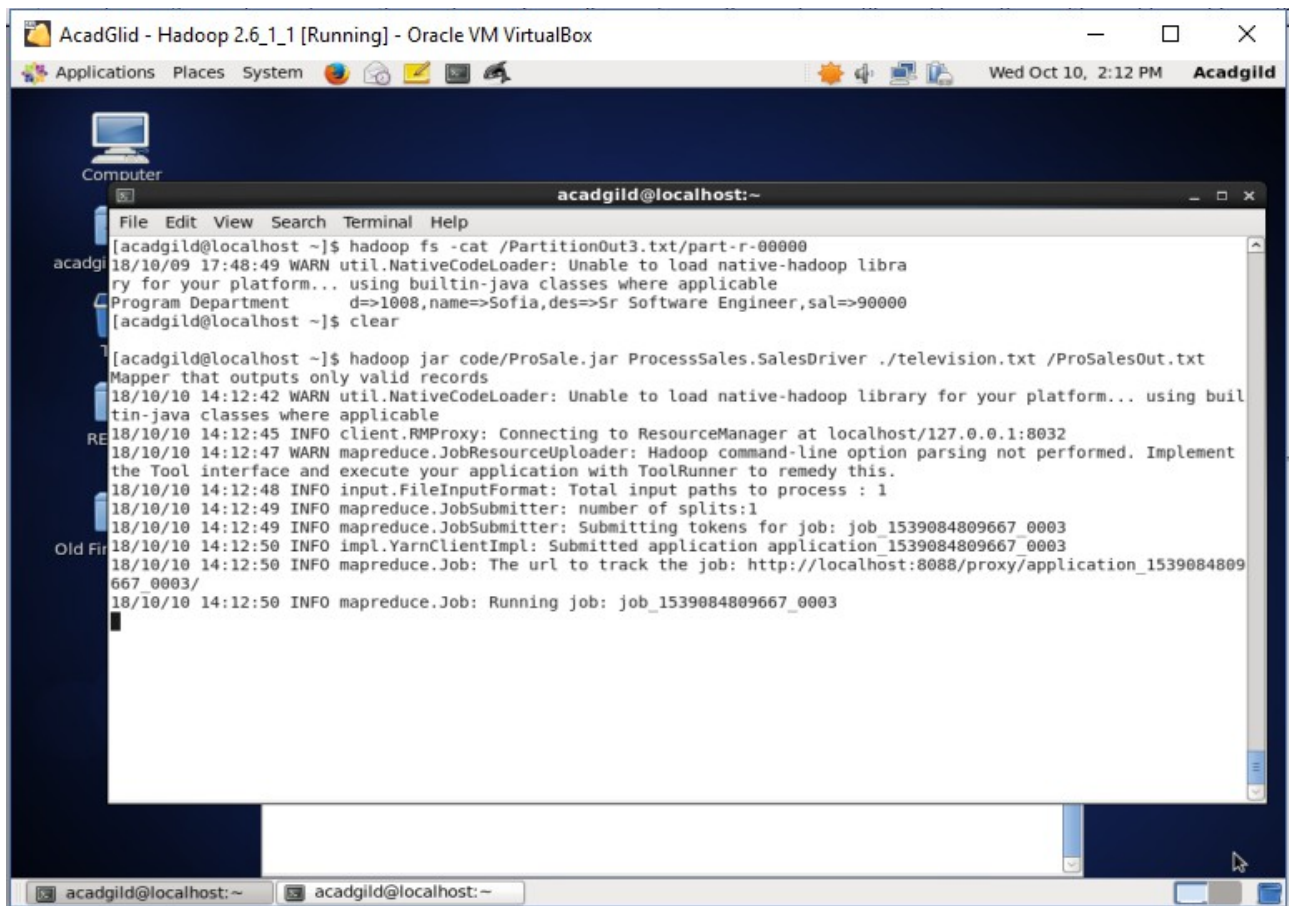
## Task 1

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



AcadGild - Hadoop 2.6.1\_1 [Running] - Oracle VM VirtualBox

Applications Places System Wed Oct 10, 2:12 PM Acadgild

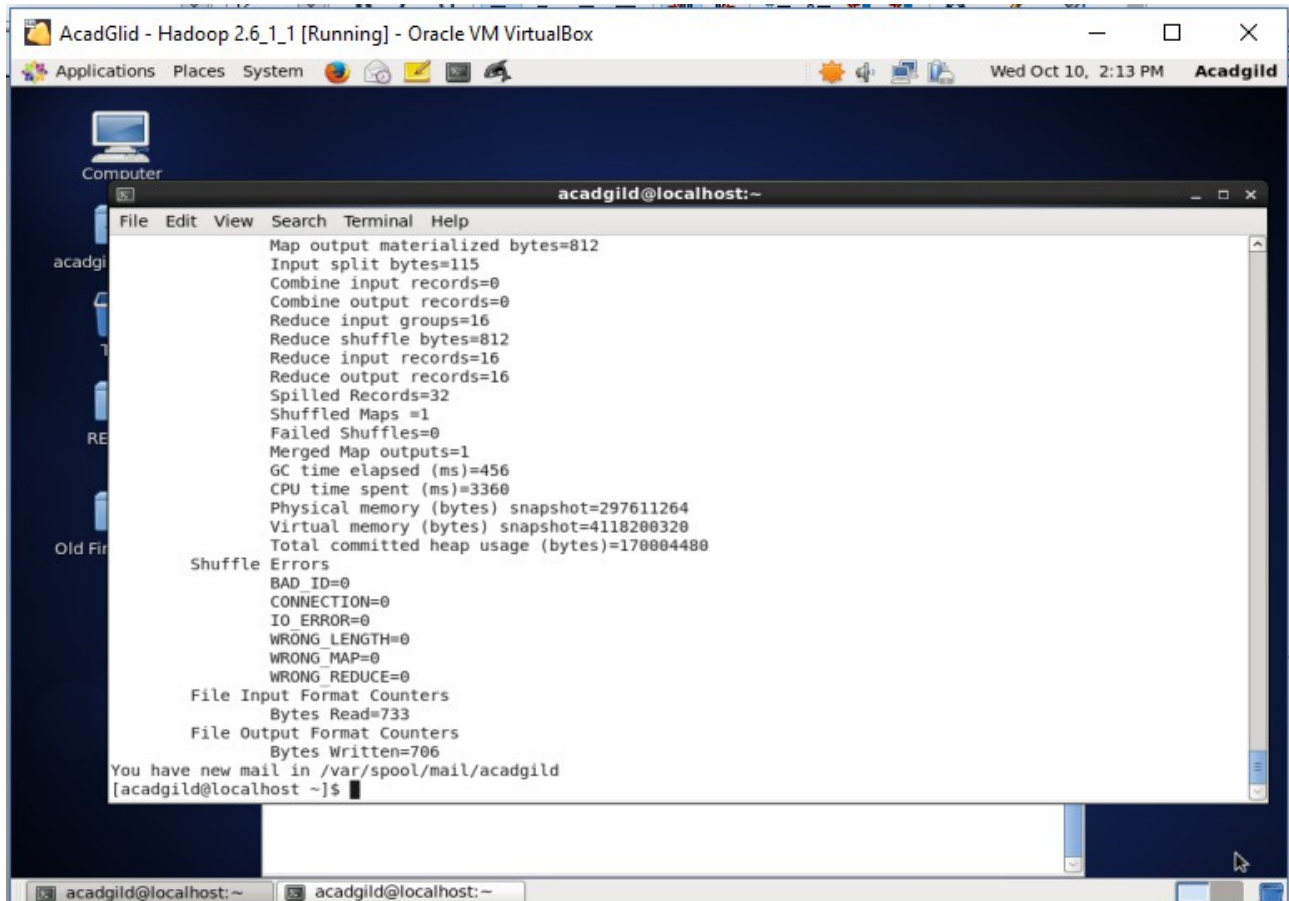
Computer

acadgild@localhost:~

```
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop fs -cat /PartitionOut3.txt/part-r-00000
18/10/09 17:48:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Program Department d=>1008,name=>Sofia,des=>Sr Software Engineer,sal=>90000
[acadgild@localhost ~]$ clear

[acadgild@localhost ~]$ hadoop jar code/ProSale.jar ProcessSales.SalesDriver ./television.txt /ProSalesOut.txt
Mapper that outputs only valid records
18/10/10 14:12:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/10/10 14:12:45 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/10 14:12:47 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 14:12:48 INFO input.FileInputFormat: Total input paths to process : 1
18/10/10 14:12:49 INFO mapreduce.JobSubmitter: number of splits:1
18/10/10 14:12:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1539084809667_0003
18/10/10 14:12:50 INFO impl.YarnClientImpl: Submitted application application_1539084809667_0003
18/10/10 14:12:50 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1539084809667_0003/
18/10/10 14:12:50 INFO mapreduce.Job: Running job: job_1539084809667_0003
```

