**ASSIGNMENT- 3.2**

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.

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

**PROGRAM – *Map\_Filter.java :***

package assignment\_1;

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 Map\_filter extends Mapper<LongWritable, Text, NullWritable, Text> {

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

throws IOException, InterruptedException {

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

if((lineArray.length>0)

&& (lineArray[0] !=null)

&& (lineArray[1] !=null)

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

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

)

{

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

}

}

}

**TASK1.java:**

package assignment\_1;

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

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.NullWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.lib.input.\*;

import org.apache.hadoop.mapreduce.lib.output.\*;

import org.apache.hadoop.mapreduce.Job;

public class Task1 {

@SuppressWarnings("deprecation")

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

Configuration config = new Configuration();

Job job = new Job(config,"Demo\_job");

job.setJarByClass(Task1.class);

job.setOutputKeyClass(NullWritable.class);

job.setOutputValueClass(Text.class);

job.setInputFormatClass(TextInputFormat.class); job.setMapperClass(Map\_filter.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);

}

}

**OUTPUT:**

