Task:- Write a Map Reduce program to filter out the invalid records.

Data set:

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

Samsung|Super|14|Maharashtra|619082|9200

Code :-

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.conf.\*;

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.io.NullWritable;

import org.apache.hadoop.io.Text;

public class Task1 {

public static void main(String[] args) throws Exception {

Configuration conf = new Configuration();

Job job = new Job(conf, "Task1");

job.setJarByClass(Task1.class);

job.setOutputKeyClass(NullWritable.class);

job.setOutputValueClass(Text.class);

job.setMapperClass(Task1Mapper.class);

job.setNumReduceTasks(0);

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);

}

}

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.\*;

public class Task1Mapper extends Mapper<LongWritable, Text, NullWritable, Text> {

public void map(LongWritable key, Text value, Context context)

throws IOException, InterruptedException {

//every line will be split based on '|' and will be stored in String array

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

//will check for not null and valid record

if((lineArray.length > 0) &&

(lineArray[0] != null) &&

(lineArray[1] != null) &&

(! lineArray[0].equalsIgnoreCase("NA")) &&

(! lineArray[1].equalsIgnoreCase("NA"))

) {

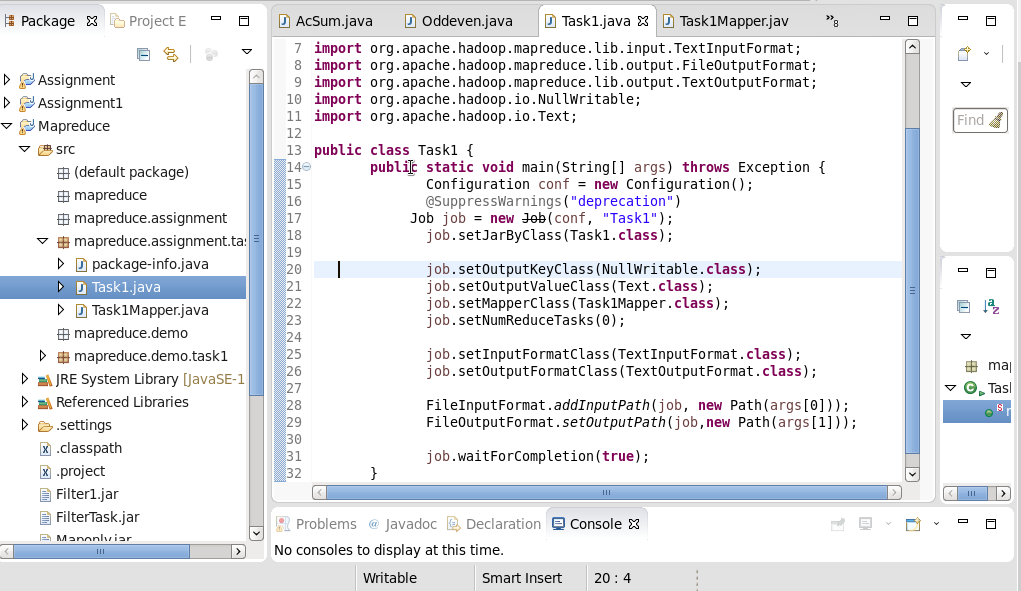
context.write(NullWritable.get(), value);

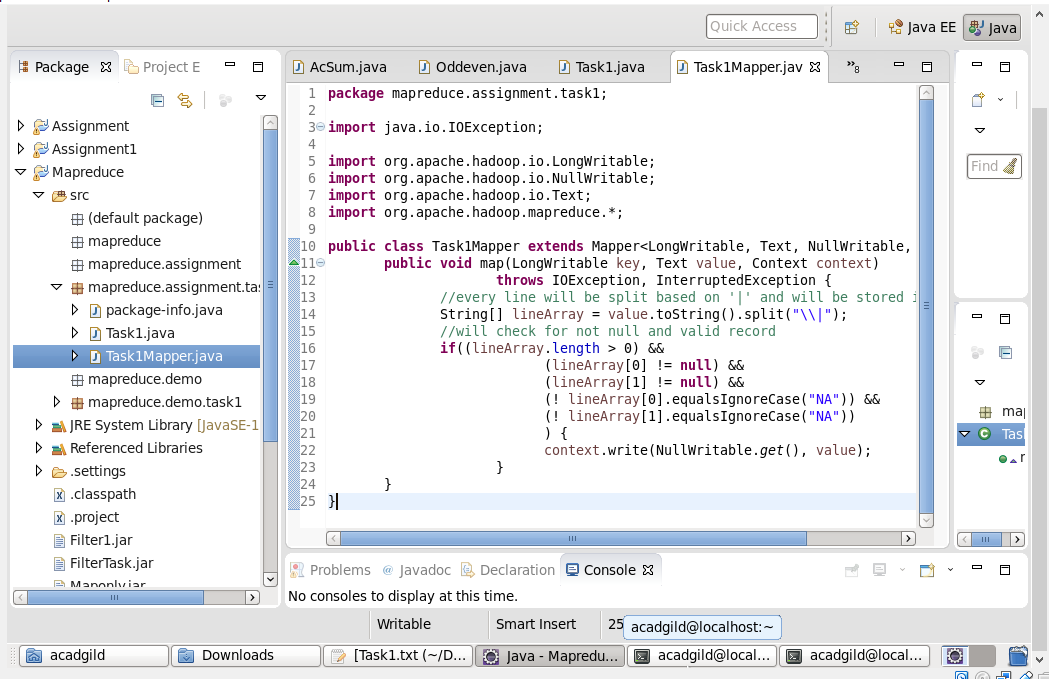
}

}

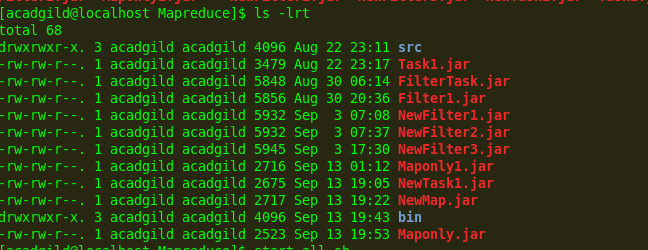
}

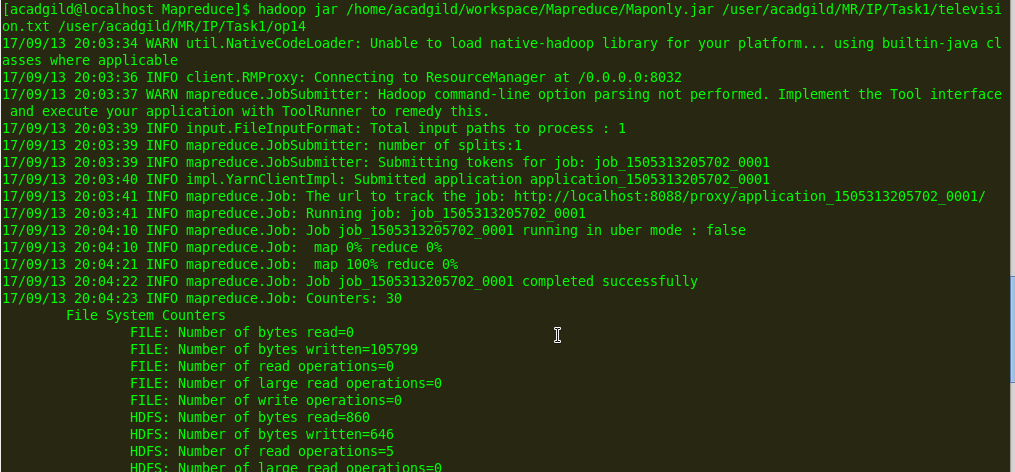
Screen Shots:-

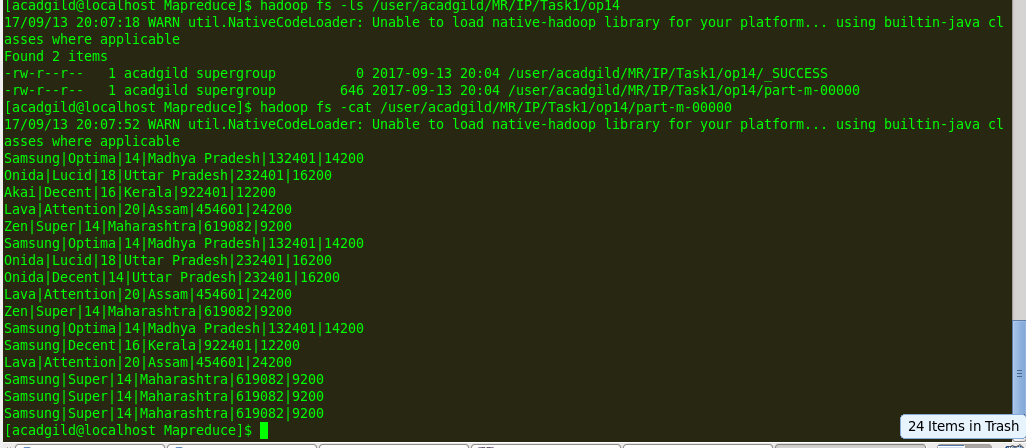




Exported Jar file named Maponly







So Invalid records are filtered.