On successful run



AcadGild - Hadoop 2.6.1\_1 [Running] - Oracle VM VirtualBox

Applications Places System Wed Oct 10, 2:13 PM Acadgild

Computer

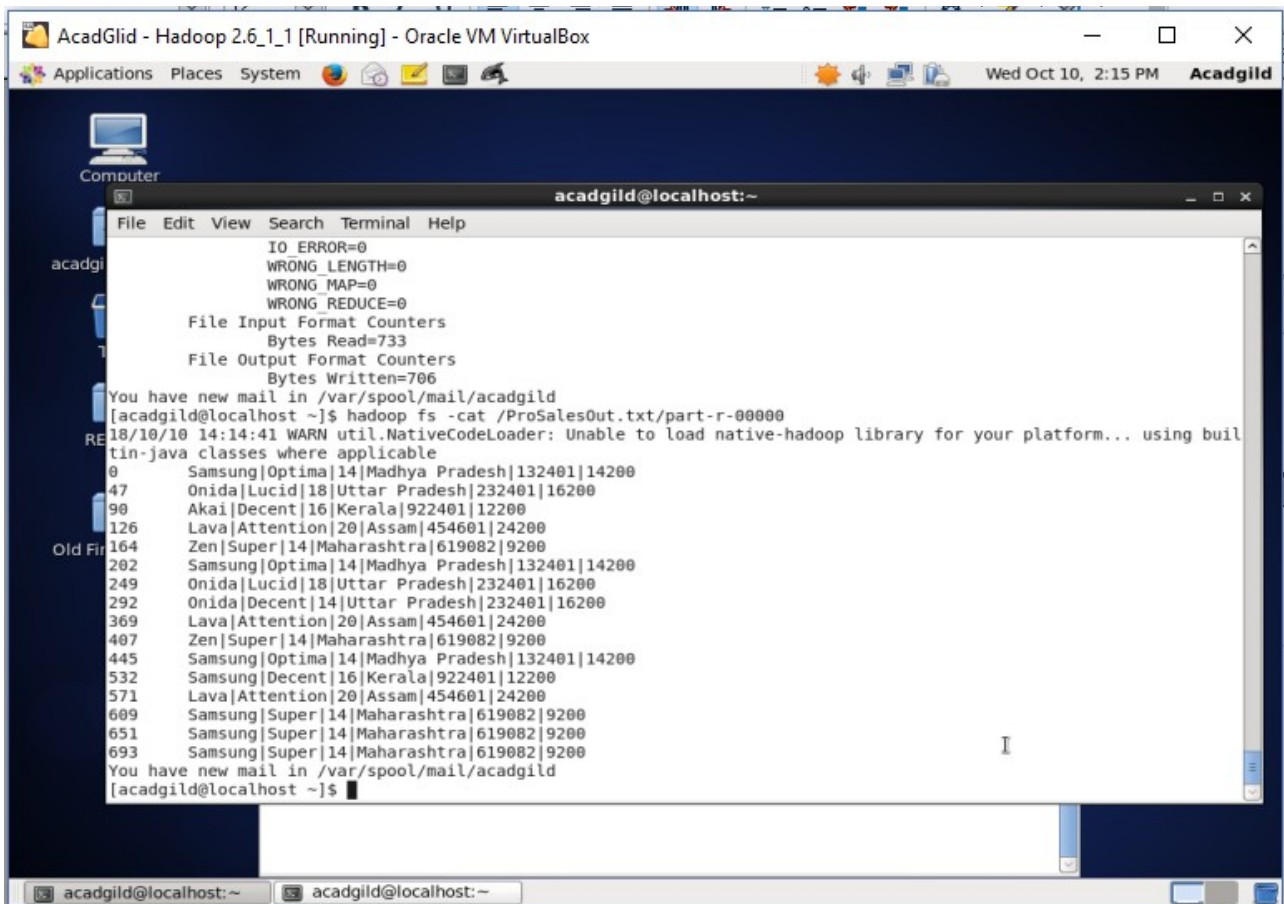
acadgild@localhost:~

```
File Edit View Search Terminal Help
Map output materialized bytes=812
Input split bytes=115
Combine input records=0
Combine output records=0
Reduce input groups=16
Reduce shuffle bytes=812
Reduce input records=16
Reduce output records=16
Spilled Records=32
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=456
CPU time spent (ms)=3360
Physical memory (bytes) snapshot=297611264
Virtual memory (bytes) snapshot=4118200320
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=706
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

## Checking the output



