Problem Statement

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

For this task I created two classes **TV_setsMapper.java (**Mapper class**)** and **TV_sets.java (**main class**).**

TV_setsMapper.java

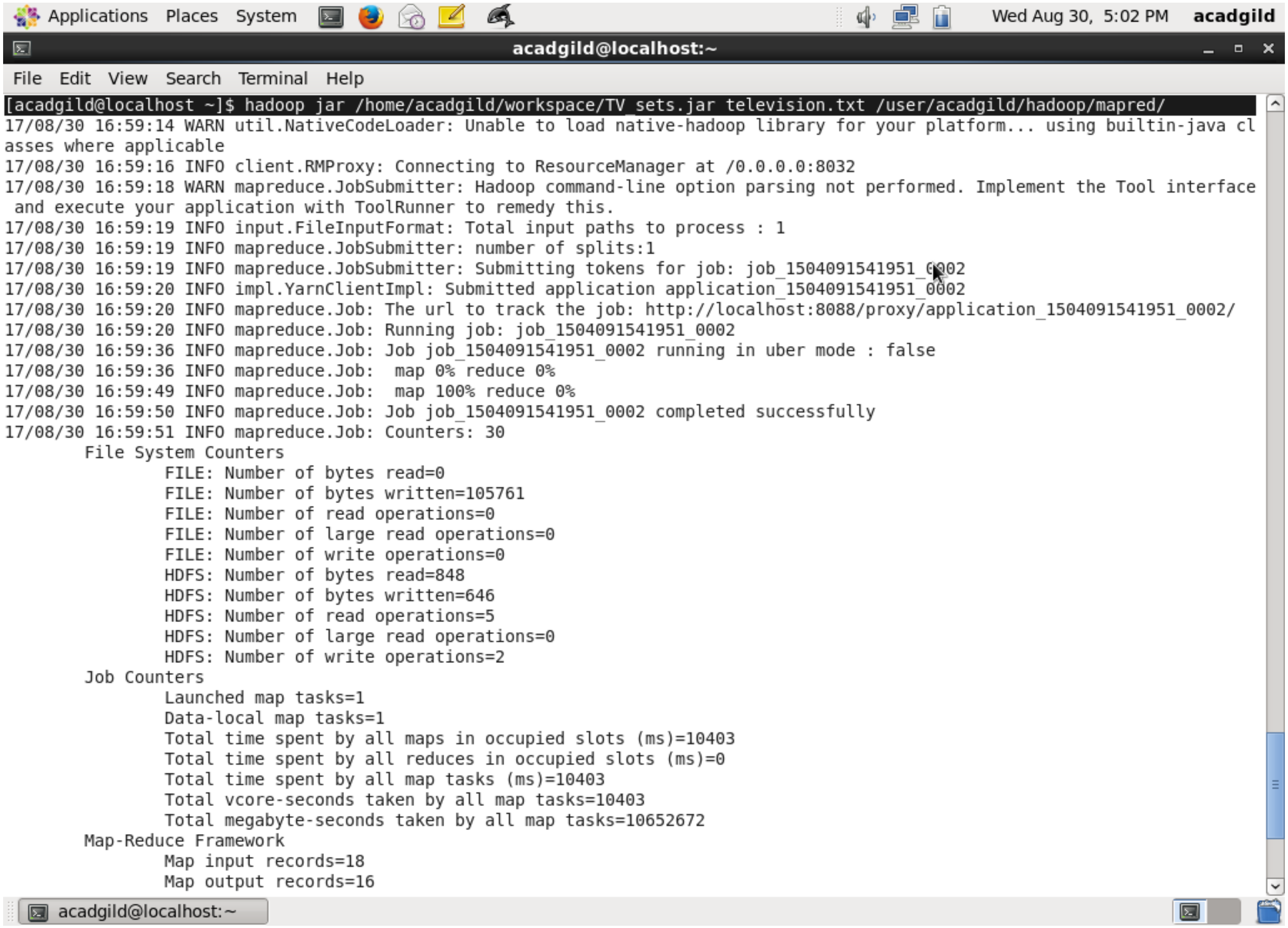
```
import org.apache.hadoop.io.*;
import java.io.IOException;
import org.apache.hadoop.mapreduce.Mapper;

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

private final static NullWritable OutVal = NullWritable.get();
private Text records =new Text();


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

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

     if(tokens[0].equals("NA") || tokens[1].equals("NA")){
          System.out.println("Invalid Record" + tokens[0] + " " + tokens[1]);
     }
     else{
          records.set(value);

          context.write(records, OutVal);
               }
      }
 }
```

