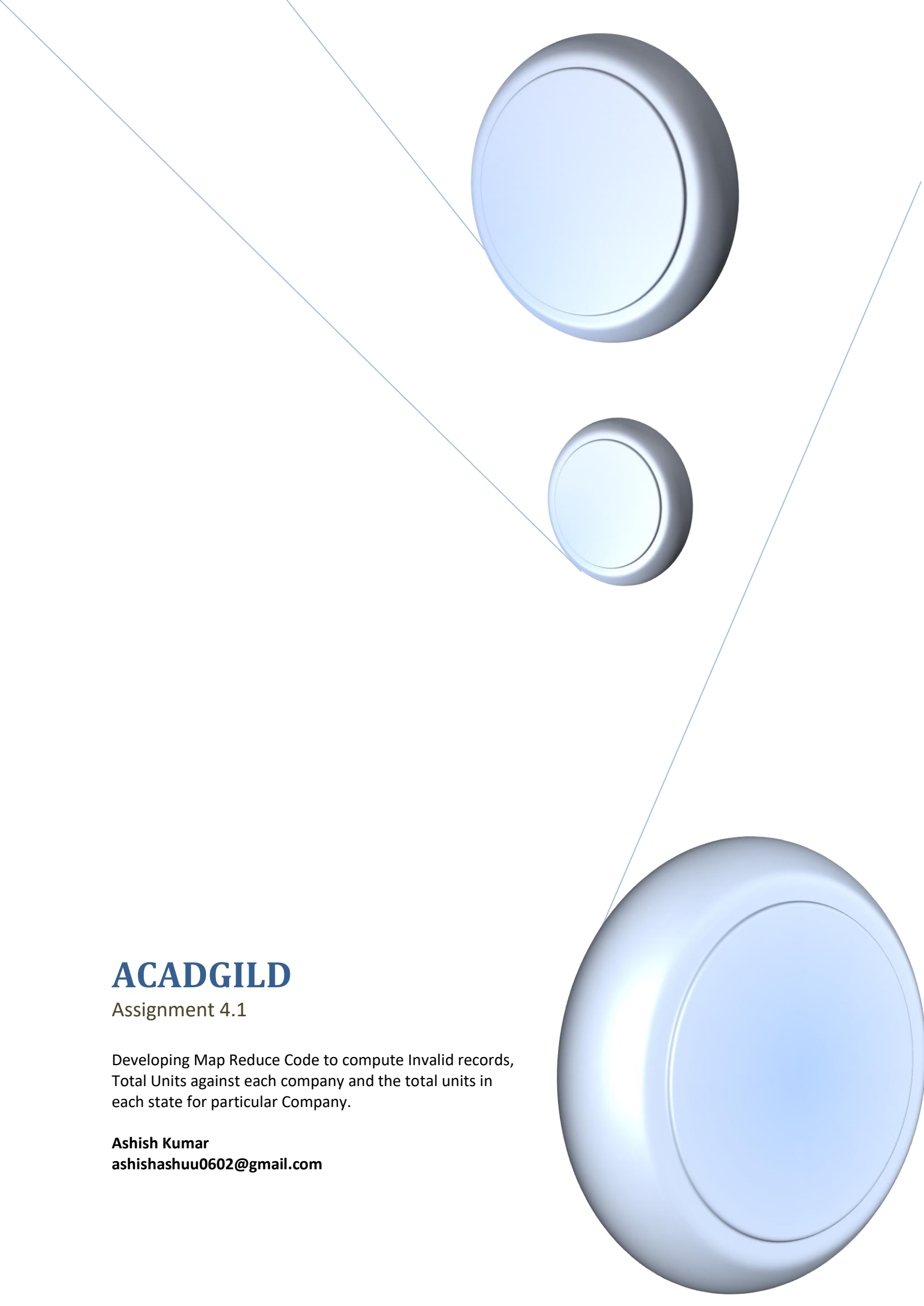


ACADGILD

Assignment 4.1

Developing Map Reduce Code to compute Invalid records, Total Units against each company and the total units in each state for particular Company.

Ashish Kumar
ashishashuu0602@gmail.com

Task 1: Map reduce program to filter out the invalid records. Map only job.

