ASSIGNMENT MAPREDUCE

PROBLEM STATEMENT:

--> We have a dataset of sales of different TV sets across different locations.

--> Records look like:
Samsung|Optima|14|Madhya Pradesh|132401|14200

--> The fields are arranged like:
Company Name|Product Name|Size in inches|State|Pin Code|Price

--> There are some invalid records which contain 'NA' in either Company Name or Product Name.

TASK 1: Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

DRIVER CODE:

```
package acadgild_Mapreduce;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
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 driver {

      public static void main(String[] args) throws Exception {
            Configuration conf = new Configuration();
            conf.addResource(new
Path("/usr/lib/hadoop/etc/hadoop/core-site.xml"));
            conf.addResource(new
Path("/usr/lib/hadoop/etc/hadoop/core-site.xml"));

            FileSystem fs = FileSystem.get(conf);

            String inPath;
            String outPath;

            if(args.length != 2){

                  System.err.println("Usage: Seq File Example <input file>
<output dir>");
                  System.out.println("Using default file: seqfile");
```

```
            inPath = "/user/acadgild/television";
            outPath = "/user/acadgild/validtv";
        }else{

            inPath = args[0];
            outPath = args[1];
        }

        @SuppressWarnings("deprecation")
        Job job = new Job(conf, "television");
        job.setJarByClass(driver.class);
        job.setMapperClass(MyMapper.class);
        job.setNumReduceTasks(0);

        job.setMapOutputKeyClass(LongWritable.class);
        job.setMapOutputValueClass(Text.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        FileInputFormat.addInputPath(job, new Path(inPath));
        if (fs.exists(new Path(outPath))) {
            fs.delete(new Path(outPath), true);
        }
        FileOutputFormat.setOutputPath(job, new Path(outPath));
        System.exit(job.waitForCompletion(true)?0:1);

    }
}
```

MAPPER CODE:

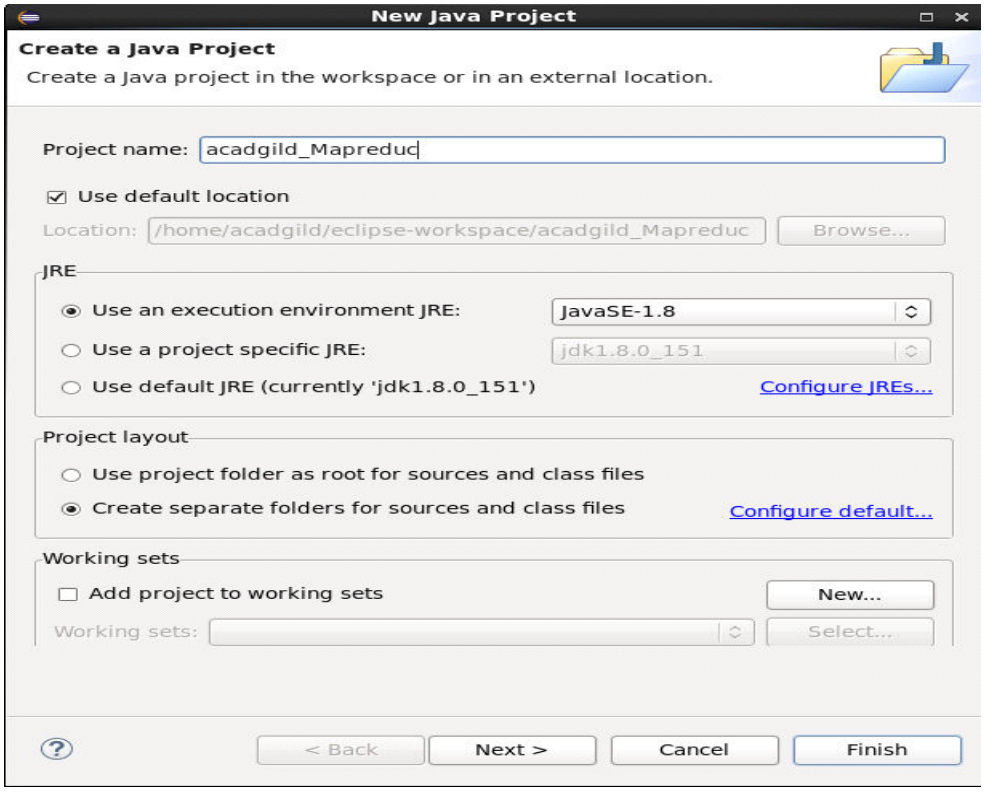
```
package acadgild_Mapreduce;

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

import java.io.IOException;

public class MyMapper extends Mapper<LongWritable, Text, Text, Text> {

    private Text records = new Text();
    private Text dummy = new Text();

    public void map(LongWritable key, Text value, Context context) throws
IOException, InterruptedException{

        System.out.println("Input key = "+key.toString() +" Input value
= " + value.toString());

        String[] token = value.toString().split("\\|");

        int flag = 0;
```

```
            System.out.println("Company "+ token[0]+"product "+token[1]);
            System.out.println("company "+token[0]+" product "+token[1]);
            if (token[0].equals("NA") || token[1].equals("NA")) {

                  flag = 1;
            }

            if (flag == 0) {

                  records.set(value);
                  dummy.set("");
                  context.write(dummy, records);
                  System.out.println(records.toString());
            }

      }

}
```

EXPLANATION: SO TO DO ABOVE TASKS FIRST DOWNLOADED THE INPUT FILE i.e OUR DATASET "television.txt" FILE AND THEN USING "put" COMMAND COPIED THE FILE INTO HDFS DIRECTORY FROM PATH "/home/acadgild/television.txt" INTO PATH "/user/acadgild/television.txt"

