Q1. MapReduce Code for averageclosingprice:

Code :

import java.io.\*;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.io.DoubleWritable;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.mapreduce.Job;

import org.apache.hadoop.mapreduce.Mapper;

import org.apache.hadoop.mapreduce.Reducer;

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

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

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

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

public class Module\_end {

public static class MapClass extends Mapper<LongWritable,Text,Text,DoubleWritable>

{

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

{

try{

String[] str = value.toString().split(",");

Double closing\_price = Double.*parseDouble*(str[6]);

context.write(new Text(str[1]),new DoubleWritable(closing\_price));

}

catch(Exception e)

{

System.***out***.println(e.getMessage());

}

}

}

public static class ReduceClass extends Reducer<Text,DoubleWritable,Text,DoubleWritable>

{

private DoubleWritable result = new DoubleWritable();

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

double sum = 0;

int count=0;

for (DoubleWritable val : values)

{

sum += val.get();

count+=1;

}

double avg = sum/count;

result.set(avg);

context.write(key, result);

}

}

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

Configuration conf = new Configuration();

Job job = Job.*getInstance*(conf, "AVG closing price for each Stock");

job.setJarByClass(Module\_end.class);

job.setMapperClass(MapClass.class);

job.setReducerClass(ReduceClass.class);

job.setNumReduceTasks(1);

job.setOutputKeyClass(Text.class);

job.setOutputValueClass(DoubleWritable.class);

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

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

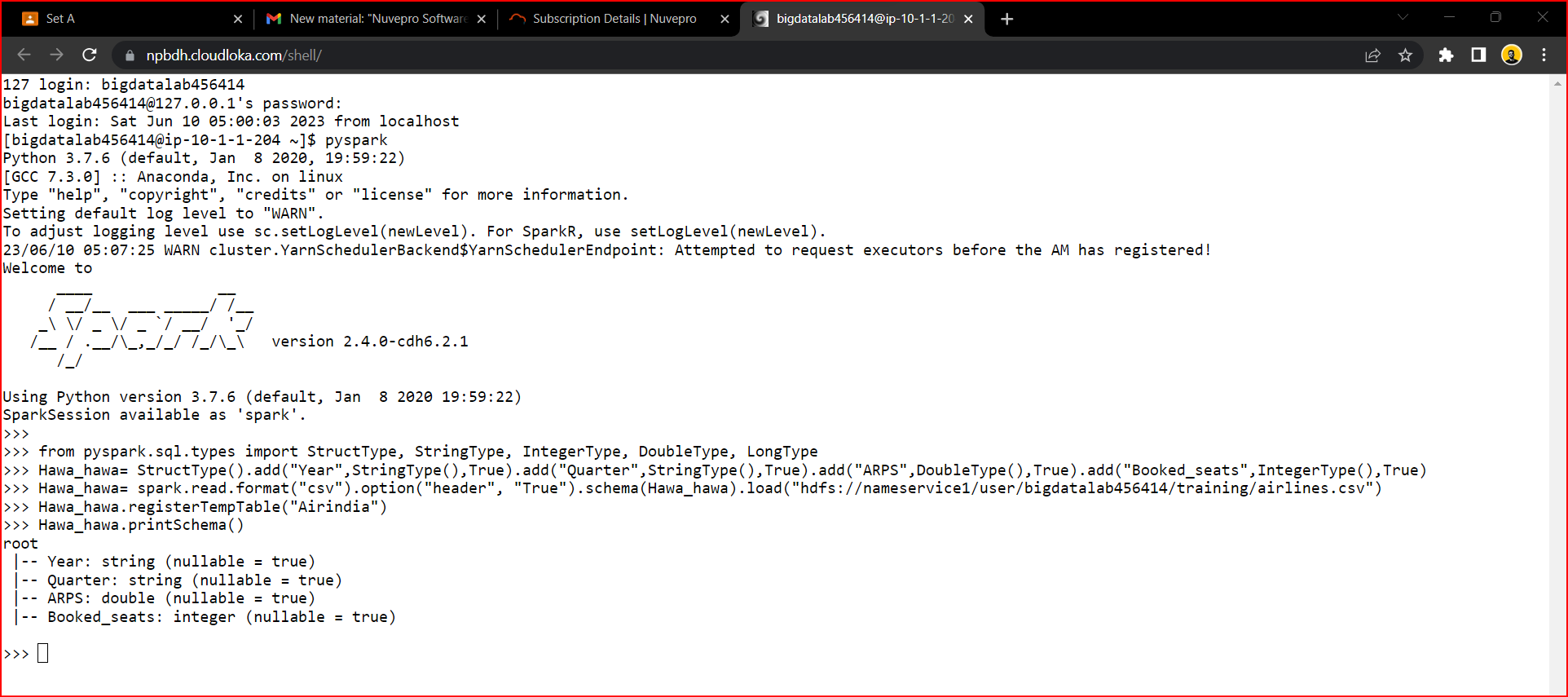
System.*exit*(job.waitForCompletion(true) ? 0 : 1);