Solution:

Expected Output/Result:
2 Records with NA should come as output.

Below is the code used to filter the Invalid Records, records having "NA". We have used this code as executable Jar file. We have placed this jar file in local filesystem and pointed to this Jar file to run the provided input file(television.txt).

Below is the Code Snippet –

Mapper Code (**InvalidRecords.java**)

```java
package com.acadgild.task1;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.NullWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;


    public class InvalidRecords extends Mapper<LongWritable, Text,NullWritable,Text>{

        NullWritable key;
        Text record;

    public void setup(Context context)
        {
            record = new Text();
        }

    public void map(LongWritable key1, Text value, Context context) throws IOException, InterruptedException{
        String totline = value.toString();
        record = new Text(totline);

        if(totline.contains("NA"))

            context.write(key,record);
        }
    }
```

Driver Class (**InvalidrecordsMain.java**)

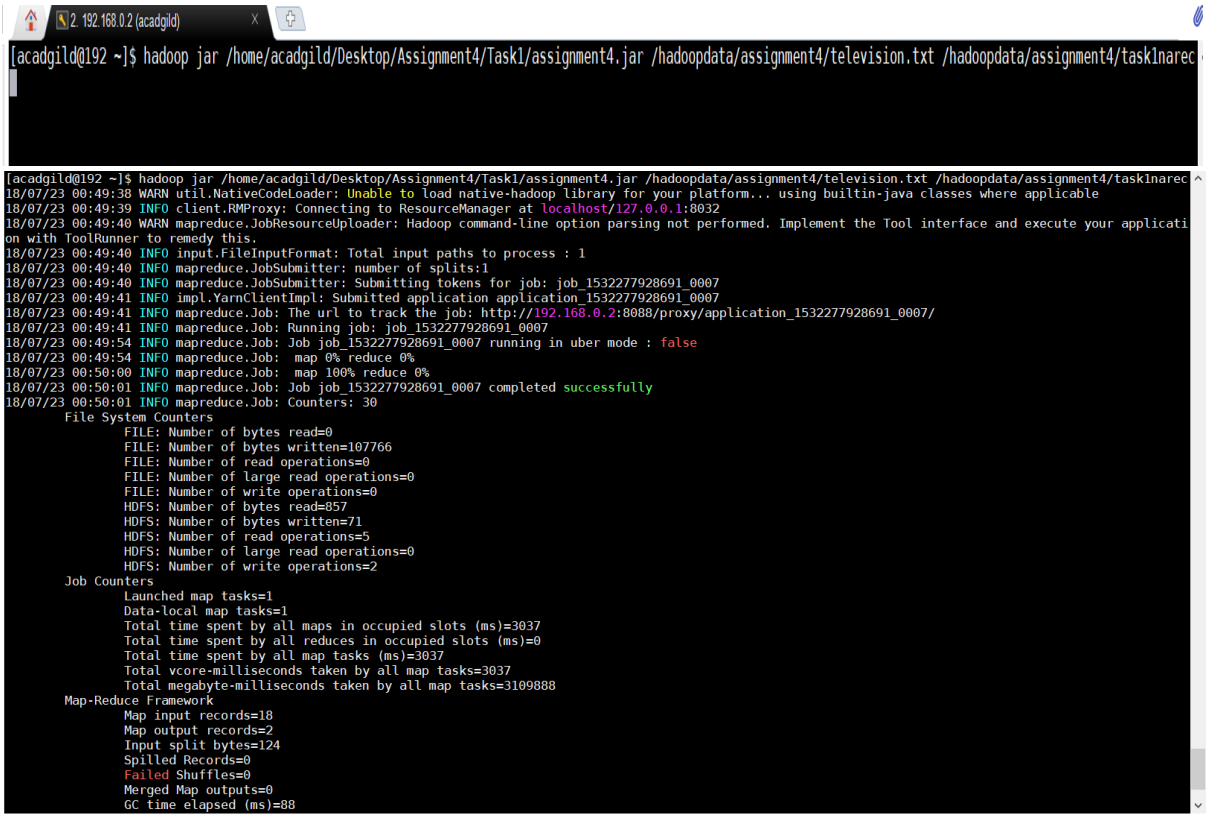
```java
package com.acadgild.task1;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.NullWritable;
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.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class InvalidrecordsMain {

    public static void main(String[] args) throws Exception, Exception, InterruptedException {
        // TODO Auto-generated method stub
                if (args.length != 2) {
                  System.err.println("Usage: Provide Input file and output path <input path> <output path>");
                  System.exit(-1);
                }

                //Job Related Configurations
                Configuration conf = new Configuration();
                @SuppressWarnings("deprecation")
                Job job = new Job(conf, "Assignment 4 - InvalidRecords");
                job.setJarByClass(InvalidrecordsMain.class);


                //Provide paths to pick the input file for the job
                FileInputFormat.setInputPaths(job, new Path(args[0]));
```

```java


                //Provide paths to pick the output file for the job, and delete it if already present
                Path outputPath = new Path(args[1]);
                FileOutputFormat.setOutputPath(job, outputPath);
                outputPath.getFileSystem(conf).delete(outputPath, true);


                //To set the mapper and reducer of this job
                job.setMapperClass(InvalidRecords.class);
                // Specify the number of reducer to 0
                job.setNumReduceTasks(0);


                //set the input and output format class
                job.setInputFormatClass(TextInputFormat.class);
                job.setOutputFormatClass(TextOutputFormat.class);

                //set up the output key and value classes
                job.setMapOutputKeyClass(NullWritable.class);
                job.setMapOutputValueClass(Text.class);

                job.setOutputKeyClass(NullWritable.class);
                job.setOutputValueClass(Text.class);

                //execute the job
                System.exit(job.waitForCompletion(true) ? 0 : 1);
            }
        }
```

