Name: Phạm Đức Thể ID: 19522253 Class: DS200.M21

BIG DATA

SUMMARY

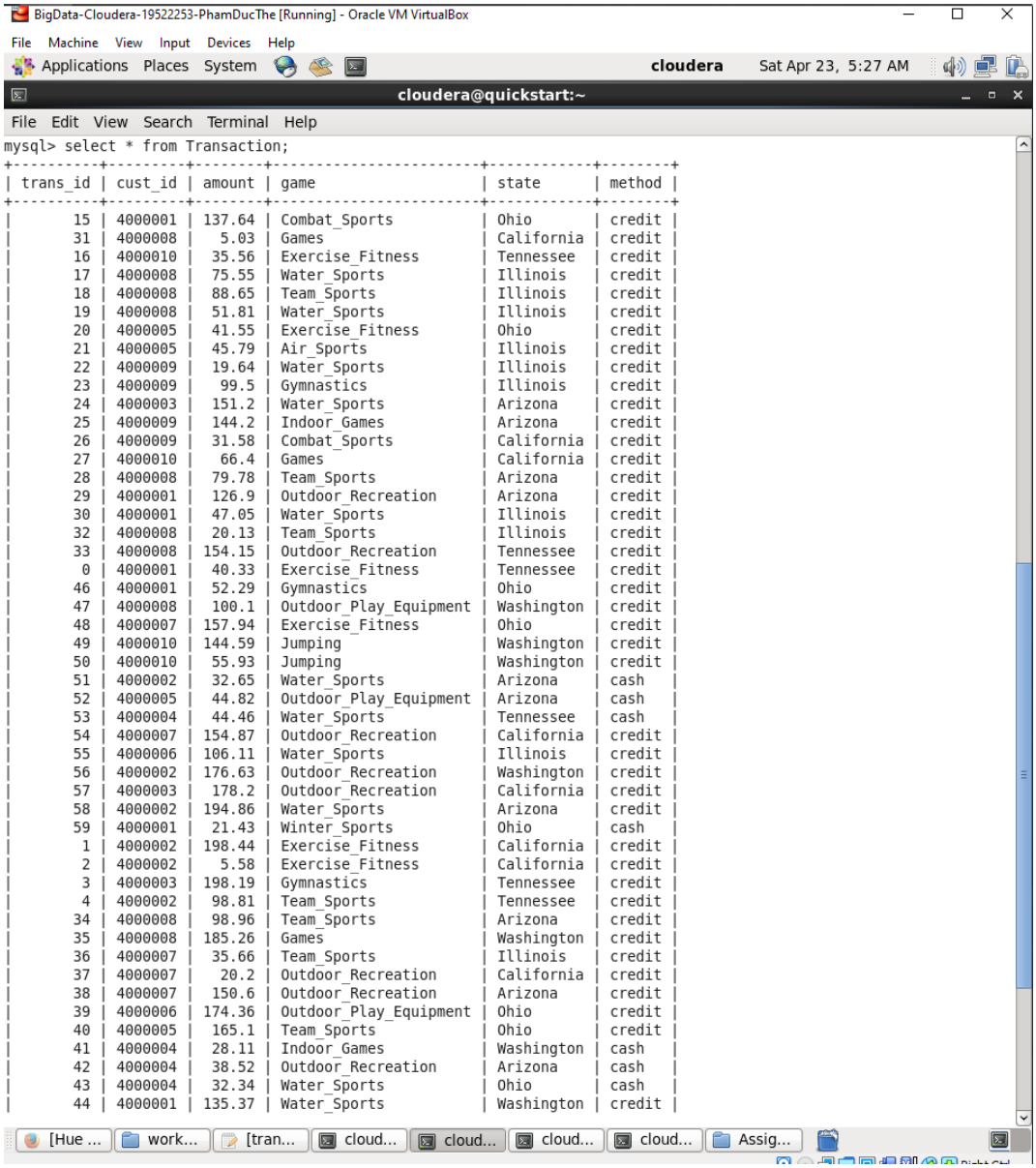
Task		Status	Page
Assignment 02	Task name 1	Hoàn thành	2
	Task name 2	Hoàn thành	4
	Task name 3	Hoàn thành	6
	Task name 4	Hoàn thành	7
	Task name 5	Hoàn thành	8
	Task name 6	Hoàn thành	10
	Task name 7	Hoàn thành	11
…	…		
	…		

