SUSHMITHA P A

1NT19IS170

C2

EXERCISE 4

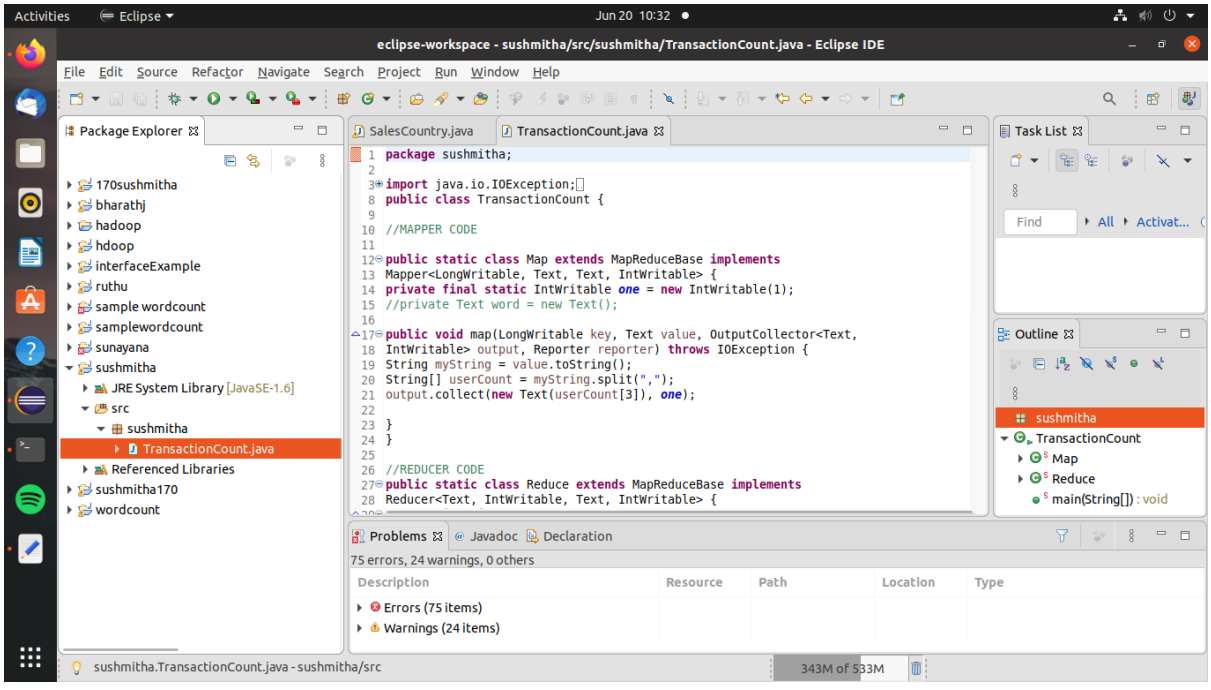
Use the Hadoop framework to write a MapReduce program to read a csv. File into a single node Hadoop cluster containing following fields

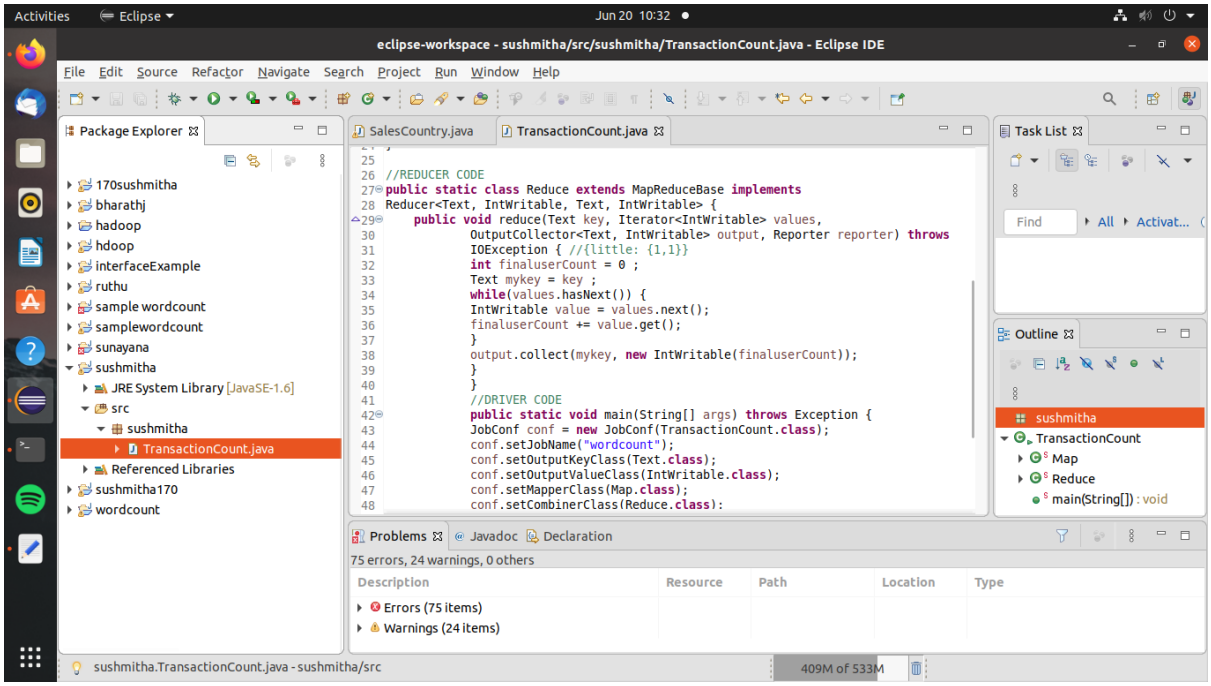
- **Sl No.**
- **Card name**
- **Username**
- **Amount**