Command to use Jar and respective input and output –

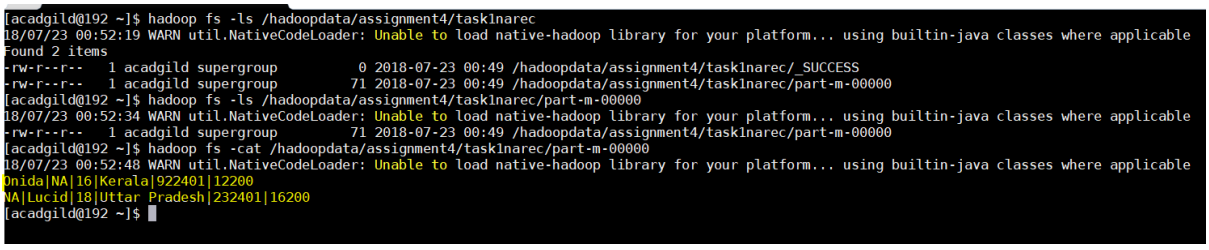
Task 1 –Input: **television.txt**

Output Directory – **task1narec**

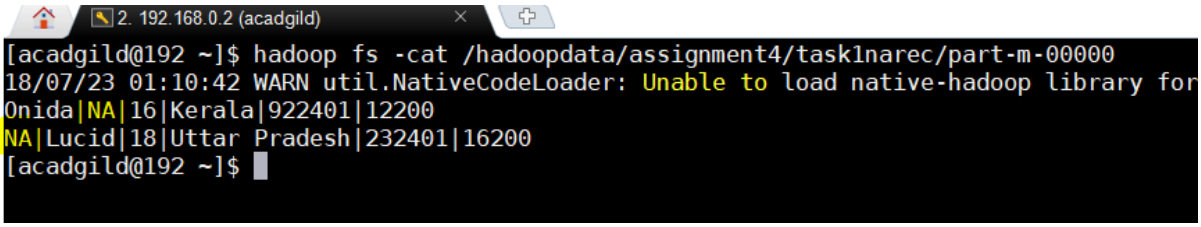


Actual Result:

The output shows only 2 Records which is having “NA” record.



Output –



Task 2: Map reduce program to calculate total units sold for each company.