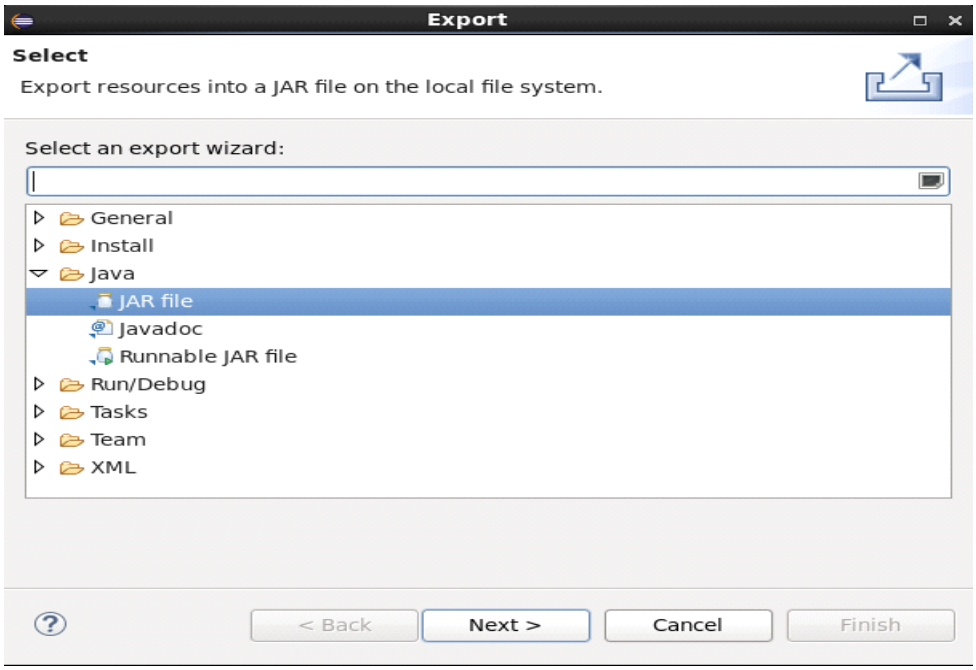
CODE: hadoop fs -put television.txt /user/acadgild

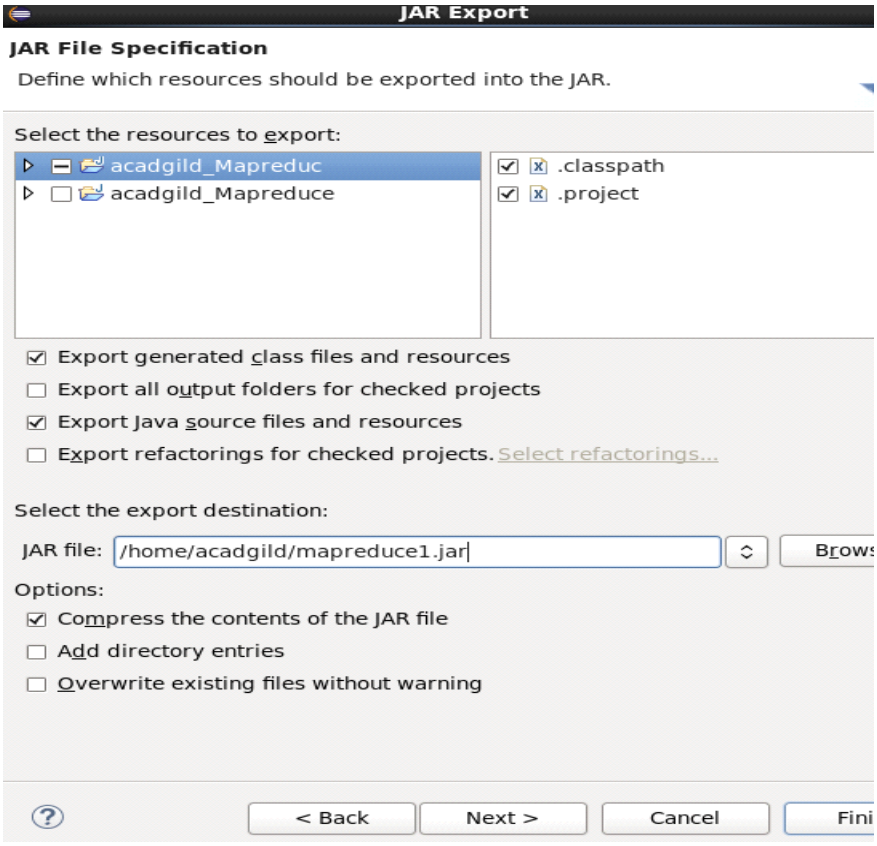
THEN CREATED A JAVA PROJECT IN ECLIPSE FOR THAT CLICK **"New --> Java Project --> ENTER NAME OF THE PROJECT --> Finish"** AS SHOWN IN THE SCREENSHOT.

