1NT19IS108
CHAITANYA P
C1 BATCH

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name UserName Amount withdrawn Implement the following,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 program4;
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 read {

//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(read.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);
}
}
```

creating input directory.
hdfs dfs -mkdir -p /input108

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ 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]
Starting datanodes
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
Starting resourcemanager
Starting nodemanagers
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
3025 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
5493 Jps
4999 ResourceManager
4811 SecondaryNameNode
5148 NodeManager
4428 NameNode
4572 DataNode
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /input108
```

Coping the 108c.csv file from local to hdfs.
Hdfs dfs -copyFromLocal /home/hdoop/Desktop/108c.csv /input108

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hdoop/Desktop/108c.csv /input108
2022-06-21 09:58:26,435 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```

Hadoop jar /home/hdoop/Desktop/chaitanya.jar /input108 /output108
By using this 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/chaitanya.jar /input108 /output108
2022-06-21 10:00:39,278 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 10:00:39,406 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 10:00:39,539 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your appli
cation with ToolRunner to remedy this.
2022-06-21 10:00:39,563 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1655785365793_0001
2022-06-21 10:00:39,669 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:00:39,771 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-21 10:00:39,830 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:00:39,881 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:00:39,905 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-21 10:00:40,013 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-21 10:00:40,029 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655785365793_0001
2022-06-21 10:00:40,029 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-21 10:00:40,165 INFO conf.Configuration: resource-types.xml not found
2022-06-21 10:00:40,165 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-21 10:00:40,328 INFO impl.YarnClientImpl: Submitted application application_1655785365793_0001
2022-06-21 10:00:40,355 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655785365793_0001/
2022-06-21 10:00:40,356 INFO mapreduce.Job: Running job: job_1655785365793_0001
2022-06-21 10:00:45,422 INFO mapreduce.Job: Job job_1655785365793_0001 running in uber mode : false
2022-06-21 10:00:45,424 INFO mapreduce.Job:  map 0% reduce 0%
2022-06-21 10:00:49,480 INFO mapreduce.Job:  map 100% reduce 0%
2022-06-21 10:00:53,501 INFO mapreduce.Job:  map 100% reduce 100%
2022-06-21 10:00:54,518 INFO mapreduce.Job: Job job_1655785365793_0001 completed successfully
2022-06-21 10:00:54,573 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=101
                FILE: Number of bytes written=677550
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=434
                HDFS: Number of bytes written=67
                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
```

```
        Launched reduce tasks=1
        Data-local map tasks=2
        Total time spent by all maps in occupied slots (ms)=3543
        Total time spent by all reduces in occupied slots (ms)=1550
        Total time spent by all map tasks (ms)=3543
        Total time spent by all reduce tasks (ms)=1550
        Total vcore-milliseconds taken by all map tasks=3543
        Total vcore-milliseconds taken by all reduce tasks=1550
        Total megabyte-milliseconds taken by all map tasks=3628032
        Total megabyte-milliseconds taken by all reduce tasks=1587200
Map-Reduce Framework
        Map input records=8
        Map output records=8
        Map output bytes=91
        Map output materialized bytes=107
        Input split bytes=182
        Combine input records=8
        Combine output records=7
        Reduce input groups=7
        Reduce shuffle bytes=107
        Reduce input records=7
        Reduce output records=7
        Spilled Records=14
        Shuffled Maps =2
        Failed Shuffles=0
        Merged Map outputs=2
        GC time elapsed (ms)=173
        CPU time spent (ms)=1150
        Physical memory (bytes) snapshot=832196608
        Virtual memory (bytes) snapshot=7619448832
        Total committed heap usage (bytes)=851443712
        Peak Map Physical memory (bytes)=338276352
        Peak Map Virtual memory (bytes)=2536869888
        Peak Reduce Physical memory (bytes)=198430720
        Peak Reduce Virtual memory (bytes)=2548649984
Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
```

```
          Peak Reduce Virtual memory (bytes)=2548649984
    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=252
    File Output Format Counters
            Bytes Written=67
```

For output we use hdfs dfs -cat /output108/part*

```
                Bytes Written=67
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output108/part*
2022-06-21 10:01:34,054 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
11100   1
25000   1
30000   2
44000   1
45000   1
56000   1
Amount withdrawn        1
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$
```