Solution:

Below is the Code Snippet –

Mapper Code (**TotaUnitCount.java**)

```java
package com.acadgild.task2;

import java.io.IOException;

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

public class TotaUnitCount extends Mapper<LongWritable, Text, Text, IntWritable> {
        IntWritable unitcou;
        Text compname;

    public void setup(Context context)
        {
            unitcou = new IntWritable(1);
            compname = new Text();
        }
    public void map(LongWritable key, Text compvalue, Context context) throws IOException, InterruptedExceptio
        {
            String[] totaline = compvalue.toString().split("[|]");

        //below condition to check NA(invalid) record, it should not be included in total units

            if(!((totaline[0].equals("NA")) || (totaline[1].equals("NA"))))
            {
                compname.set((totaline[0]));
                context.write(compname, unitcou);
            }


        }
```

Reducer Code (**TotaUnitCountReducer.java**)

```java
package com.acadgild.task2;

import java.io.IOException;

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

public class TotaUnitCountReducer extends Reducer<Text, IntWritable, Text, IntWritable>
    {
        public void reduce(Text compname, Iterable<IntWritable> counter,
                Context context)
                        throws IOException, InterruptedException {
                        System.out.println("From The Reducer=>"+compname) ;

                        int sum = 0;
                        for (IntWritable unitcou : counter)
                         {
                            sum+=unitcou.get();
                         }

                        context.write(compname, new IntWritable(sum));
                    }
    }

```

Driver Code (**TotalUnitsCouMain.java**)

```java
package com.acadgild.task2;

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.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;


public class TotalUnitsCouMain {

    public static void main(String[] args) throws Exception, IOException, InterruptedException {
        // TODO Auto-generated method stub
          if (args.length != 2) {
              System.err.println("Usage: Provide Input file path and output path <input path> <output path>");
              System.exit(-1);
           }

           //Job Related Configurations
           Configuration conf = new Configuration();
           @SuppressWarnings("deprecation")
           Job job = new Job(conf, "Assignment 4 Task 2- Total Units");
           System.out.println("Assignment 4 Task 2 - Total Units for each Commpany");
           job.setJarByClass(TotalUnitsCouMain.class);


           //Provide paths to pick the input file for the job
           FileInputFormat.setInputPaths(job, new Path(args[0]));


           //Provide paths to pick the output file for the job, and delete it if already present
           Path outputPath = new Path(args[1]);
           FileOutputFormat.setOutputPath(job, outputPath);
           outputPath.getFileSystem(conf).delete(outputPath, true);


           //To set the mapper and reducer of this job
```

```java
              job.setMapperClass(TotaUnitCount.class);
              job.setReducerClass(TotaUnitCountReducer.class);
              // Specify the number of reducer to 2
              job.setNumReduceTasks(2);


              //set the input and output format class
              job.setInputFormatClass(TextInputFormat.class);
              job.setOutputFormatClass(TextOutputFormat.class);

              //set up the output key and value classes
              job.setMapOutputKeyClass(Text.class);
              job.setMapOutputValueClass(IntWritable.class);

              job.setOutputKeyClass(Text.class);
              job.setOutputValueClass(IntWritable.class);

              //execute the job
              System.exit(job.waitForCompletion(true) ? 0 : 1);
           }
        }

```

Command to run the jar file as below –

```
[acadgild@192 ~]$ hadoop jar /home/acadgild/Desktop/Assignment4/Task2/assign4task2.jar  /hadoopdata/assignment4/television.txt /hadoopdata/assignment4/totaunit
rec
18/07/23 02:47:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Assignment 4 Task 2 - Total Units for each Commpany
18/07/23 02:47:13 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/23 02:47:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your applicati
on with ToolRunner to remedy this.
18/07/23 02:47:14 INFO input.FileInputFormat: Total input paths to process : 1
18/07/23 02:47:14 INFO mapreduce.JobSubmitter: number of splits:1
18/07/23 02:47:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532277928691_0011
18/07/23 02:47:15 INFO impl.YarnClientImpl: Submitted application application_1532277928691_0011
18/07/23 02:47:15 INFO mapreduce.Job: The url to track the job: http://192.168.0.2:8088/proxy/application_1532277928691_0011/
18/07/23 02:47:15 INFO mapreduce.Job: Running job: job_1532277928691_0011
18/07/23 02:47:22 INFO mapreduce.Job: Job job_1532277928691_0011 running in uber mode : false
18/07/23 02:47:22 INFO mapreduce.Job:  map 0% reduce 0%
18/07/23 02:47:27 INFO mapreduce.Job:  map 100% reduce 0%
18/07/23 02:47:34 INFO mapreduce.Job:  map 100% reduce 50%
18/07/23 02:47:35 INFO mapreduce.Job:  map 100% reduce 100%
18/07/23 02:47:36 INFO mapreduce.Job: Job job_1532277928691_0011 completed successfully
18/07/23 02:47:36 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=210
                FILE: Number of bytes written=324652
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=857
                HDFS: Number of bytes written=38
                HDFS: Number of read operations=9
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=2
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=2841
                Total time spent by all reduces in occupied slots (ms)=9688
                Total time spent by all map tasks (ms)=2841
                Total time spent by all reduce tasks (ms)=9688
                Total vcore-milliseconds taken by all map tasks=2841
                Total vcore-milliseconds taken by all reduce tasks=9688
                Total megabyte-milliseconds taken by all map tasks=2909184
```