```
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=706
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /ProSalesOut.txt/part-r-00000
18/10/10 14:14:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built-in-java classes where applicable
0   Samsung|Optima|14|Madhya Pradesh|132401|14200
47  Onida|Lucid|18|Uttar Pradesh|232401|16200
90  Akai|Decent|16|Kerala|922401|12200
126 Lava|Attention|20|Assam|454601|24200
164 Zen|Super|14|Maharashtra|619082|9200
202 Samsung|Optima|14|Madhya Pradesh|132401|14200
249 Onida|Lucid|18|Uttar Pradesh|232401|16200
292 Onida|Decent|14|Uttar Pradesh|232401|16200
369 Lava|Attention|20|Assam|454601|24200
407 Zen|Super|14|Maharashtra|619082|9200
445 Samsung|Optima|14|Madhya Pradesh|132401|14200
532 Samsung|Decent|16|Kerala|922401|12200
571 Lava|Attention|20|Assam|454601|24200
609 Samsung|Super|14|Maharashtra|619082|9200
651 Samsung|Super|14|Maharashtra|619082|9200
693 Samsung|Super|14|Maharashtra|619082|9200
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

## 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;

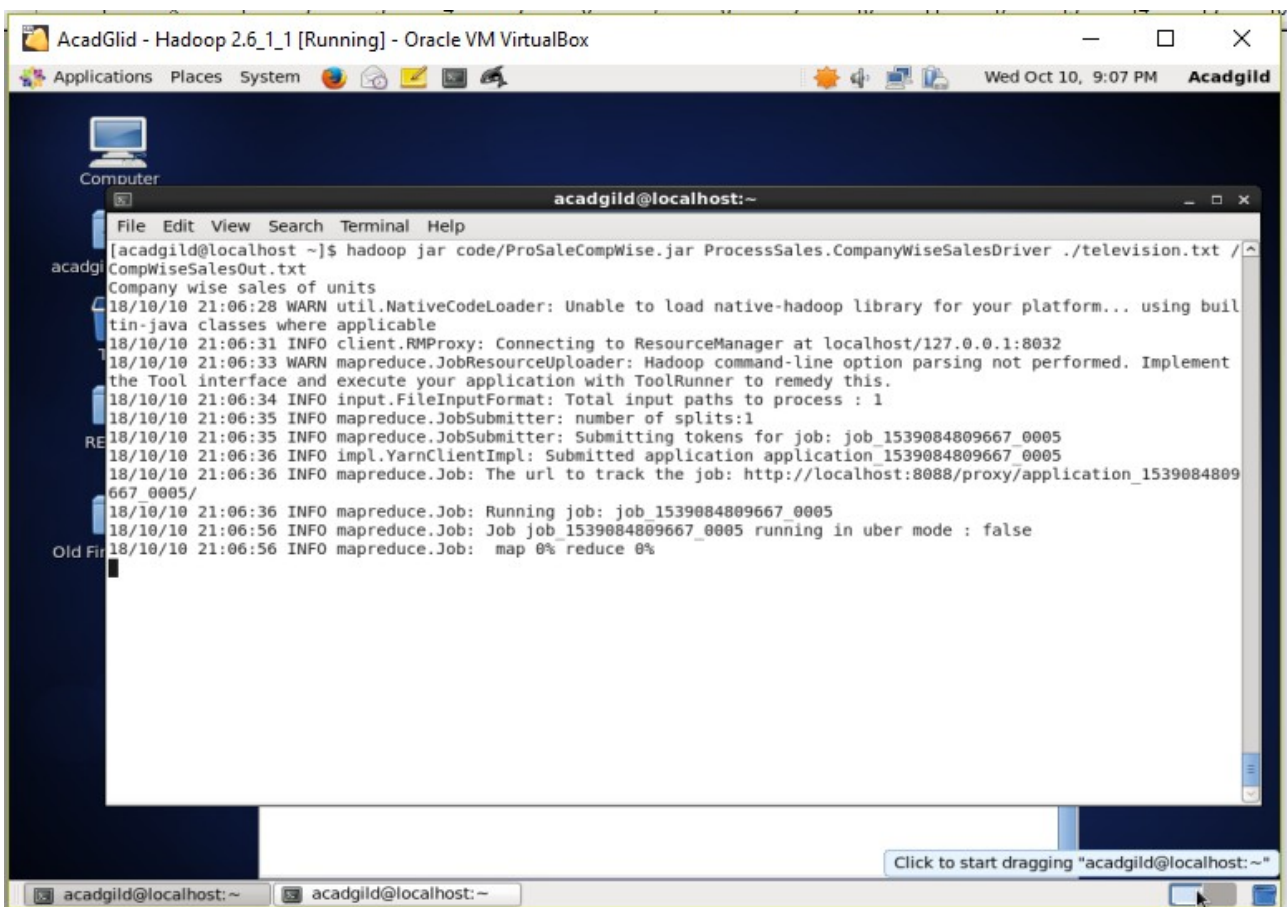
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.Reducer.Context;