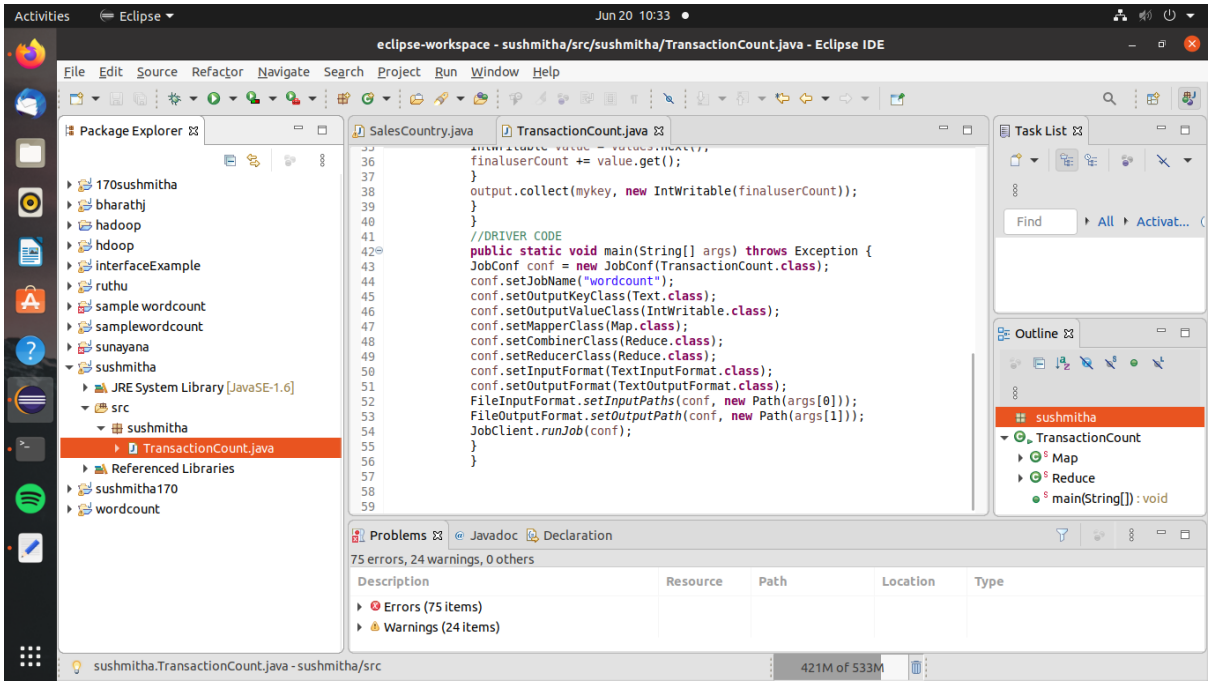
```
package sushmitha;

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);
FileInputFormat.setInputPaths(conf, new Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
JobClient.runJob(conf);
}}
```

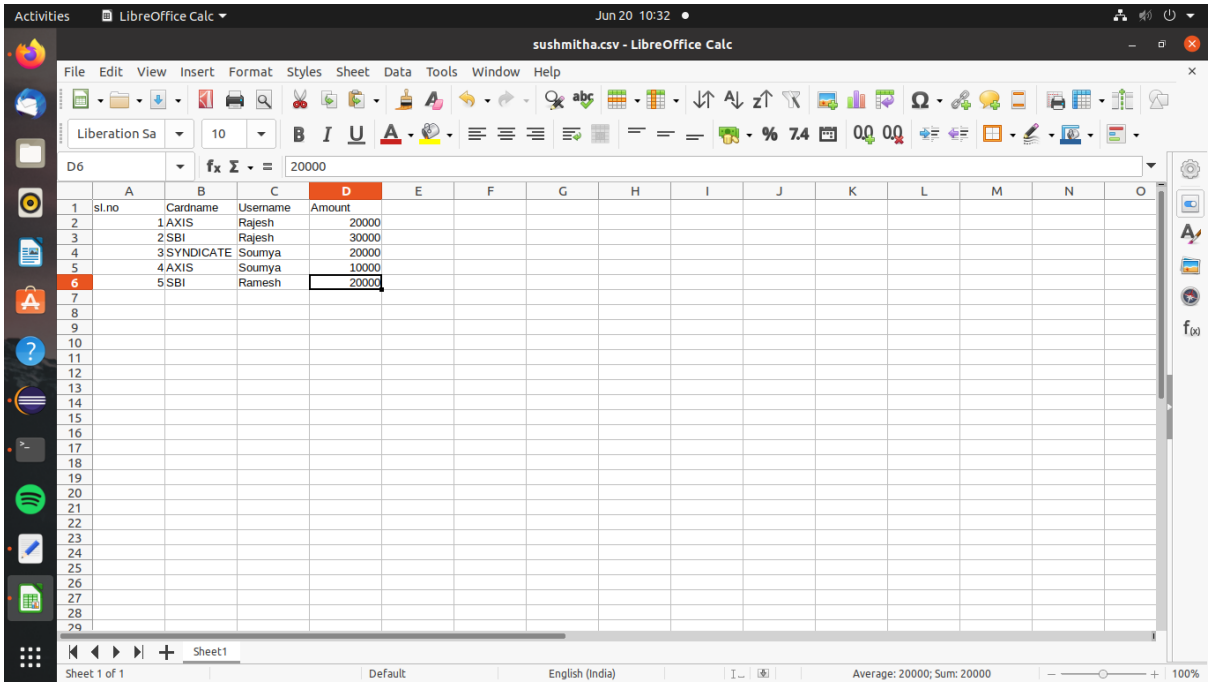
Compiling the program on eclipse




Activities
Eclipse
Jun 20 10:32
eclipse-workspace - sushmitha/src/sushmitha/TransactionCount.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
170sushmitha
bharathj
hadoop
hdoop
interfaceExample
ruthu
sample wordcount
samplewordcount
sunayana
sushmitha
JRE System Library [JavaSE-1.6]
src
sushmitha
TransactionCount.java
Referenced Libraries
sushmitha170
wordcount
SalesCountry.java
TransactionCount.java
//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);
Task List
Find All Activat...
Outline
sushmitha
TransactionCount
Map
Reduce
main(String[]) : void
Problems
Javadoc
Declaration
75 errors, 24 warnings, 0 others
Description
Resource
Path
Location
Type
Errors (75 items)
Warnings (24 items)
sushmitha.TransactionCount.java - sushmitha/src
409M of 533M


Activities
Eclipse
Jun 20 10:33
eclipse-workspace - sushmitha/src/sushmitha/TransactionCount.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
170sushmitha
bharathj
hadoop
hdoop
interfaceExample
ruthu
sample wordcount
samplewordcount
sunayana
sushmitha
JRE System Library [JavaSE-1.6]
src
sushmitha
TransactionCount.java
Referenced Libraries
sushmitha170
wordcount
SalesCountry.java
TransactionCount.java
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);
FileInputFormat.setInputPaths(conf, new Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
JobClient.runJob(conf);
}
}
Task List
Find All Activat...
Outline
sushmitha
TransactionCount
Map
Reduce
main(String[]) : void
Problems
Javadoc
Declaration
75 errors, 24 warnings, 0 others
Description
Resource
Path
Location
Type
Errors (75 items)
Warnings (24 items)
sushmitha.TransactionCount.java - sushmitha/src
421M of 533M

Creating a csv file




```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
7776 NodeManager
7044 NameNode
7626 ResourceManager
14365 Jps
5773 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
7438 SecondaryNameNode
7199 DataNode
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 7044.  Stop it first.
Starting datanodes
localhost: datanode is running as process 7199.  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 7438.  Stop it first.
Starting resourcemanager
resourcemanager is running as process 7626.  Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 7776.  Stop it first.
```

Creating an input directory

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /170input
```

Copying the contents of the csv file to HDFS

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hdoop/Desktop/sushmitha.csv /input70
2022-06-20 10:26:25,833 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
```

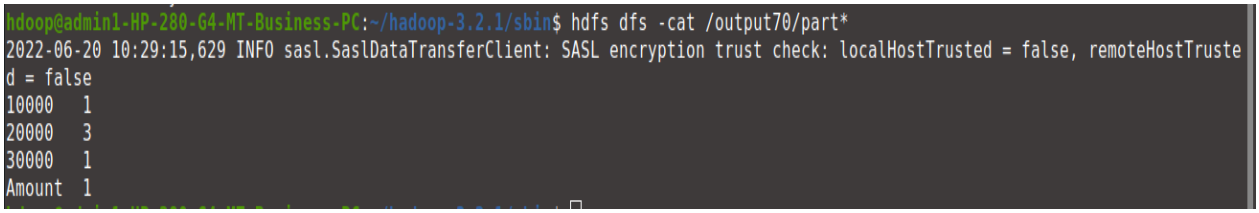
Running the job by passing in the input and output directories

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/sushmitha170.jar /input70 /output70
2022-06-20 10:28:04,102 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-20 10:28:04,221 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-20 10:28:04,347 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy this.
2022-06-20 10:28:04,374 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.st
aging/job_1655698442856_0004
2022-06-20 10:28:04,472 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
2022-06-20 10:28:04,602 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-20 10:28:04,656 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
2022-06-20 10:28:04,698 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
2022-06-20 10:28:04,714 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-20 10:28:04,814 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
2022-06-20 10:28:05,240 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655698442856_0004
2022-06-20 10:28:05,240 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-20 10:28:05,381 INFO conf.Configuration: resource-types.xml not found
2022-06-20 10:28:05,381 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-20 10:28:05,423 INFO impl.YarnClientImpl: Submitted application application_1655698442856_0004
2022-06-20 10:28:05,447 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/applicati
on_1655698442856_0004/
2022-06-20 10:28:05,448 INFO mapreduce.Job: Running job: job_1655698442856_0004
2022-06-20 10:28:10,515 INFO mapreduce.Job: Job job_1655698442856_0004 running in uber mode : false
2022-06-20 10:28:10,517 INFO mapreduce.Job:  map 0% reduce 0%
2022-06-20 10:28:13,632 INFO mapreduce.Job:  map 100% reduce 0%
2022-06-20 10:28:17,664 INFO mapreduce.Job:  map 100% reduce 100%
2022-06-20 10:28:17,685 INFO mapreduce.Job: Job job_1655698442856_0004 completed successfully
2022-06-20 10:28:17,775 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=67
                FILE: Number of bytes written=677593
```

```
        File System Counters
                FILE: Number of bytes read=67
                FILE: Number of bytes written=677593
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=363
                HDFS: Number of bytes written=33
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=3552
                Total time spent by all reduces in occupied slots (ms)=1508
                Total time spent by all map tasks (ms)=3552
                Total time spent by all reduce tasks (ms)=1508
                Total vcore-milliseconds taken by all map tasks=3552
                Total vcore-milliseconds taken by all reduce tasks=1508
                Total megabyte-milliseconds taken by all map tasks=3637248
                Total megabyte-milliseconds taken by all reduce tasks=1544192
        Map-Reduce Framework
                Map input records=6
                Map output records=6
                Map output bytes=61
                Map output materialized bytes=73
                Input split bytes=162
                Combine input records=6
                Combine output records=5
                Reduce input groups=4
                Reduce shuffle bytes=73
                Reduce input records=5
```

```
                Reduce input groups=4
                Reduce shuffle bytes=73
                Reduce input records=5
                Reduce output records=4
                Spilled Records=10
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=128
                CPU time spent (ms)=1180
                Physical memory (bytes) snapshot=819392512
                Virtual memory (bytes) snapshot=7611478016
                Total committed heap usage (bytes)=854589440
                Peak Map Physical memory (bytes)=336121856
                Peak Map Virtual memory (bytes)=2533154816
                Peak Reduce Physical memory (bytes)=192483328
                Peak Reduce Virtual memory (bytes)=2545389568
        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=201
        File Output Format Counters
                Bytes Written=33
```

Output


hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output70/part*
2022-06-20 10:29:15,629 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTruste
d = false
10000 1
20000 3
30000 1
Amount 1