**BIG DATA PROJECT REPORT** 

**Country Sales Analysis**

Prepared By:

Aishwarya R Shanbhag-PES2201800006

Navya P Hegde-PES2201800017

Shrushti Hegde-PES2201800652

For our big data project we have chosen the country sales dataset .Our dataset consists of the columns such as Transaction\_Id ,Product,Price,Country.By making use of simple map reduce program written in java,we are trying to figure out the total number of products sold in each country. In this project we are making use of various big data tools like hadoop and hive .We are using hive inorder to get better insights of our dataset.Using hive queries ,any random person can get to know the about the dataset.Here is a snippet of our code:

import java.io.IOException;

import java.util.StringTokenizer;

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.Mapper;

import org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class Sales {

public static class Map extends Mapper<Object, Text, Text, IntWritable>{

private final static IntWritable one = new IntWritable(1);

private Text word = new Text();

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

String Line = value.toString();

StringTokenizer itr = new StringTokenizer(Line,",");

while (itr.hasMoreTokens()) {

itr.nextToken();

itr.nextToken();

itr.nextToken();

word.set(itr.nextToken());

context.write(word, one);

}

}

}

public static class Reduce extends Reducer<Text,IntWritable,Text,IntWritable> {

private IntWritable result = new IntWritable();

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

int sum = 0;

for (IntWritable val : values) {

sum += val.get();

}

result.set(sum);

context.write(key, result);

}

}

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

Configuration conf = new Configuration();

Job job = Job.getInstance(conf, "Country Sales");

job.setJarByClass(Sales.class);

job.setMapperClass(Map.class);

job.setCombinerClass(Reduce.class);

job.setReducerClass(Reduce.class);

job.setOutputKeyClass(Text.class);

job.setOutputValueClass(IntWritable.class);

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

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

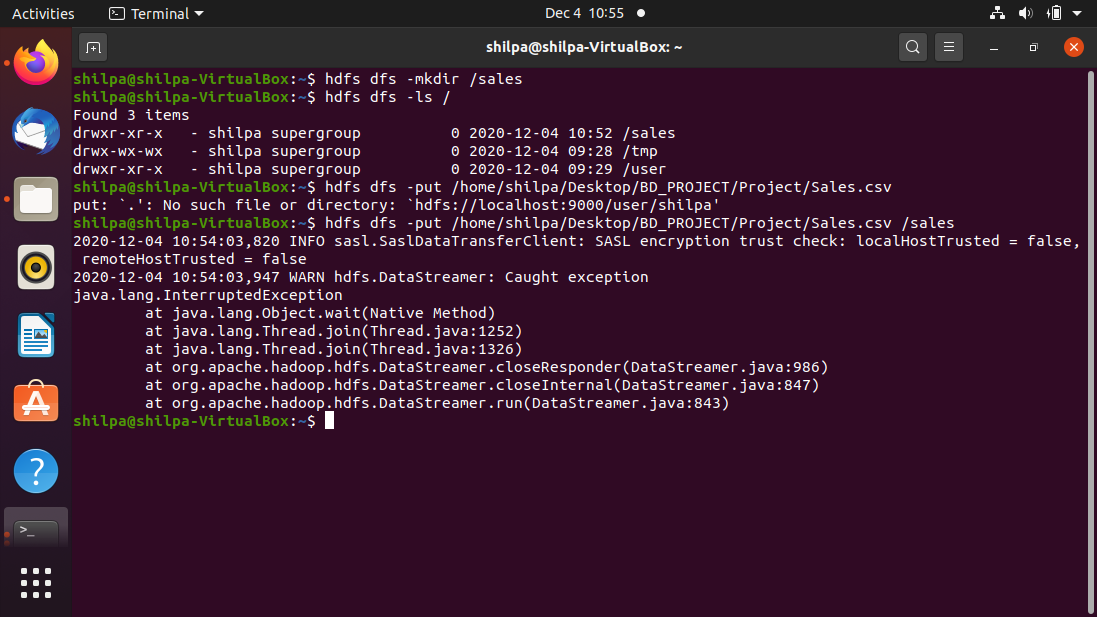
System.exit(job.waitForCompletion(true) ? 0 : 1);