The Output of the above Code is as below –

Input File – **television.txt**

Output Directory – **totaunitrec**

```
[acadgild@192 ~]$ hadoop fs -cat /hadoopdata/assignment4/totaunitrec/part-r-00000
18/07/23 02:51:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y
Onida   3
Zen     2
[acadgild@192 ~]$ hadoop fs -cat /hadoopdata/assignment4/totaunitrec/part-r-00001
18/07/23 02:51:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y
Akai    1
Lava    3
Samsung 7
[acadgild@192 ~]$
```

```
18/07/23 02:55:11 WARN util.NativeCodeLoader: Unable to load nati
Onida   3
Zen     2
Akai    1
Lava    3
Samsung 7
[acadgild@192 ~]$
```

Task 3: Map reduce program to calculate total units in each state for ONIDA Company.

Solution:

Below is the Code Snippet –

Mapper Code (**OnidaTotUnits.java**)

```
1  package com.acadgild.assign4.task3;
2
3  import java.io.IOException;
9
10 public class OnidaTotUnits extends Mapper<LongWritable, Text,Text, IntWritable> {
11
12                 Text statename = new Text();
13                 IntWritable onidacou = new IntWritable(1);
14
15         public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedExceptio
16
17                 System.out.println("This is Mapper" + key.toString());
18                 String[] compname = value.toString().split("[|]");
19
20                 //below is the condition check to count units only for Onida Company.
21
22                 if(compname[0].equals("Onida"))
23                         {
24
25                             String statename1 = new String (compname[3]);
26                             String companynm = new String (compname[0]);
27                             String space = new String ("      ");
28                             String stco = companynm.concat(space).concat(statename1);
29                             Text stcomp = new Text(stco);
30
31                             statename.set(stcomp);
32                             context.write(statename, onidacou);
33                         }
34 }
35 }
```

Reducer Code (**OnidaTotUnitsReducer.java**)

```java
package com.acadgild.assign4.task3;

import java.io.IOException;

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

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

    public void reduce(Text statename, Iterable<IntWritable> counter,
              Context context)
                      throws IOException, InterruptedException {
                      System.out.println("From The Reducer=>"+ statename) ;

                      int sum = 0;
                      for (IntWritable onidacou : counter)
                       {
                          sum+=onidacou.get();
                       }

                      context.write(statename, new IntWritable(sum));
                  }
}
```