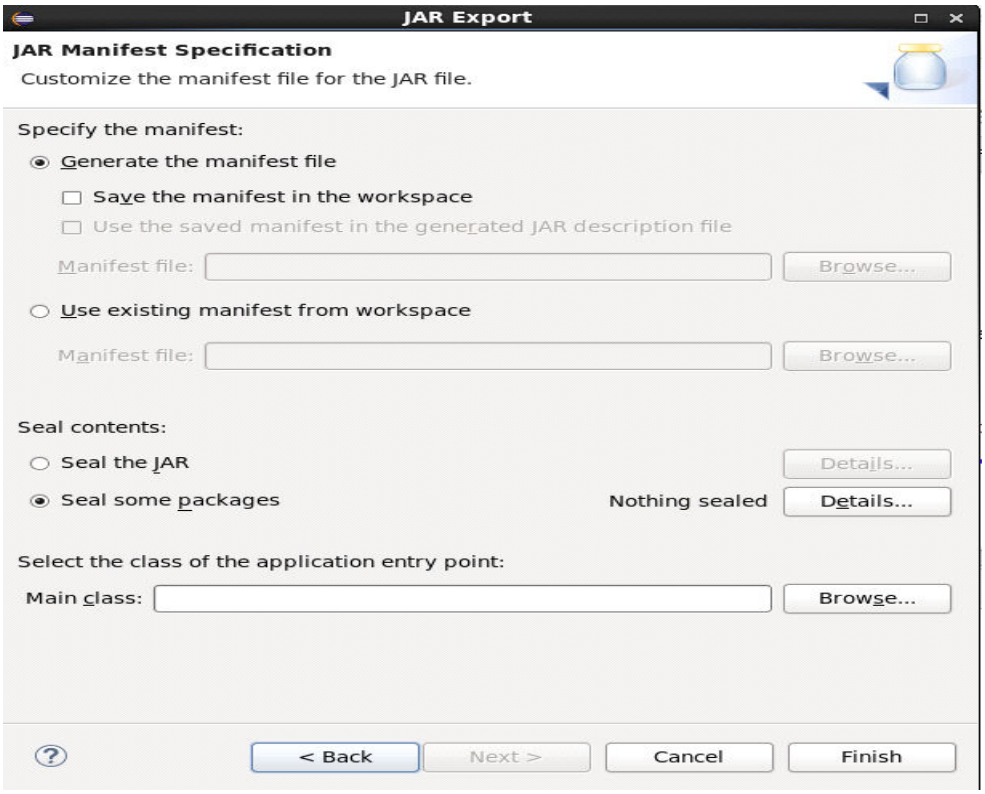


NOW U CAN SEE A JAVA PROJECT IS CREATED WHICH CAN BE SEEN ON LEFT HAND SIDE. NOW WE NEED TO CREATE A JAVA CLASS FOR OUR "MAPPER CODE". SO WE RIGHT CLICK ON OUR CREATED JAVA PROJECT AND GO TO **"New --> CLASS --> ENTER NAME OF THE CLASS --> Finish".**

NOW ENTERED THE CODE. NOW AFTER ENTERING THE CODE U WILL SEE MANY ERRORS. ITS BECAUSE WE NEED TO ADD LIBRARIES i.e JAR FILES FOR OUR MAP-REDUCE ENVIRONMNET TO WORK. SO FOR THAT RIGHT CLICK ON THE CREATED JAVA PROJECT AND THEN CLICK ON **"Build Path --> Configure Build Path --. GO TO Libraries --> Add External JARS"** THEN BROWSE THE LOCATION AND THEN ADD IT.

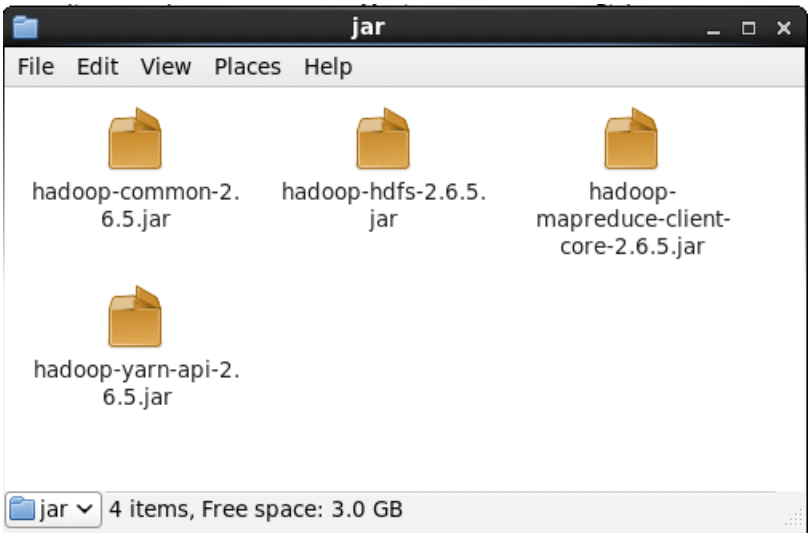






JAR FILES DOWNLOADED FROM PATH "/home/acadgild/jar".

THE JAR FILES REQUIRED ARE :



CAN BE DOWNLOADED FROM LINK "https://mvnrepository.com/artifact/org.apache.hadoop"

AFTER ADDING THE NECESSARY JAR FILES THE ERRORS ARE ELIMINATED. THE SAME WAY WE CREATE JAVA CLASS FOR "DRIVER CODE". AS THE TASK IS ONLY FOR MAP ONLY JOBS SO WE DONT REQUIRE CODE FOR "REDUCER".

AFTER ALL THESE WE THEN NEED TO EXPORT THE THESE JAVA CLASS AS A "JAR" FILE INTO OUR LINUX MACHINE SO THAT WE CAN RUN THE MAP REDUCE PROGRAM DIRECTLY THROUGH COMMANDLINE.