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





```
AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@localhost:~
File Edit View Search Terminal Help
Map output materialized bytes=204
Input split bytes=115
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=204
Reduce input records=16
Reduce output records=5
Spilled Records=32
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=418
CPU time spent (ms)=3450
Physical memory (bytes) snapshot=293453824
Virtual memory (bytes) snapshot=4117905408
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=38
You have new mail in /var/spool/mail/acadmild
[acadgild@localhost ~]$
```

Output of mapreduce program

```
AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@localhost:~
File Edit View Search Terminal Help
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=418
CPU time spent (ms)=3450
Physical memory (bytes) snapshot=293453824
Virtual memory (bytes) snapshot=4117905408
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=38
You have new mail in /var/spool/mail/acadmild
[acadgild@localhost ~]$ hadoop fs -cat /CompWiseSalesOut.txt/part-r-00000
18/10/10 21:08:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built-in java classes where applicable
Akai 1
Lava 3
Onida 3
Samsung 7
Zen 2
You have new mail in /var/spool/mail/acadmild
[acadgild@localhost ~]$
```

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
```

AcadGild - Hadoop 2.6\_1\_1 [Running] - Oracle VM VirtualBox

Applications Places System Wed Oct 10, 10:27 PM Acadgild

Computer

acadmild@localhost:~

```
File Edit View Search Terminal Help
[acadmild@localhost ~]$ hadoop jar code/ProSaleOnidaCompStateWise.jar ProcessSales.OnidaByStateSalesDriver ./telev
ision.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 buil
tin-java classes where applicable
18/10/10 22:26:48 INFO client.RMPProxy: 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 1539084809667_0007
18/10/10 22:26:53 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1539084809
667_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%