Driver Class (**OnidaTotUnitsMain.java**)

```java
package com.acadgild.assign4.task3;

import java.io.IOException;


public class OnidaTotUnitsMain {

    public static void main(String[] args) throws IOException, ClassNotFoundException, InterruptedException {
        // TODO Auto-generated method stub
          if (args.length != 2) {
              System.err.println("Usage: Provide Input file path and output path <input path> <output path>");
              System.exit(-1);
           }

          //Job Related Configurations
          Configuration conf = new Configuration();
          @SuppressWarnings("deprecation")
          Job job = new Job(conf, "Assignment 4 Task 3- Total Units only for ONIDA company in each state");
          System.out.println("Assignment 4 Task 3 - Total Units for each State only for ONIDA company");
          job.setJarByClass(OnidaTotUnitsMain.class);


          //Provide paths to pick the input file for the job
          FileInputFormat.setInputPaths(job, new Path(args[0]));


          //Provide paths to pick the output file for the job, and delete it if already present
          Path outputPath = new Path(args[1]);
          FileOutputFormat.setOutputPath(job, outputPath);
          outputPath.getFileSystem(conf).delete(outputPath, true);
```

```
42
43              //To set the mapper and reducer of this job
44              job.setMapperClass(OnidaTotUnits.class);
45              job.setReducerClass(OnidaTotUnitsReducer.class);
46              // Specify the number of reducer to 1
47              job.setNumReduceTasks(1);
48
49
50              //set the input and output format class
51              job.setInputFormatClass(TextInputFormat.class);
52              job.setOutputFormatClass(TextOutputFormat.class);
53
54              //set up the output key and value classes
55              job.setMapOutputKeyClass(Text.class);
56              job.setMapOutputValueClass(IntWritable.class);
57
58              job.setOutputKeyClass(Text.class);
59              job.setOutputValueClass(IntWritable.class);
60
61              //execute the job
62              System.exit(job.waitForCompletion(true) ? 0 : 1);
63            }
64
65      }
66
67
```

Command which runs the jar for expected output –

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/Assignment4/Task3/assign4task3.jar /hadoopdata/assignment4/television.txt /hadoopdata/assignment4/Oni
datask3
18/07/24 02:10:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Assignment 4 Task 3 - Total Units for each State only for ONIDA company
18/07/24 02:10:22 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/24 02:10:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your applicati
on with ToolRunner to remedy this.
18/07/24 02:10:23 INFO input.FileInputFormat: Total input paths to process : 1
18/07/24 02:10:23 INFO mapreduce.JobSubmitter: number of splits:1
18/07/24 02:10:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532370396751_0004
18/07/24 02:10:23 INFO impl.YarnClientImpl: Submitted application application_1532370396751_0004
18/07/24 02:10:23 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532370396751_0004/
18/07/24 02:10:23 INFO mapreduce.Job: Running job: job_1532370396751_0004
18/07/24 02:10:31 INFO mapreduce.Job: Job job_1532370396751_0004 running in uber mode : false
18/07/24 02:10:31 INFO mapreduce.Job:  map 0% reduce 0%
18/07/24 02:10:36 INFO mapreduce.Job:  map 100% reduce 0%
18/07/24 02:10:42 INFO mapreduce.Job:  map 100% reduce 100%
18/07/24 02:10:42 INFO mapreduce.Job: Job job_1532370396751_0004 completed successfully
18/07/24 02:10:42 INFO mapreduce.Job: Counters: 49
       File System Counters
              FILE: Number of bytes read=123
              FILE: Number of bytes written=216499
              FILE: Number of read operations=0
```

```
       File System Counters
              FILE: Number of bytes read=123
              FILE: Number of bytes written=216499
              FILE: Number of read operations=0
              FILE: Number of large read operations=0
              FILE: Number of write operations=0
              HDFS: Number of bytes read=857
              HDFS: Number of bytes written=47
              HDFS: Number of read operations=6
              HDFS: Number of large read operations=0
              HDFS: Number of write operations=2
       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)=2964
              Total time spent by all reduces in occupied slots (ms)=3010
              Total time spent by all map tasks (ms)=2964
              Total time spent by all reduce tasks (ms)=3010
              Total vcore-milliseconds taken by all map tasks=2964
              Total vcore-milliseconds taken by all reduce tasks=3010
              Total megabyte-milliseconds taken by all map tasks=3035136
              Total megabyte-milliseconds taken by all reduce tasks=3082240
```

The Output:

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/assignment4/Onidatask3
18/07/24 02:13:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bu
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-07-24 02:10 /hadoopdata/assignment4/Onidatask3/_SUCCESS
-rw-r--r--   1 acadgild supergroup         47 2018-07-24 02:10 /hadoopdata/assignment4/Onidatask3/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/assignment4/Onidatask3/part-r-00000
18/07/24 02:14:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bu
Onida      Kerala      1
Onida      Uttar Pradesh       3
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```