SO FOR CREATING A JAR FILE. SELECT BOTH THE CLASSES THEN RIGHT CLICK. THEN CLICK ON **"Export --> JAR FILE --> ENTER JAR FILE NAME --> NEXT --> THEN GO TO OPTION "MAIN CLASS" AND BROWSE THE CLASS NAME i.e "driver" --> FINISH.** WITH THIS OUR JAR FILE WITH NAME "mapreduce1.jar" IS CREATED. USING THE BELOW CODE WE CAN CHEKC THE JAR FILE CREATED. FIRST OPEN THE TERMINAL THEN ENTER:-

CODE: ls

SOLUTION REPORT:

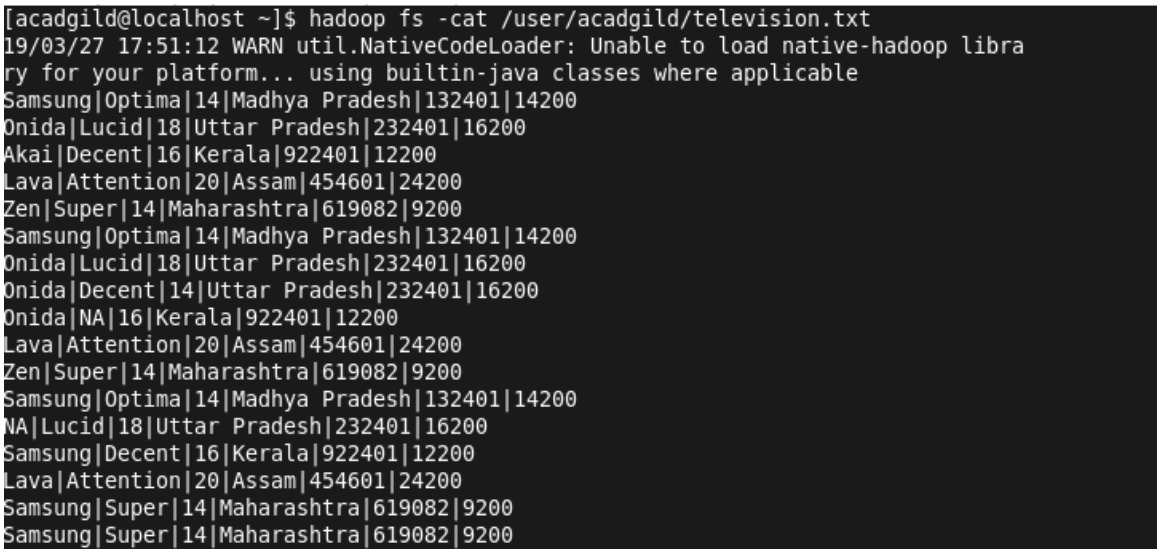
```
[acadgild@localhost ~]$ ls
CaseStudyIUseCasesDriver.java          eclipse             Pictures
CaseStudyIUseCasesMoviesMapper.java    eclipse-workspace   project
CaseStudyIUseCasesRatingsMapper.java   install             Public
CaseStudyIUseCasesReducer.java         jar                 session7.jar
commandline                            mapreduce.jar       Templates
Desktop                                Music               Videos
Documents                              new file
Downloads                              new file~
```

EXPLANATION: TO CHECK THE DATASET "television.txt" ENTER AS BELOWS:-

CODE: hadoop fs -cat /user/acadgild/television.txt

SOLUTION REPORT:

```
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/television.txt
19/03/27 17:51:12 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar Pradesh|232401|16200
Onida|NA|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
NA|Lucid|18|Uttar Pradesh|232401|16200
Samsung|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
```



EXPLANATION: NOW WE WILL RUN OUR JAR FILE. SO BEFORE THAT WE WILL START ALL THE HADOOP DAEMONS. USING CODE "start-all.sh" AND WE CAN CHECK THE STATUS OF ALL THE DAEMONS USING "jps". THEN WE ENTER THE BELOW CODE TO RUN THE JAR FILE.

CODE: hadoop jar mapreduce1.jar /user/acadgild/television.txt /user/acadgild/output

EXPLANATION: HERE **"PATH: /user/acadgild/television.txt"** IS THE LOCATION OF THE INPUT FILE TO BE EXECUTED FOR OUR MAP-REDUCE PROGRAM. THE OUTPUT WILL

BE SAVED IN
"PATH:/user/acadgild/output". WE CAN CHECK THE OUTPUT BY GOING INTO OUR OUTPUT LOCATION SAVED IN HDFS DIRECTORY BY ENTERING AS BELOWS:

CODE: hadoop fs -ls /user/acadgild/output

SOLUTION REPORT:

```
[acadgild@localhost ~]$ hadoop jar mapreduce1.jar
/user/acadgild/television.txt /user/acadgild/output
19/03/27 17:39:25 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/27 17:39:26 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
19/03/27 17:39:27 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
19/03/27 17:39:27 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/27 17:39:27 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 17:39:28 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553680429517_0002
19/03/27 17:39:29 INFO impl.YarnClientImpl: Submitted application
application_1553680429517_0002
19/03/27 17:39:29 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553680429517_0002/
19/03/27 17:39:29 INFO mapreduce.Job: Running job: job_1553680429517_0002
19/03/27 17:39:50 INFO mapreduce.Job: Job job_1553680429517_0002 running
in uber mode : false
19/03/27 17:39:50 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 17:39:58 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 17:39:59 INFO mapreduce.Job: Job job_1553680429517_0002 completed
successfully
19/03/27 17:39:59 INFO mapreduce.Job: Counters: 30
      File System Counters
            FILE: Number of bytes read=0
            FILE: Number of bytes written=107305
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=848
            HDFS: Number of bytes written=662
            HDFS: Number of read operations=5
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=2
      Job Counters
            Launched map tasks=1
            Data-local map tasks=1
            Total time spent by all maps in occupied slots (ms)=5761
            Total time spent by all reduces in occupied slots (ms)=0
            Total time spent by all map tasks (ms)=5761
            Total vcore-milliseconds taken by all map tasks=5761
            Total megabyte-milliseconds taken by all map tasks=5899264
      Map-Reduce Framework
```

```
            Map input records=18
            Map output records=16
            Input split bytes=115
            Spilled Records=0
            Failed Shuffles=0
            Merged Map outputs=0
            GC time elapsed (ms)=42
            CPU time spent (ms)=540
            Physical memory (bytes) snapshot=108277760
            Virtual memory (bytes) snapshot=2056761344
            Total committed heap usage (bytes)=62980096
      File Input Format Counters
            Bytes Read=733
      File Output Format Counters
            Bytes Written=662

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output
19/03/27 17:40:29 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 17:39
/user/acadgild/output/_SUCCESS
-rw-r--r--   1 acadgild supergroup        662 2019-03-27 17:39
/user/acadgild/output/part-m-00000

[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/output/part-m-00000
19/03/27 17:40:46 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
      Samsung|Optima|14|Madhya Pradesh|132401|14200
      Onida|Lucid|18|Uttar Pradesh|232401|16200
      Akai|Decent|16|Kerala|922401|12200
      Lava|Attention|20|Assam|454601|24200
      Zen|Super|14|Maharashtra|619082|9200
      Samsung|Optima|14|Madhya Pradesh|132401|14200
      Onida|Lucid|18|Uttar Pradesh|232401|16200
      Onida|Decent|14|Uttar Pradesh|232401|16200
      Lava|Attention|20|Assam|454601|24200
      Zen|Super|14|Maharashtra|619082|9200
      Samsung|Optima|14|Madhya Pradesh|132401|14200
      Samsung|Decent|16|Kerala|922401|12200
      Lava|Attention|20|Assam|454601|24200
      Samsung|Super|14|Maharashtra|619082|9200
      Samsung|Super|14|Maharashtra|619082|9200
      Samsung|Super|14|Maharashtra|619082|9200
```

```
[acadgild@localhost ~]$ hadoop jar mapreduce1.jar /user/acadgild/television.txt
/user/acadgild/output
19/03/27 17:39:25 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
19/03/27 17:39:26 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0
.1:8032
19/03/27 17:39:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option
 parsing not performed. Implement the Tool interface and execute your applicatio
n with ToolRunner to remedy this.
19/03/27 17:39:27 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 17:39:27 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 17:39:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
53680429517_0002
19/03/27 17:39:29 INFO impl.YarnClientImpl: Submitted application application_15
53680429517_0002
19/03/27 17:39:29 INFO mapreduce.Job: The url to track the job: http://localhost
:8088/proxy/application_1553680429517_0002/
19/03/27 17:39:29 INFO mapreduce.Job: Running job: job_1553680429517_0002
19/03/27 17:39:50 INFO mapreduce.Job: Job job_1553680429517_0002 running in uber
 mode : false
19/03/27 17:39:50 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 17:39:58 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 17:39:59 INFO mapreduce.Job: Job job_1553680429517_0002 completed succe
ssfully
19/03/27 17:39:59 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=107305
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=848
                HDFS: Number of bytes written=662
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
```

```
19/03/27 17:39:59 INFO mapreduce.Job: Job job_1553680429517_0002 completed succe
ssfully
19/03/27 17:39:59 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=107305
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=848
                HDFS: Number of bytes written=662
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=5761
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=5761
                Total vcore-milliseconds taken by all map tasks=5761
                Total megabyte-milliseconds taken by all map tasks=5899264
        Map-Reduce Framework
                Map input records=18
                Map output records=16
                Input split bytes=115
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=42
                CPU time spent (ms)=540
                Physical memory (bytes) snapshot=108277760
                Virtual memory (bytes) snapshot=2056761344
                Total committed heap usage (bytes)=62980096
        File Input Format Counters
                Bytes Read=733
        File Output Format Counters
                Bytes Written=662
```

OUTPUT: FROM THE SCREENSHOT WE CAN SEE THAT ALL THE INVALID RECORDS "NA" HAS BEEN FILTERED OUT

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output
19/03/27 17:40:29 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 17:39 /user/acadgild/ou
tput/_SUCCESS
-rw-r--r--   1 acadgild supergroup        662 2019-03-27 17:39 /user/acadgild/ou
tput/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/output/part-m-00000
19/03/27 17:40:46 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
        Samsung|Optima|14|Madhya Pradesh|132401|14200
        Onida|Lucid|18|Uttar Pradesh|232401|16200
        Akai|Decent|16|Kerala|922401|12200
        Lava|Attention|20|Assam|454601|24200
        Zen|Super|14|Maharashtra|619082|9200
        Samsung|Optima|14|Madhya Pradesh|132401|14200
        Onida|Lucid|18|Uttar Pradesh|232401|16200
        Onida|Decent|14|Uttar Pradesh|232401|16200
        Lava|Attention|20|Assam|454601|24200
        Zen|Super|14|Maharashtra|619082|9200
        Samsung|Optima|14|Madhya Pradesh|132401|14200
        Samsung|Decent|16|Kerala|922401|12200
        Lava|Attention|20|Assam|454601|24200
        Samsung|Super|14|Maharashtra|619082|9200
        Samsung|Super|14|Maharashtra|619082|9200
        Samsung|Super|14|Maharashtra|619082|9200
```

TASK 2: Write a Map Reduce program to calculate the total units sold for each Company.

SOLUTION:

DRIVER CODE:

```
package acadgild_mapreduceOne;

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 driver1 {

      public static void main(String[] args) throws Exception {
            Configuration conf = new Configuration();
            @SuppressWarnings("deprecation")
            Job job = new Job(conf, "salesunit");

            job.setJarByClass(driver1.class);
            job.setMapperClass(MyMapperOne.class);
            job.setReducerClass(MyReducerOne.class);

            job.setMapOutputKeyClass(Text.class);
            job.setMapOutputValueClass(IntWritable.class);