TV_sets.java

```
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;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;


public class TV_sets {

     public static void main(String[] args) throws Exception {
  if (args.length != 2 ){
       System.err.println("usage : TV_sets <input path> <output path>") ;
       System.exit(-1);
 }
  Configuration conf =new Configuration();
  @SuppressWarnings("deprecation")
Job job = new Job(conf, " my TV-sets");
  job.setJarByClass(TV_sets.class);
  FileInputFormat.setInputPaths(job, args[0]);
  Path outputPath = new Path(args[1]);
  FileOutputFormat.setOutputPath(job, outputPath);
  outputPath.getFileSystem(conf).delete(outputPath, true);
  job.setMapperClass(TV_setsMapper.class);
  job.setInputFormatClass(TextInputFormat.class);
  job.setOutputFormatClass(TextOutputFormat.class);
  job.setMapOutputKeyClass(Text.class);
  job.setMapOutputValueClass(NullWritable.class);
  job.setNumReduceTasks(0);
  //execute the job
  System.exit(job.waitForCompletion(true) ? 0 :1);
  }

}
```

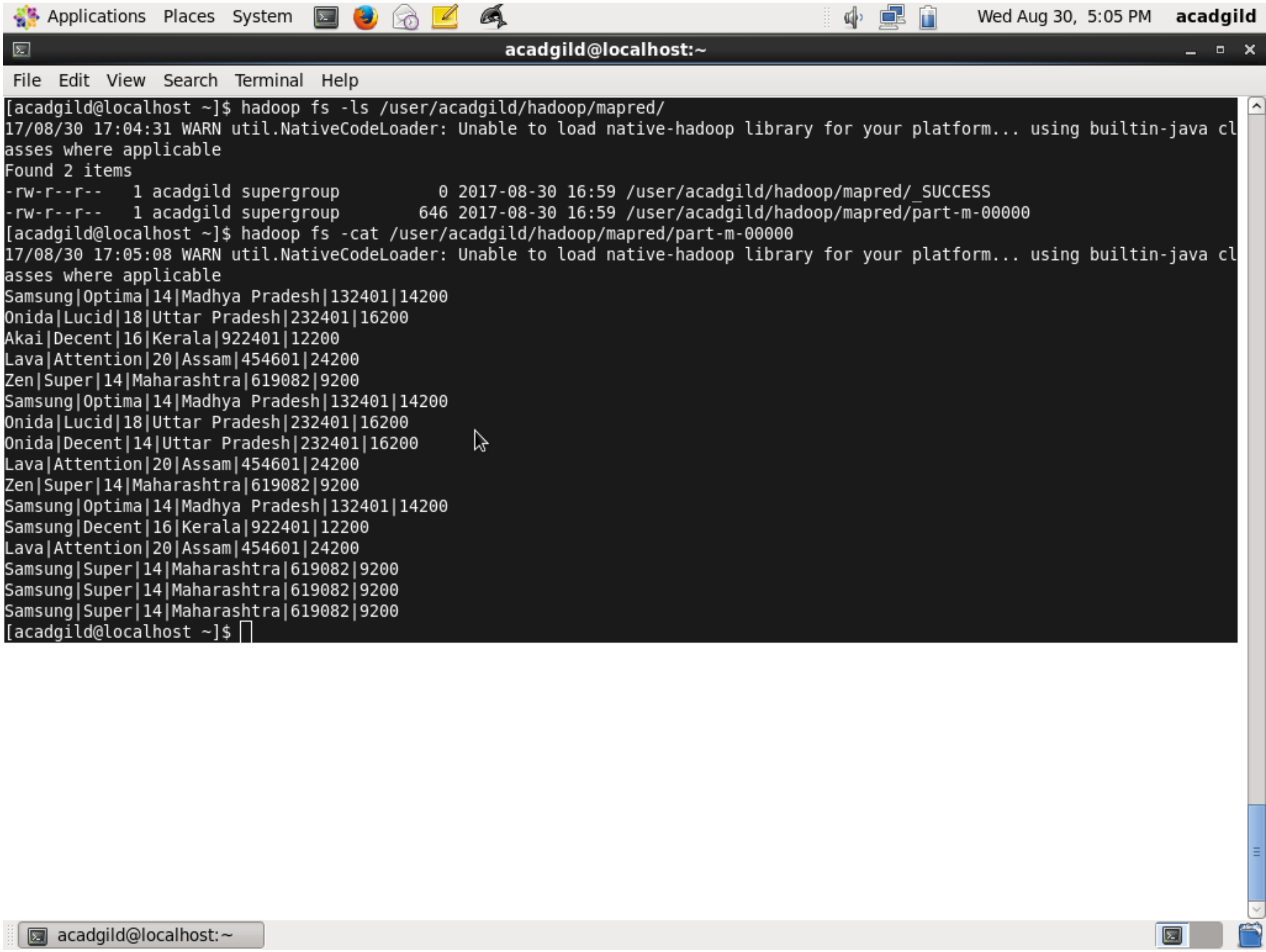
- Then we need to export the .jar file   from Eclipse and in the command shell we need to type   :

