TotalUnits.java

import java.io.IOException;

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 TotalUnits {

@SuppressWarnings("deprecation")

public static void main(String[] args) throws IOException, ClassNotFoundException, InterruptedException {

// TODO Auto-generated method stub

Configuration conf = new Configuration();

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

job.setJarByClass(TotalUnits.class);

job.setMapOutputKeyClass(Text.class);

job.setMapOutputValueClass(IntWritable.class);

job.setOutputKeyClass(Text.class);

job.setOutputValueClass(IntWritable.class);

job.setMapperClass(TotalUnitsMapper.class);

job.setReducerClass(TotalUnitsReducer.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);

}

}

TotalUnitsMapper.java

import java.io.IOException;

import java.util.regex.Pattern;

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

public void map(LongWritable Key, Text Value, Context context) throws IOException, InterruptedException {

String[] lineArray = Value.toString().split(Pattern.quote("|"));

if (!lineArray[0].equals("NA")) {

Text company\_Name = new Text(lineArray[0]);

IntWritable count = new IntWritable(1);

context.write(company\_Name, count);

}

}

}

TotalUnitsReducer.java

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Reducer;

public class TotalUnitsReducer extends Reducer<Text, IntWritable, Text, IntWritable>

{

private IntWritable total = new IntWritable();

@Override

public void reduce(Text key, Iterable<IntWritable> values,Context context) throws IOException, InterruptedException

{

Integer count = 0;

for ( IntWritable value : values )

{

count++;

}

total.set(count);

context.write(key, total);

}

}