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

            job.setInputFormatClass(TextInputFormat.class);
            job.setOutputFormatClass(TextOutputFormat.class);

            FileInputFormat.addInputPath(job, new Path(args[0]));
            FileOutputFormat.setOutputPath(job, new Path(args[1]));

            job.waitForCompletion(true);

      }
}
```

MAPPER CODE:

```
Package acadgild_mapreduceOne;

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 MyMapperOne extends Mapper<LongWritable, Text, Text,
IntWritable> {

                Text outKey = new Text();
                IntWritable outValue = new IntWritable();

                public void map(LongWritable key, Text value, Context con)
throws IOException, InterruptedException{

                     String[] line = value.toString().split("\\|");
                     outKey.set(line[0]);
                     outValue.set(1);
                     con.write(outKey, outValue);

      }
}
```

REDUCER CODE:

```
package acadgild_mapreduceOne;

import java.io.IOException;

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

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

          public void reduce(Text keyin, Iterable<IntWritable> values,
Context con) throws IOException, InterruptedException{

                int sum = 0;
                for(IntWritable value : values){
                     sum += value.get();
                }

                con.write(keyin, new IntWritable(sum));
     }
}
```

EXPLANATION: WE DO ALL THE SAME ABOVE OPERATIONS AND CODE WHICH WE USED IN "TASK 1". SO CREATED A JAR FILE AS "mapReduce1.jar" AND IS IN **"PATH:/home/acadgild/mapReduce1.jar"**. OUR INPUT FILE IS SAME i.e "television.txt". SO HERE WE HAVE CREATED A MAPREDUCE PROGRAM TO CALCULATE THE TOTAL UNITS SOLD BY EACH COMPANY. THE OUTPUT WILL BE SAVED IN HDFS DIRECTORY IN
"PATH: /user/acadgild/output1"

SOLUTION REPORT:

```
[acadgild@localhost ~]$ ls
CaseStudyIUseCasesDriver.java          eclipse            new file~
CaseStudyIUseCasesMoviesMapper.java    eclipse-workspace  Pictures
CaseStudyIUseCasesRatingsMapper.java   install            project
CaseStudyIUseCasesReducer.java         jar                Public
commandline                            mapreduce.jar     session7.jar
Desktop                                mapReduce1.jar      Templates
Documents                              Music              Videos
Downloads                              new file

[acadgild@localhost ~]$ hadoop jar  mapReduce1.jar
/user/acadgild/television.txt /user/acadgild/output1
19/03/27 19:25:16 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/27 19:25:19 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
19/03/27 19:25:21 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
19/03/27 19:25:22 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/27 19:25:22 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 19:25:23 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553680429517_0003
19/03/27 19:25:23 INFO impl.YarnClientImpl: Submitted application
application_1553680429517_0003
19/03/27 19:25:23 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553680429517_0003/
19/03/27 19:25:23 INFO mapreduce.Job: Running job: job_1553680429517_0003
19/03/27 19:25:46 INFO mapreduce.Job: Job job_1553680429517_0003 running
in uber mode : false
19/03/27 19:25:46 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 19:26:09 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 19:26:27 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 19:26:28 INFO mapreduce.Job: Job job_1553680429517_0003 completed
successfully
19/03/27 19:26:29 INFO mapreduce.Job: Counters: 49
      File System Counters
            FILE: Number of bytes read=225
            FILE: Number of bytes written=216359
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=848
            HDFS: Number of bytes written=43
            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)=19993
            Total time spent by all reduces in occupied slots (ms)=14858
            Total time spent by all map tasks (ms)=19993
            Total time spent by all reduce tasks (ms)=14858
            Total vcore-milliseconds taken by all map tasks=19993
            Total vcore-milliseconds taken by all reduce tasks=14858
            Total megabyte-milliseconds taken by all map tasks=20472832
            Total megabyte-milliseconds taken by all reduce tasks=15214592
      Map-Reduce Framework
            Map input records=18
            Map output records=18
            Map output bytes=183
            Map output materialized bytes=225
            Input split bytes=115
            Combine input records=0
            Combine output records=0
            Reduce input groups=6
            Reduce shuffle bytes=225
            Reduce input records=18
            Reduce output records=6
            Spilled Records=36
            Shuffled Maps =1
            Failed Shuffles=0
            Merged Map outputs=1
            GC time elapsed (ms)=352
            CPU time spent (ms)=3860
            Physical memory (bytes) snapshot=320524288
            Virtual memory (bytes) snapshot=4118192128
            Total committed heap usage (bytes)=222429184
      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=43
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output1
19/03/27 19:26:55 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 19:26
/user/acadgild/output1/_SUCCESS
-rw-r--r--   1 acadgild supergroup         43 2019-03-27 19:26
/user/acadgild/output1/part-r-00000

[acadgild@localhost ~]$ hadoop fs -cat
```

/user/acadgild/output1/part-r-00000
19/03/27 19:27:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
Lava 3
NA 1
Onida 4
Samsung 7
Zen 2