Hadoop jar /home/acadgild/workspace/TV_sets.jar television.txt user/acadgild/hadoop/mapred/



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/workspace/TV_sets.jar television.txt /user/acadgild/hadoop/mapred/
17/08/30 16:59:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
17/08/30 16:59:16 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/08/30 16:59:18 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface
 and execute your application with ToolRunner to remedy this.
17/08/30 16:59:19 INFO input.FileInputFormat: Total input paths to process : 1
17/08/30 16:59:19 INFO mapreduce.JobSubmitter: number of splits:1
17/08/30 16:59:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1504091541951_0002
17/08/30 16:59:20 INFO impl.YarnClientImpl: Submitted application application_1504091541951_0002
17/08/30 16:59:20 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1504091541951_0002/
17/08/30 16:59:20 INFO mapreduce.Job: Running job: job_1504091541951_0002
17/08/30 16:59:36 INFO mapreduce.Job: Job job_1504091541951_0002 running in uber mode : false
17/08/30 16:59:36 INFO mapreduce.Job:  map 0% reduce 0%
17/08/30 16:59:49 INFO mapreduce.Job:  map 100% reduce 0%
17/08/30 16:59:50 INFO mapreduce.Job: Job job_1504091541951_0002 completed successfully
17/08/30 16:59:51 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=105761
                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=646
                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)=10403
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=10403
                Total vcore-seconds taken by all map tasks=10403
                Total megabyte-seconds taken by all map tasks=10652672
        Map-Reduce Framework
                Map input records=18
                Map output records=16
```

- To check the results obtained on HDFS we can Type
  Hadoop fs –cat /user/acadgild/hadoop/mapred/part-m-00000



```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop/mapred/
17/08/30 17:04:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2017-08-30 16:59 /user/acadgild/hadoop/mapred/_SUCCESS
-rw-r--r--   1 acadgild supergroup        646 2017-08-30 16:59 /user/acadgild/hadoop/mapred/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/mapred/part-m-00000
17/08/30 17:05:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses 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 ~]$ 
```