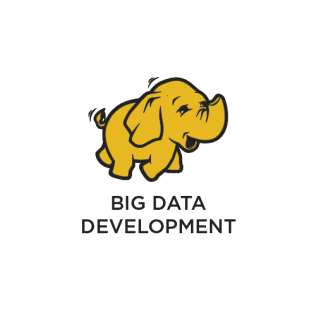
*Big Data and Hadoop Development*



Session 3: Exploring MapReduce

Assignment 2

*Downloaded datasets from given link:-*

https://drive.google.com/file/d/0Bxr27gVaXO5sVjQ5QW0wQ3RCTUU/view?usp=sharing

# Problem Statement

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.

**FilterRecords.java**

**package** mapreduce.television;

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

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

**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.output.FileOutputFormat;

**public** **class** FilterRecords {

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

Configuration c = **new** Configuration();

Job mapOnlyJob = **new** Job(c, "FilterRecords");

mapOnlyJob.setJarByClass(FilterRecords.**class**);

mapOnlyJob.setMapperClass(MapForFilterRecords.**class**);

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

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

mapOnlyJob.setNumReduceTasks(0);

FileInputFormat.*addInputPath*(mapOnlyJob, **new** Path(args[0]));

FileOutputFormat.*setOutputPath*(mapOnlyJob, **new** Path(args[1]));

System.*exit*(mapOnlyJob.waitForCompletion(**true**) ? 0 : 1);

}

}

MapForFilterRecords.java

**package** mapreduce.television;

**import** java.io.IOException;

**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** MapForFilterRecords **extends** Mapper<LongWritable, Text, Text, IntWritable> {

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

String line = value.toString();

String NA = "NA";

String[] words = line.split("\\|");

String companyName = words[0];

String productName = words[1];

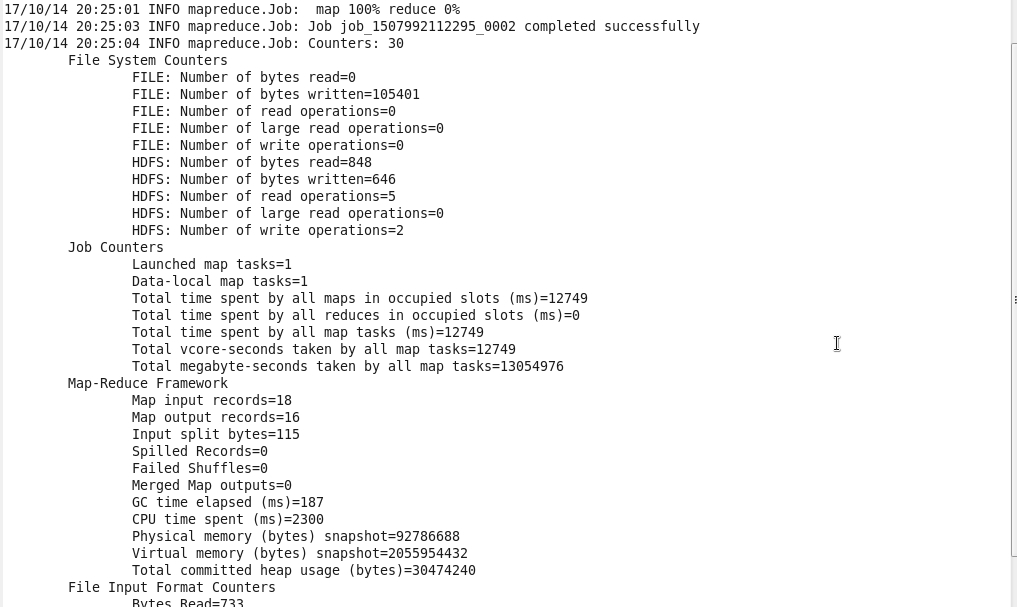
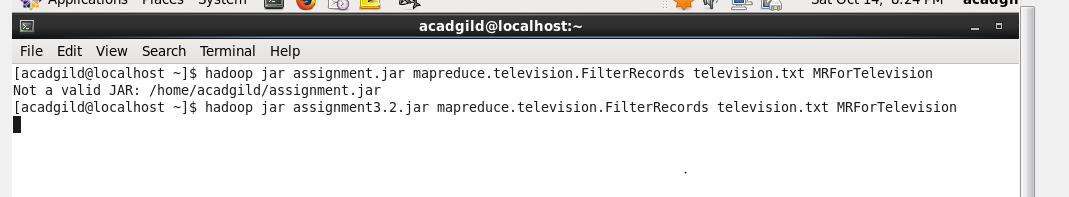
**if** (!(companyName.equals(NA) || productName.equals(NA))) {

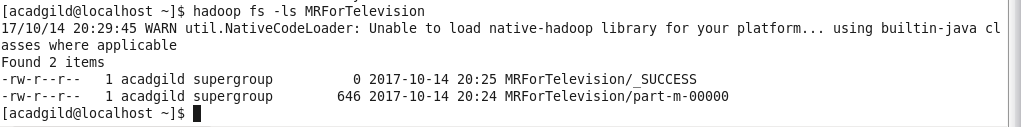
context.write(value, **null**);

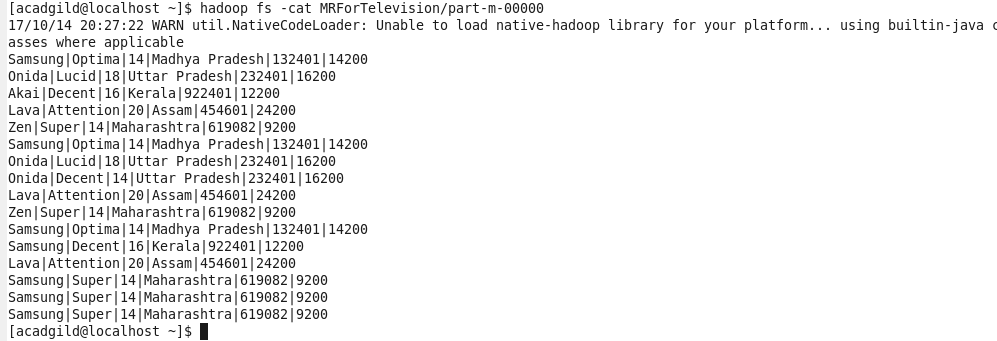
}

}

}

****

**For the output **

****