```
[acadgild@localhost ~]$ hadoop jar  mapReduce1.jar /user/acadgild/television.txt
 /user/acadgild/output1
19/03/27 19:25:16 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
19/03/27 19:25:19 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0
.1:8032
19/03/27 19:25:21 WARN mapreduce.JobResourceUploader: Hadoop command-line option
 parsing not performed. Implement the Tool interface and execute your applicatio
n with ToolRunner to remedy this.
19/03/27 19:25:22 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 19:25:22 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 19:25:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
53680429517_0003
19/03/27 19:25:23 INFO impl.YarnClientImpl: Submitted application application_15
53680429517_0003
19/03/27 19:25:23 INFO mapreduce.Job: The url to track the job: http://localhost
:8088/proxy/application_1553680429517_0003/
19/03/27 19:25:23 INFO mapreduce.Job: Running job: job_1553680429517_0003
19/03/27 19:25:46 INFO mapreduce.Job: Job job_1553680429517_0003 running in uber
 mode : false
19/03/27 19:25:46 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 19:26:09 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 19:26:27 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 19:26:28 INFO mapreduce.Job: Job job_1553680429517_0003 completed succe
ssfully
19/03/27 19:26:29 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=225
                FILE: Number of bytes written=216359
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=848
                HDFS: Number of bytes written=43
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                CPU time spent (ms)=3860
                Physical memory (bytes) snapshot=320524288
                Virtual memory (bytes) snapshot=4118192128
                Total committed heap usage (bytes)=222429184
        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=43
You have new mail in /var/spool/mail/acadgild
```

```
                FILE: Number of write operations=0
                HDFS: Number of bytes read=848
                HDFS: Number of bytes written=43
                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)=19993
                Total time spent by all reduces in occupied slots (ms)=14858
                Total time spent by all map tasks (ms)=19993
                Total time spent by all reduce tasks (ms)=14858
                Total vcore-milliseconds taken by all map tasks=19993
                Total vcore-milliseconds taken by all reduce tasks=14858
                Total megabyte-milliseconds taken by all map tasks=20472832
                Total megabyte-milliseconds taken by all reduce tasks=15214592
        Map-Reduce Framework
                Map input records=18
                Map output records=18
                Map output bytes=183
                Map output materialized bytes=225
                Input split bytes=115
                Combine input records=0
                Combine output records=0
                Reduce input groups=6
                Reduce shuffle bytes=225
                Reduce input records=18
                Reduce output records=6
                Spilled Records=36
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=352
                CPU time spent (ms)=3860
                Physical memory (bytes) snapshot=320524288
                Virtual memory (bytes) snapshot=4118192128
                Total committed heap usage (bytes)=222429184
```

OUPTUT: SO FROM THE SCREENSHOT WE CAN SEE THE TOTAL UNITS SOLD BY EACH COMPANY.

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output1
19/03/27 19:26:55 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 19:26 /user/acadgild/ou
tput1/_SUCCESS
-rw-r--r--   1 acadgild supergroup         43 2019-03-27 19:26 /user/acadgild/ou
tput1/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/output1/part-m-00000
19/03/27 19:27:15 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
cat: `/user/acadgild/output1/part-m-00000': No such file or directory
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/output1/part-r-00000
19/03/27 19:27:27 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Akai    1
Lava    3
NA      1
Onida   4
Samsung 7
Zen     2
You have new mail in /var/spool/mail/acadgild
```

Task 3: Write a Map Reduce program to calculate the total units sold

in each state for Onida company.

SOLUTION:

EXPLANATION: WE DO ALL THE SAME ABOVE OPERATIONS AND CODE WHICH WE USED IN "TASK 1". SO CREATED A JAR FILE AS "mapReduce2.jar" AND IS IN **"PATH:/home/acadgild/mapReduce2.jar".** OUR INPUT FILE IS SAME i.e "television.txt". SO HERE WE HAVE CREATED A MAPREDUCE PROGRAM TO CALCULATE THE TOTAL UNITS SOLD IN EACH STATE FOR ONIDA COMPANY. THE OUTPUT WILL BE SAVED IN HDFS DIRECTORY IN **"PATH: /user/acadgild/output2".**

DRIVER CODE:

```
package acadgild_mapreduceTwo;

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 driver2 {

            public static void main(String[] args) throws Exception {
                  Configuration conf = new Configuration();
                  @SuppressWarnings("deprecation")
                  Job job = new Job(conf, "task3");

                  job.setJarByClass(driver2.class);
                  job.setMapperClass(MyMapperTwo.class);
                  job.setReducerClass(MyReducerTwo.class);

                  job.setMapOutputKeyClass(Text.class);
                  job.setMapOutputValueClass(IntWritable.class);

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

                  job.setInputFormatClass(TextInputFormat.class);
                  job.setOutputFormatClass(TextOutputFormat.class);

                  Path outputPath = new Path(args[1]);
                  FileInputFormat.addInputPath(job, new Path(args[0]));
                  FileOutputFormat.setOutputPath(job, outputPath);
                  outputPath.getFileSystem(conf).delete(outputPath,
true);

                  job.waitForCompletion(true);
            }
```

```
}
```

MAPPER CODE:

```
package acadgild_mapreduceTwo;

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 MyMapperTwo extends Mapper<LongWritable, Text, Text,
IntWritable> {

            Text outKey = new Text();
            IntWritable outValue = new IntWritable();

            public void map(LongWritable key, Text value, Context con)
throws IOException, InterruptedException{

                  String[] line = value.toString().split("\\|");
                  if (line[0].equalsIgnoreCase("Onida")) {
                        outKey.set(line[3]);
                        outValue.set(1);
                  }
                  con.write(outKey, outValue);
            }
      }
```

REDUCER CODE:

```
package acadgild_mapreduceTwo;

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

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

            public void reduce(Text keyin, Iterable<IntWritable> values,
Context con) throws IOException, InterruptedException{

                  int sum = 0;
                  for(IntWritable value : values){
                        sum += value.get();
                  }

                  con.write(keyin, new IntWritable(sum));
      }
}
```