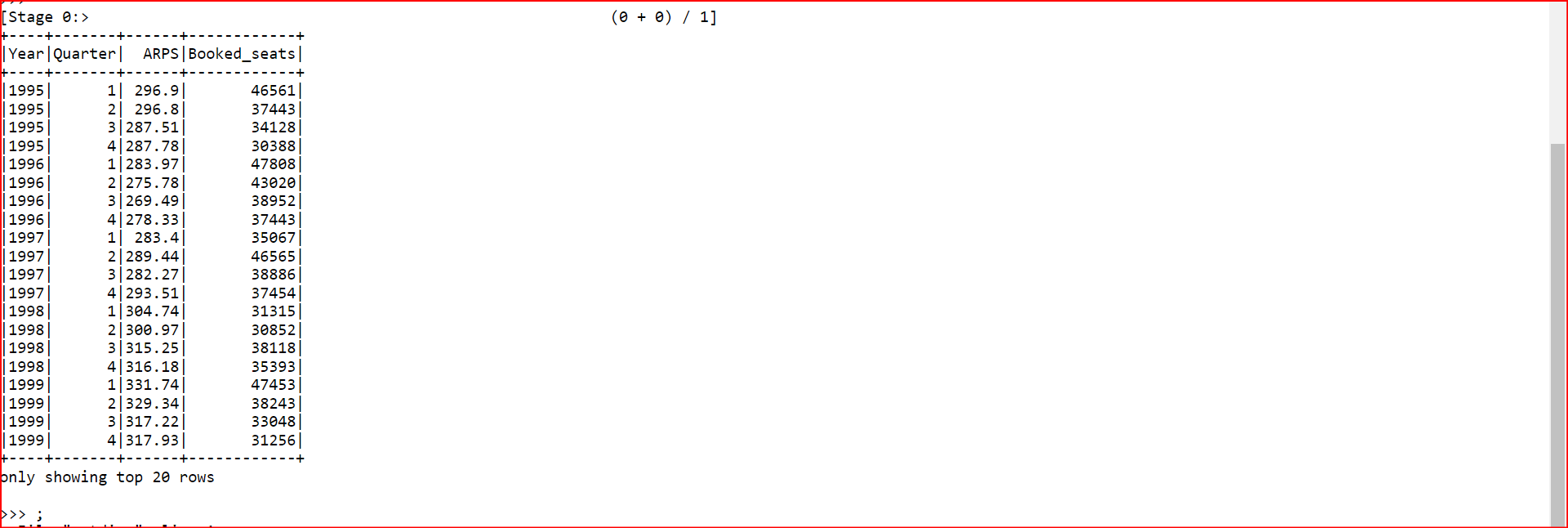
}

}

Jar File uploaded : on git hub link mentioned below,

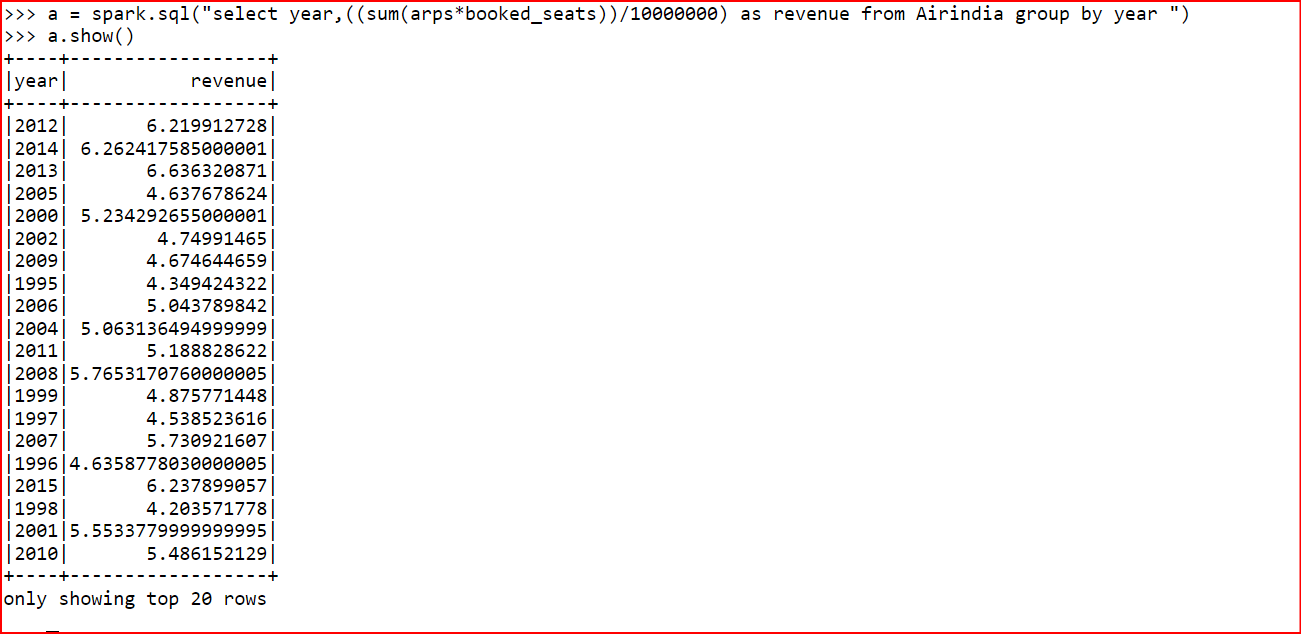
Q3. (SPARK ) Use Airlines dataset for the Below Questions: [ Marks: 10 ]





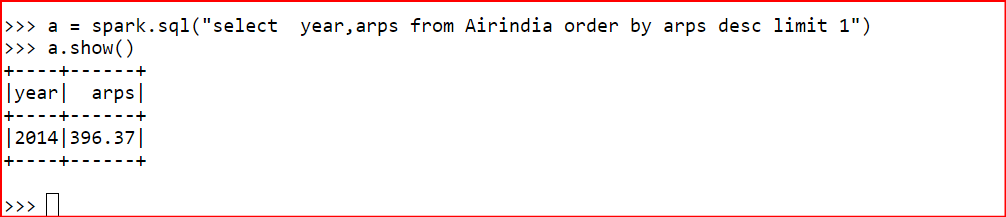
1. What is the total revenue generated in each year?

Command : a = spark.sql("select year,((sum(arps\*booked\_seats))/10000000) as revenue from airlines group by year ")



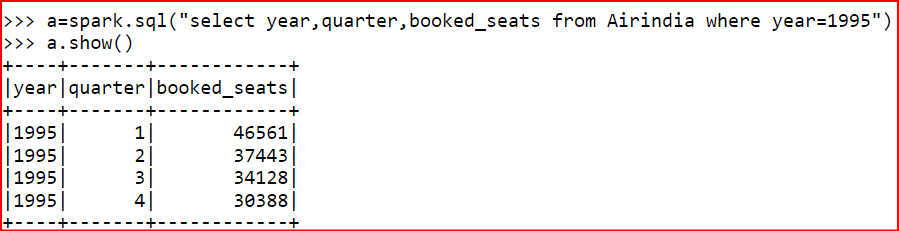
2.

A=spark.sql(“select year,arps from Airindia order by arps desc limit 1”)



3. What is the total number of booked seats for each quarter in a given year?

a=spark.sql(“select year,quarter,booked\_seats from Airindia where year=1995”)



1)Which airports have the highest altitude ?

=> select airport\_id,name,country,altitude from airport order by altitude desc limit 1;

2)How many routes are operated by active airlines from the United States ?

=> select distinct a.name from airlines a join routes r

on a.airline\_id = r.airline\_id join airport ar1 on ar1.airport\_id = src\_airport\_id

join

airport ar2

on

ar2.airport\_id = dest\_airport\_id where trim(upper(ar1.country)) = "UNITED STATES" or trim(upper(ar2.country)) = "UNITED STATES";

3)Which airlines operate routes that have less than 3 stops number of stops

top 10 alphabetically?

* select air.airline\_id,air.name,country,r.stops from airlines as air inner join routes r on air.airline\_id = r.airline\_id where stops<3 order by air.name limit 10;

4)How many airlines have a specific IATA code ‘W9’?

* Select count(distinct(airline\_id)) as count from airlines where iata='W9';

5)Find the airlines that operate routes with a specific equipment as ‘AN4’

and codeshare enabled.

* Select distinct(a.name) from airlines a join routes r on a.airline\_id=r.airline\_id where equipment='AN4' and codeshare='Y';