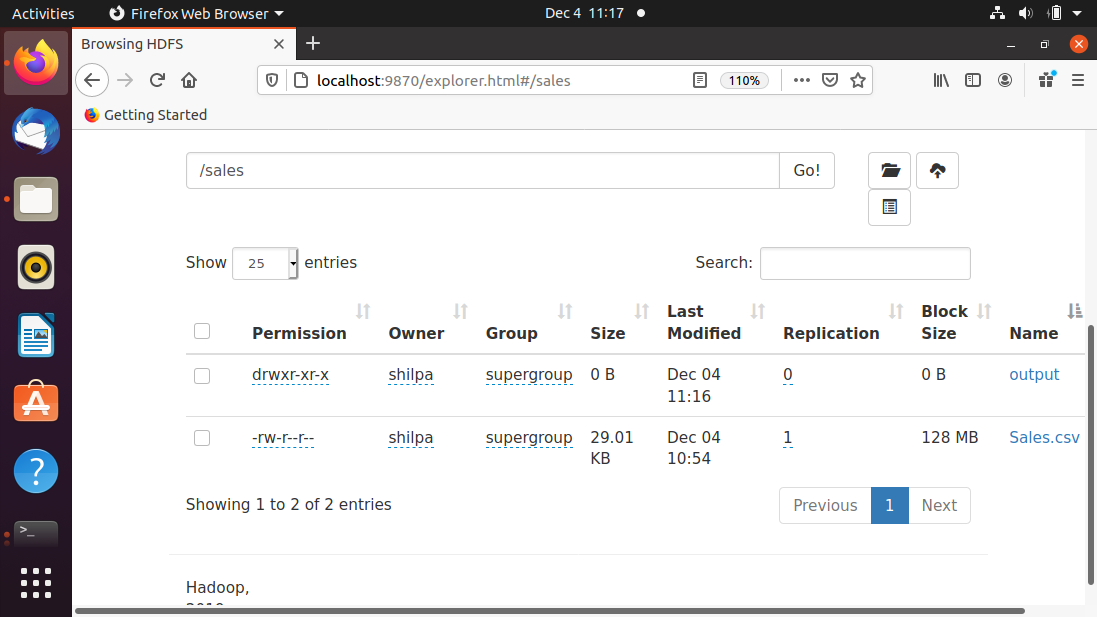
}

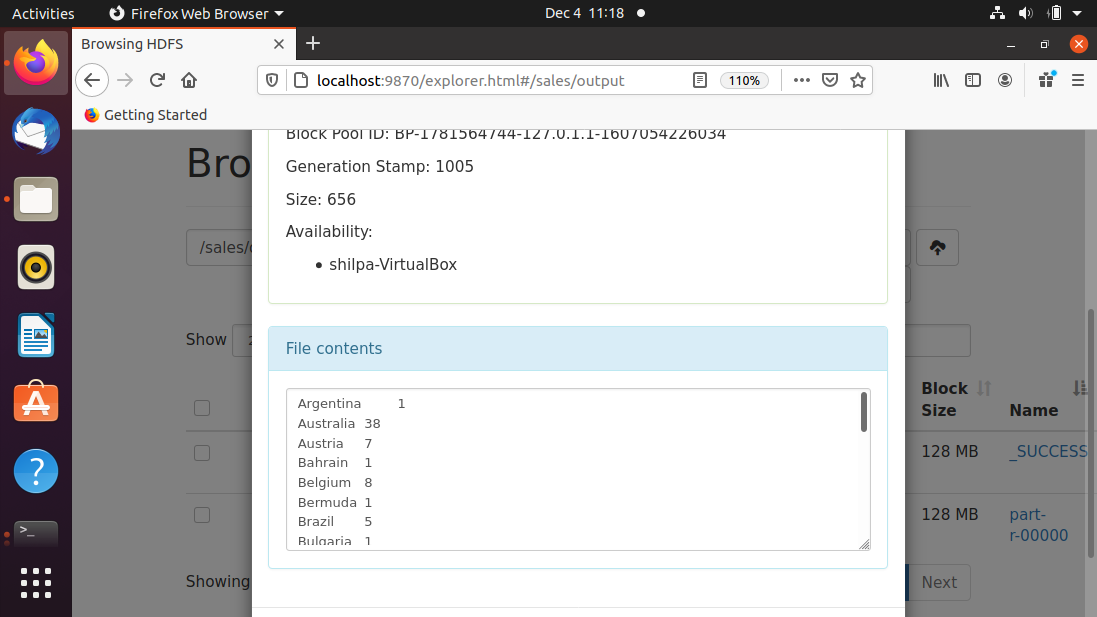
}

