BIG DATA

Name : S L A Laisha

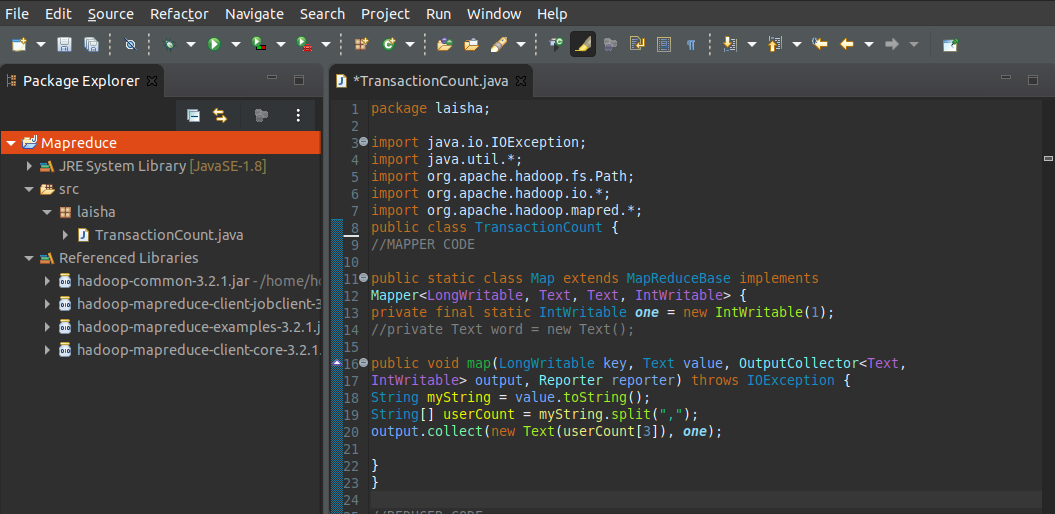
USN : 1NT19IS147

Sec: C1

Date: 21.06.2022

4.TH PROGRAM

Run the code in eclipse:



CODE:

package laisha;

import java.io.IOException;

import java.util.\*;

import org.apache.hadoop.fs.Path;

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

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

public class TransactionCount {

//MAPPER CODE

public static class Map extends MapReduceBase implements

Mapper<LongWritable, Text, Text, IntWritable> {

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

//private Text word = new Text();

public void map(LongWritable key, Text value, OutputCollector<Text,

IntWritable> output, Reporter reporter) throws IOException {

String myString = value.toString();

String[] userCount = myString.split(",");

output.collect(new Text(userCount[3]), one);

}

}

//REDUCER CODE

public static class Reduce extends MapReduceBase implements

Reducer<Text, IntWritable, Text, IntWritable> {

public void reduce(Text key, Iterator<IntWritable> values,

OutputCollector<Text, IntWritable> output, Reporter reporter) throws

IOException { //{little: {1,1}}

int finaluserCount = 0 ;

Text mykey = key ;

while(values.hasNext()) {

IntWritable value = values.next();

finaluserCount += value.get();

}

output.collect(mykey, new IntWritable(finaluserCount));

}

}

//DRIVER CODE

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

JobConf conf = new JobConf(TransactionCount.class);

conf.setJobName("wordcount");

conf.setOutputKeyClass(Text.class);

conf.setOutputValueClass(IntWritable.class);

conf.setMapperClass(Map.class);

conf.setCombinerClass(Reduce.class);

conf.setReducerClass(Reduce.class);

conf.setInputFormat(TextInputFormat.class);

conf.setOutputFormat(TextOutputFormat.class); // hadoop jar

//jarname classpath inputfolder outputfolder

FileInputFormat.setInputPaths(conf, new Path(args[0]));

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

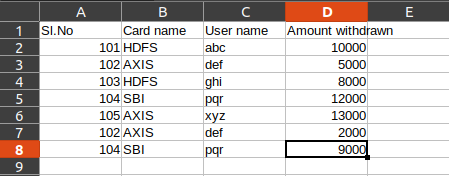
JobClient.runJob(conf);

}

}

Create a new jar file(laishanew.jar) and csv file(4laisha.csv):

CSV File:(excel)

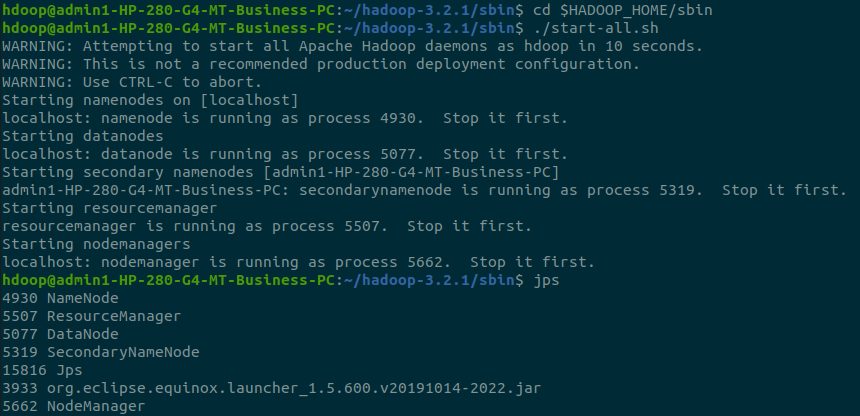


Terminal:

> cd $HADOOP\_HOME/sbin

> ./start-all.sh

> jps



Creating a input file (give123)

Copying the 4laisha.csv file from local to hdfs

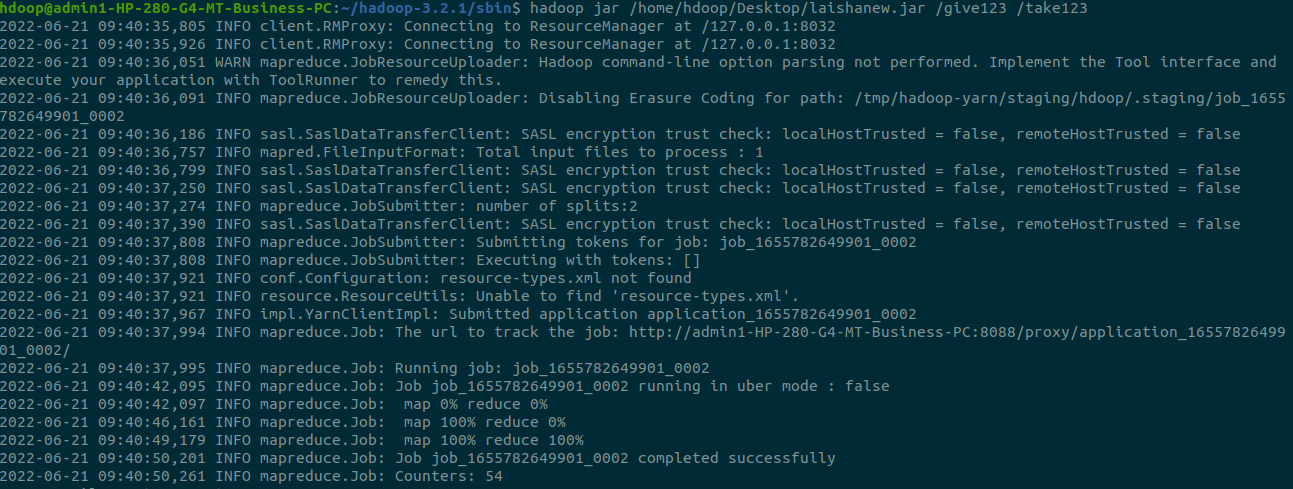
> hdfs dfs -mkdir -p /give123

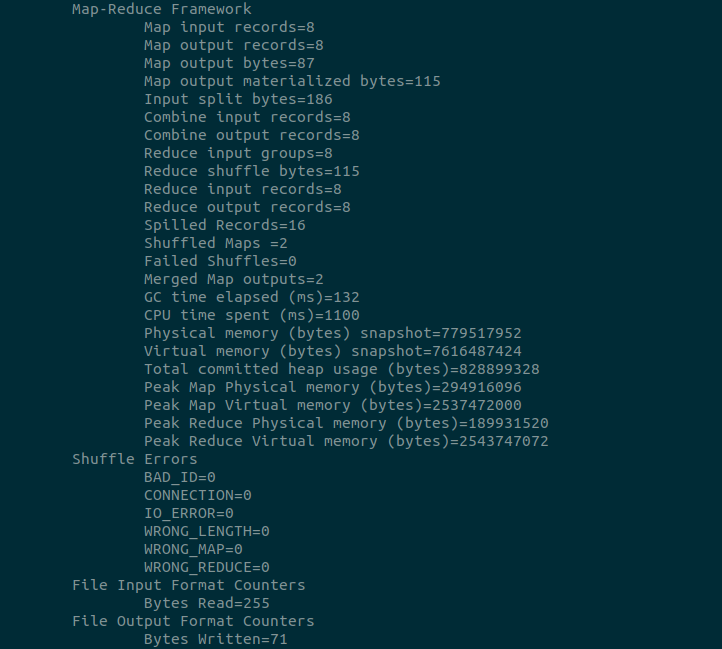
> hdfs dfs -copyFromLocal /home/hdoop/Desktop/4laisha.csv /give123



> hadoop jar /home/hdoop/Desktop/laishanew.jar /give123 /take123

By using the above command, we can print the classpath needed to get the Hadoop jar and required libraries.





To get output:

> hdfs dfs -cat /take123/part\*

