**ASSIGNMENT3.2**

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

**context.**

**Job Program:**

**package filterinvalid;**

**import org.apache.hadoop.fs.Path;**

**public class FilterInvalidCls {**

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

**Configuration conf = new Configuration();**

**Job job = new Job(conf, "Map Reduce Assg3.2");**

**job.setJarByClass(FilterInvalidCls.class);**

**job.setMapOutputKeyClass(Text.class);**

**job.setMapOutputValueClass(Text.class);**

**job.setOutputKeyClass(Text.class);**

**job.setOutputValueClass(Text.class);**

**job.setMapperClass(FilterInvalidCls.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]));**

**/\***

**Path out=new Path(args[1]);**

**out.getFileSystem(conf).delete(out);**

**\*/**

**job.waitForCompletion(true);**

**}**

**}**

**Mapper Program**

package Filterinvalid;

import java.io.IOException;

public class FilterInvalidMapper extends Mapper<LongWritable, Text, IntWritable, IntWritable> {

Nullwritable key;

Text line;

@Override

public void setup(Context context) {

//key= new Nullwritable();

line=new Text();

}

@Override

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

throws IOException, InterruptedException {

String[] lineArray = value.toString();

line=new Text(lineArray);

if(!lineArray.contains("NA")){

context.write(key,line);

}

}

}