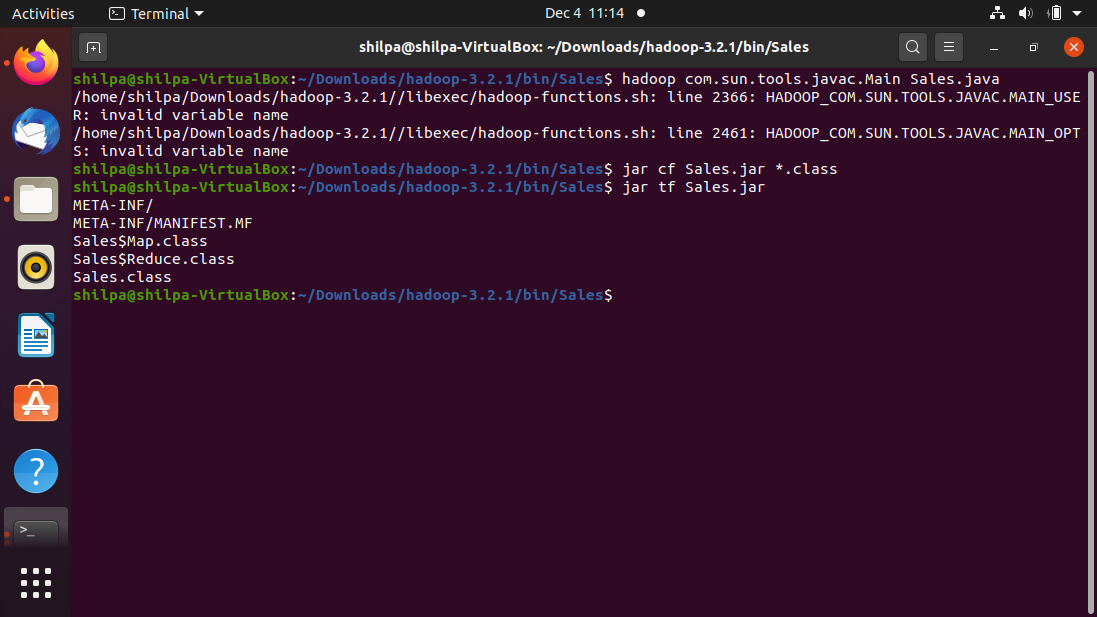
Dataset link: <https://drive.google.com/file/d/1o84zkFxy9sG74aayxJzxnRLhm9bzqage/view>