SOLUTION REPORT:

```
[acadgild@localhost ~]$ ls
CaseStudyIUseCasesDriver.java         CaseStudyIUseCasesReducer.java
Documents  eclipse-workspace  mapReduce1.jar  Music      Pictures
session7.jar
CaseStudyIUseCasesMoviesMapper.java   commandline
Downloads  install            mapReduce2.jar  new file   project
Templates
CaseStudyIUseCasesRatingsMapper.java  Desktop
eclipse    jar                mapReduce.jar   new file~  Public
Videos

[acadgild@localhost ~]$ hadoop jar  mapReduce2.jar
/user/acadgild/television.txt /user/acadgild/output2
19/03/27 19:49:52 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/27 19:49:55 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
19/03/27 19:49:57 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
19/03/27 19:49:58 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/27 19:49:59 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 19:49:59 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553680429517_0004
19/03/27 19:50:00 INFO impl.YarnClientImpl: Submitted application
application_1553680429517_0004
19/03/27 19:50:00 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553680429517_0004/
19/03/27 19:50:00 INFO mapreduce.Job: Running job: job_1553680429517_0004
19/03/27 19:50:23 INFO mapreduce.Job: Job job_1553680429517_0004 running
in uber mode : false
19/03/27 19:50:23 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 19:50:39 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 19:50:52 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 19:50:53 INFO mapreduce.Job: Job job_1553680429517_0004 completed
successfully
19/03/27 19:50:54 INFO mapreduce.Job: Counters: 49
      File System Counters
            FILE: Number of bytes read=283
            FILE: Number of bytes written=216467
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=848
            HDFS: Number of bytes written=29
            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)=12487
            Total time spent by all reduces in occupied slots (ms)=10401
            Total time spent by all map tasks (ms)=12487
            Total time spent by all reduce tasks (ms)=10401
            Total vcore-milliseconds taken by all map tasks=12487
            Total vcore-milliseconds taken by all reduce tasks=10401
            Total megabyte-milliseconds taken by all map tasks=12786688
            Total megabyte-milliseconds taken by all reduce tasks=10650624
      Map-Reduce Framework
            Map input records=18
            Map output records=18
            Map output bytes=241
            Map output materialized bytes=283
            Input split bytes=115
            Combine input records=0
            Combine output records=0
            Reduce input groups=3
            Reduce shuffle bytes=283
            Reduce input records=18
            Reduce output records=3
            Spilled Records=36
            Shuffled Maps =1
            Failed Shuffles=0
            Merged Map outputs=1
            GC time elapsed (ms)=222
            CPU time spent (ms)=3670
            Physical memory (bytes) snapshot=317607936
            Virtual memory (bytes) snapshot=4117905408
            Total committed heap usage (bytes)=222429184
      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=29
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output2
19/03/27 19:51:15 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 19:50
/user/acadgild/output2/_SUCCESS
-rw-r--r--   1 acadgild supergroup         29 2019-03-27 19:50
/user/acadgild/output2/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat
/user/acadgild/output2/part-r-00000
19/03/27 19:51:34 WARN util.NativeCodeLoader: Unable to load native-hadoop
```

```
library for your platform... using builtin-java classes where applicable
      0
Kerala     10
Uttar Pradesh    7
```

```
^C[acadgild@localhost ~]$ hadoop jar  mapReduce2.jar /user/acadgild/television.txt /user/acadgild/output2
19/03/27 19:49:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/03/27 19:49:55 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
19/03/27 19:49:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
19/03/27 19:49:58 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 19:49:59 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 19:49:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553680429517_0004
19/03/27 19:50:00 INFO impl.YarnClientImpl: Submitted application application_1553680429517_0004
19/03/27 19:50:00 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1553680429517_0004/
19/03/27 19:50:00 INFO mapreduce.Job: Running job: job_1553680429517_0004
19/03/27 19:50:23 INFO mapreduce.Job: Job job_1553680429517_0004 running in uber mode : false
19/03/27 19:50:23 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 19:50:39 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 19:50:52 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 19:50:53 INFO mapreduce.Job: Job job_1553680429517_0004 completed successfully
19/03/27 19:50:54 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=283
                FILE: Number of bytes written=216467
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=848
                HDFS: Number of bytes written=29
                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)=12487
                Total time spent by all reduces in occupied slots (ms)=10401
                Total time spent by all map tasks (ms)=12487
                Total time spent by all reduce tasks (ms)=10401
                Total vcore-milliseconds taken by all map tasks=12487
                Total vcore-milliseconds taken by all reduce tasks=10401
                Total megabyte-milliseconds taken by all map tasks=12786688
                Total megabyte-milliseconds taken by all reduce tasks=10650624
        Map-Reduce Framework
                Map input records=18
                Map output records=18
                Map output bytes=241
                Map output materialized bytes=283
                Input split bytes=115
                Combine input records=0
                Combine output records=0
                Reduce input groups=3
                Reduce shuffle bytes=283
                Reduce input records=18
                Reduce output records=3
                Spilled Records=36
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=222
                CPU time spent (ms)=3670
                Physical memory (bytes) snapshot=317607936
                Virtual memory (bytes) snapshot=4117905408
                Total committed heap usage (bytes)=222429184
        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=29
```

OUTPUT: FROM SCREENSHOT WE CAN SEE THE TOTAL UNITS SOLD BY ONIDA COMPANY IN EACH STATES.

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/output2
19/03/27 19:51:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-27 19:50 /user/acadgild/output2/_SUCCESS
-rw-r--r--   1 acadgild supergroup         29 2019-03-27 19:50 /user/acadgild/output2/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/output2/part-r-00000
19/03/27 19:51:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
        0
Kerala  10
Uttar Pradesh   7
```