You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
```

acadmild@localhost:~ acadmild@localhost:~

No Items in Trash

AcadGild - Hadoop 2.6\_1\_1 [Running] - Oracle VM VirtualBox

Applications Places System Wed Oct 10, 10:28 PM Acadgild

Computer

acadmild@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

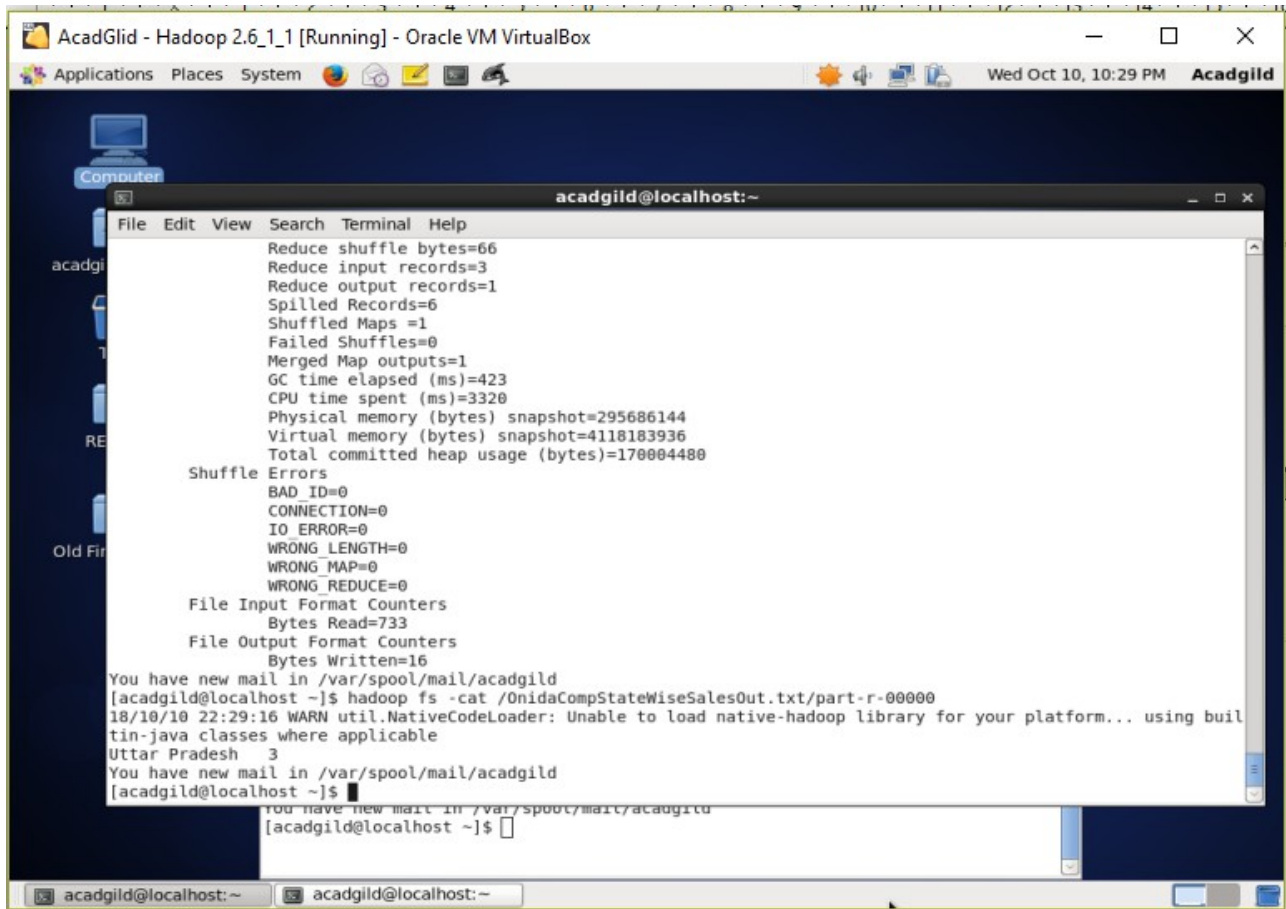
File Input Format Counters
Bytes Read=733
File Output Format Counters
Bytes Written=16

You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
```

acadmild@localhost:~ acadmild@localhost:~

## Output

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



The screenshot shows a terminal window titled 'acadgild@localhost:~' within an Oracle VM VirtualBox environment. The terminal displays the output of a Hadoop MapReduce job. The output includes various statistics such as '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', and 'Total committed heap usage (bytes)=170004480'. It also shows 'Shuffle Errors' with values for 'BAD\_ID=0', 'CONNECTION=0', 'IO\_ERROR=0', 'WRONG\_LENGTH=0', 'WRONG\_MAP=0', and 'WRONG\_REDUCE=0'. The 'File Input Format Counters' show 'Bytes Read=733' and the 'File Output Format Counters' show 'Bytes Written=16'. The terminal also displays a warning message: '18/10/10 22:29:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable'. The terminal prompt is '[acadgild@localhost ~]\$'.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
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 ~]$ hadoop fs -cat /OnidaCompStateWiseSalesOut.txt/part-r-000000  
18/10/10 22:29:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built-in-java classes where applicable  
Uttar Pradesh 3  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```