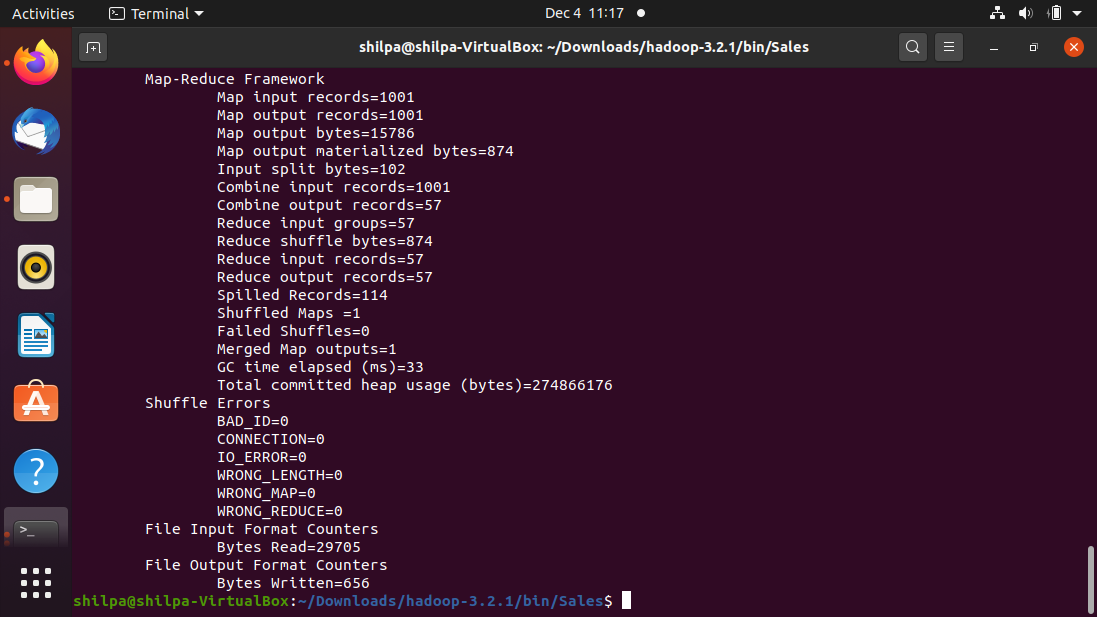
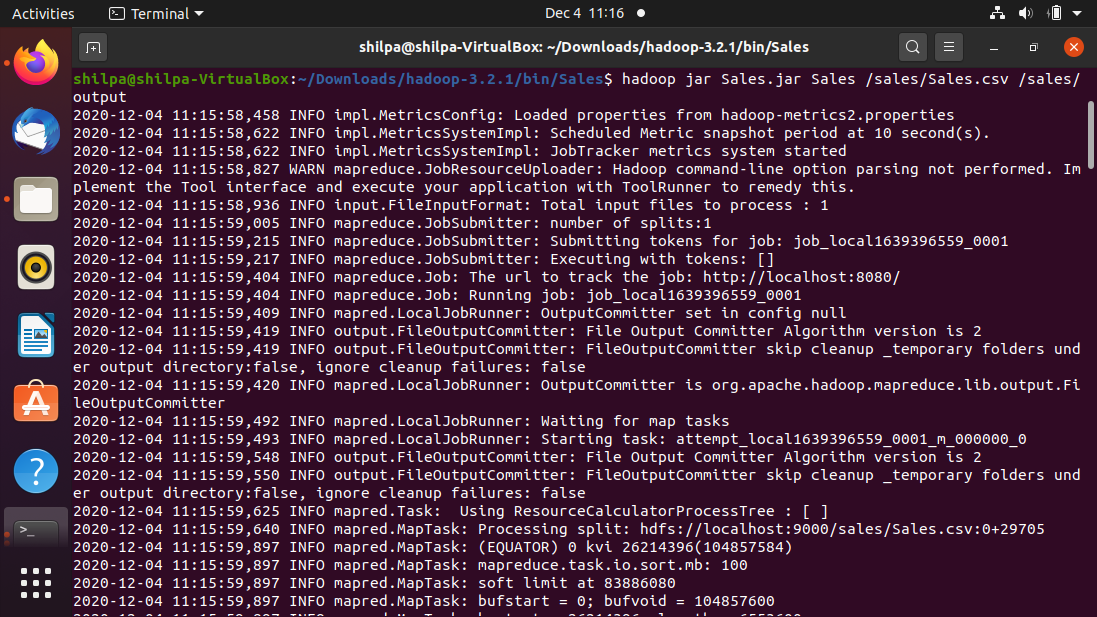
Output screenshots of the Hadoop tool:











Output screenshots of the hive queries:

