

Shwetha
1NT19IS152
C1 batch

Exercise-4: Map Reduce (Programs)

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Amount withdrawn

Implement the following,

1. Count the Number of transactions done by each user
2. Find the total amount of money transacted by each user

```
package shwethahebbbar;  
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);

}

}

```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME/sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
12721 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
14722 Jps
14373 NodeManager
13641 NameNode
14217 ResourceManager
14029 SecondaryNameNode
13790 DataNode
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /input152

```

Copying shwetha.csv file from local to hdfs.

```
$ hdfs dfs -copyFromLocal /home/hadoop/Desktop/shwetha.csv /input152
```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hadoop/Desktop/shwetha.csv /input152
2022-06-21 10:00:27,599 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

```

By using this above command, we can print the class path needed to get the Hadoop jar and required libraries

```
n$ hadoop jar /home/hadoop/Desktop/shwethahebbbar.jar /input152 /output152
```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/shwethahebbbar.jar /input152 /output152
2022-06-21 10:09:55,087 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 10:09:55,214 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 10:09:55,347 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-06-21 10:09:55,369 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1655785624817_0001/
2022-06-21 10:09:55,457 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:09:55,559 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-21 10:09:55,593 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:09:55,644 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:09:55,668 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-21 10:09:55,776 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:09:55,792 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655785624817_0001
2022-06-21 10:09:55,792 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-21 10:09:55,907 INFO conf.Configuration: resource-types.xml not found
2022-06-21 10:09:55,908 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-21 10:09:56,256 INFO impl.YarnClientImpl: Submitted application application_1655785624817_0001
2022-06-21 10:09:56,314 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655785624817_0001/
2022-06-21 10:09:56,316 INFO mapreduce.Job: Running job: job_1655785624817_0001
2022-06-21 10:10:02,385 INFO mapreduce.Job: Job job_1655785624817_0001 running in uber mode : false
2022-06-21 10:10:02,387 INFO mapreduce.Job: map 0% reduce 0%
2022-06-21 10:10:06,471 INFO mapreduce.Job: map 100% reduce 0%
2022-06-21 10:10:09,506 INFO mapreduce.Job: map 100% reduce 100%
2022-06-21 10:10:10,534 INFO mapreduce.Job: Job job_1655785624817_0001 completed successfully
2022-06-21 10:10:10,623 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=95
  FILE: Number of bytes written=677637
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=419
  HDFS: Number of bytes written=40
  HDFS: Number of read operations=11
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

```

```

Total vcore-milliseconds taken by all reduce tasks=1506
Total megabyte-milliseconds taken by all map tasks=3632128
Total megabyte-milliseconds taken by all reduce tasks=1542144
Map-Reduce Framework
  Map input records=7
  Map output records=7
  Map output bytes=75
  Map output materialized bytes=101
  Input split bytes=188
  Combine input records=7
  Combine output records=7
  Reduce input groups=4
  Reduce shuffle bytes=101
  Reduce input records=7
  Reduce output records=4
  Spilled Records=14
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=150
  CPU time spent (ms)=1060
  Physical memory (bytes) snapshot=882737152
  Virtual memory (bytes) snapshot=7624015872
  Total committed heap usage (bytes)=879755264
  Peak Map Physical memory (bytes)=350687232
  Peak Map Virtual memory (bytes)=2544566272
  Peak Reduce Physical memory (bytes)=192282624
  Peak Reduce Virtual memory (bytes)=2543751168
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=231
File Output Format Counters

```

For output

\$ hdfs dfs -cat /output152/part*

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output152/part*
2022-06-21 10:10:46,188 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
1000 2
2000 2
3000 2
Amount withdrawn 1

```