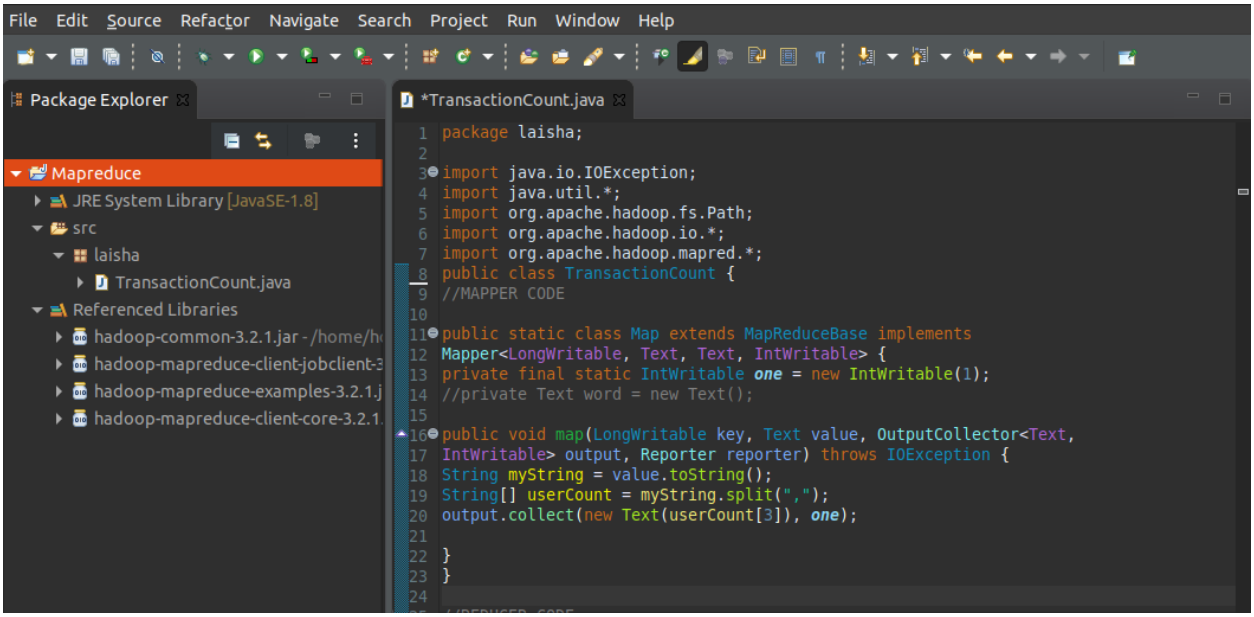
BIG DATA

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)

	A	B	C	D	E
1	Sl.No	Card name	User name	Amount withdrawn	
2	101	HDFS	abc	10000	
3	102	AXIS	def	5000	
4	103	HDFS	ghi	8000	
5	104	SBI	pqr	12000	
6	105	AXIS	xyz	13000	
7	102	AXIS	def	2000	
8	104	SBI	pqr	9000	
9					

Terminal:

> cd $HADOOP_HOME/sbin
> ./start-all.sh
> jps

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ cd $HADOOP_HOME/sbin
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 4930.  Stop it first.
Starting datanodes
localhost: datanode is running as process 5077.  Stop it first.
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process 5319.  Stop it first.
Starting resourcemanager
resourcemanager is running as process 5507.  Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 5662.  Stop it first.
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
4930 NameNode
5507 ResourceManager
5077 DataNode
5319 SecondaryNameNode
15816 Jps
3933 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
5662 NodeManager
```

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

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /give123
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hdoop/Desktop/4laisha.csv /give123
2022-06-21 09:39:48,567 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```

> 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.

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/laishanew.jar /give123 /take123
2022-06-21 09:40:35,805 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 09:40:35,926 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 09:40:36,051 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2022-06-21 09:40:36,091 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1655
782649901_0002
2022-06-21 09:40:36,186 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 09:40:36,757 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-21 09:40:36,799 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 09:40:37,250 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 09:40:37,274 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-21 09:40:37,390 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 09:40:37,808 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655782649901_0002
2022-06-21 09:40:37,808 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-21 09:40:37,921 INFO conf.Configuration: resource-types.xml not found
2022-06-21 09:40:37,921 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-21 09:40:37,967 INFO impl.YarnClientImpl: Submitted application application_1655782649901_0002
2022-06-21 09:40:37,994 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_16557826499
01_0002/
2022-06-21 09:40:37,995 INFO mapreduce.Job: Running job: job_1655782649901_0002
2022-06-21 09:40:42,095 INFO mapreduce.Job: Job job_1655782649901_0002 running in uber mode : false
2022-06-21 09:40:42,097 INFO mapreduce.Job:  map 0% reduce 0%
2022-06-21 09:40:46,161 INFO mapreduce.Job:  map 100% reduce 0%
2022-06-21 09:40:49,179 INFO mapreduce.Job:  map 100% reduce 100%
2022-06-21 09:40:50,201 INFO mapreduce.Job: Job job_1655782649901_0002 completed successfully
2022-06-21 09:40:50,261 INFO mapreduce.Job: Counters: 54
```

```
        Map-Reduce Framework
                Map input records=8
                Map output records=8
                Map output bytes=87
                Map output materialized bytes=115
                Input split bytes=186
                Combine input records=8
                Combine output records=8
                Reduce input groups=8
                Reduce shuffle bytes=115
                Reduce input records=8
                Reduce output records=8
                Spilled Records=16
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=132
                CPU time spent (ms)=1100
                Physical memory (bytes) snapshot=779517952
                Virtual memory (bytes) snapshot=7616487424
                Total committed heap usage (bytes)=828899328
                Peak Map Physical memory (bytes)=294916096
                Peak Map Virtual memory (bytes)=2537472000
                Peak Reduce Physical memory (bytes)=189931520
                Peak Reduce Virtual memory (bytes)=2543747072
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=255
        File Output Format Counters
                Bytes Written=71
```

To get output:

> hdfs dfs -cat /take123/part*

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /take123/part*
2022-06-21 09:41:51,504 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
10000   1
12000   1
13000   1
2000    1
5000    1
8000    1
9000    1
Amount withdrawn        1
```