Assignment 02

1. Task name 1: Load the content from trans.csv to this table using Sqoop and show the content of the table.

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/Assignment02 --username root --password clou
dera --table Transaction --export-dir /user/cloudera/trans.csv
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/04/23 05:21:14 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/04/23 05:21:14 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P
 instead.
22/04/23 05:21:14 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/04/23 05:21:14 INFO tool.CodeGenTool: Beginning code generation
22/04/23 05:21:16 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Transaction` AS t LIMIT 1
22/04/23 05:21:16 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Transaction` AS t LIMIT 1
22/04/23 05:21:16 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/ab7e9c86d44fd9fa818e1c88caff2b57/Transaction.java uses or overrides a deprecated
API.
Note: Recompile with -Xlint:deprecation for details.
22/04/23 05:21:22 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/ab7e9c86d44fd9fa818e1c
88caff2b57/Transaction.jar
22/04/23 05:21:23 INFO mapreduce.ExportJobBase: Beginning export of Transaction
22/04/23 05:21:23 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtrack
er.address
22/04/23 05:21:24 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/04/23 05:21:27 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead,
 use mapreduce.reduce.speculative
22/04/23 05:21:27 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, us
e mapreduce.map.speculative
22/04/23 05:21:27 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/04/23 05:21:28 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/04/23 05:21:30 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:21:31 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:21:31 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:21:31 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
```

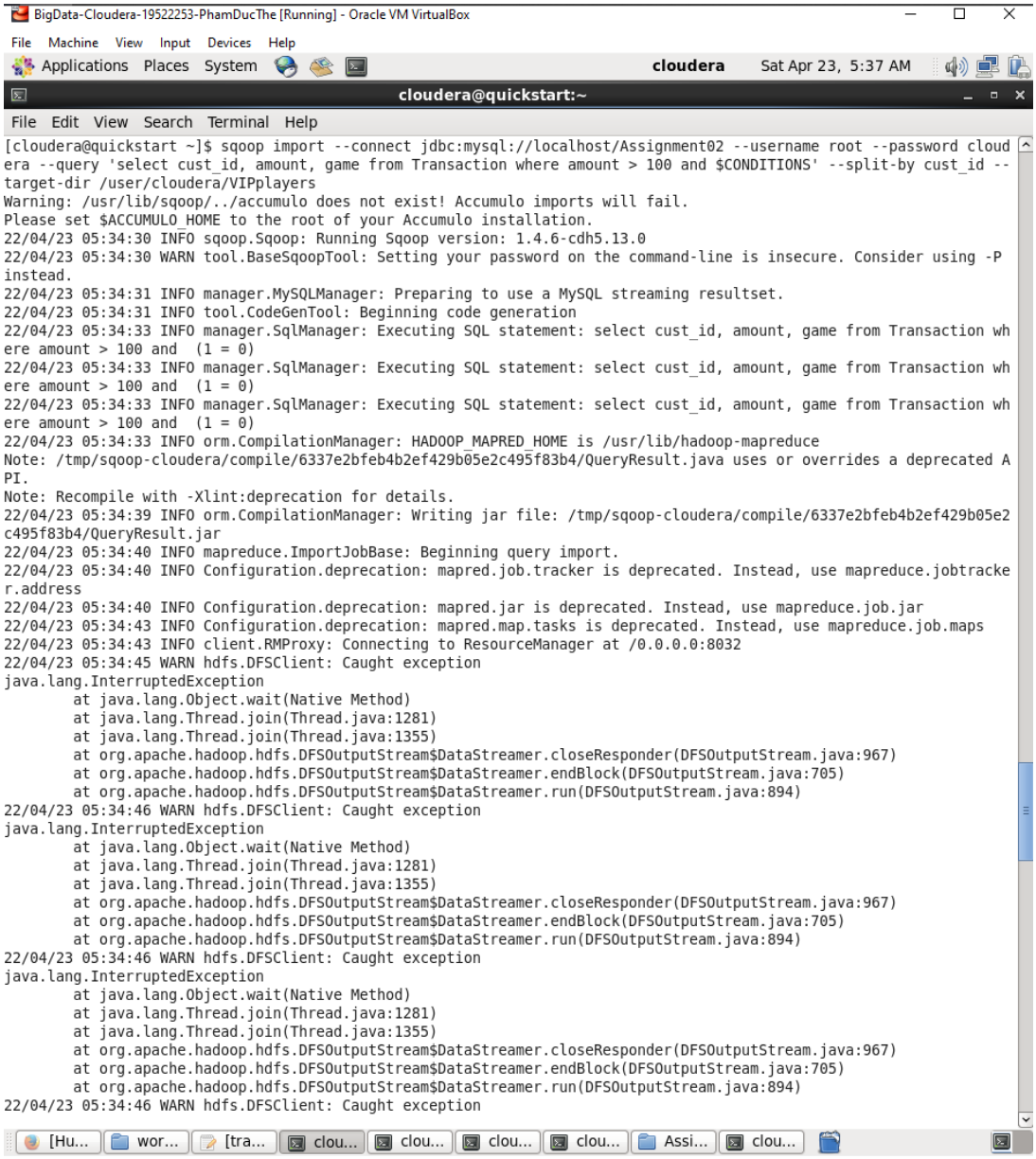
Figure 1: screenshot of sqoop command



```
mysql> select * from Transaction;
+----------+---------+--------+-----------------------+------------+--------+
| trans_id | cust_id | amount | game                  | state      | method |
+----------+---------+--------+-----------------------+------------+--------+
|       15 | 4000001 | 137.64 | Combat_Sports         | Ohio       | credit |
|       31 | 4000008 |   5.03 | Games                 | California | credit |
|       16 | 4000010 |  35.56 | Exercise_Fitness      | Tennessee  | credit |
|       17 | 4000008 |  75.55 | Water_Sports          | Illinois   | credit |
|       18 | 4000008 |  88.65 | Team_Sports           | Illinois   | credit |
|       19 | 4000008 |  51.81 | Water_Sports          | Illinois   | credit |
|       20 | 4000005 |  41.55 | Exercise_Fitness      | Ohio       | credit |
|       21 | 4000005 |  45.79 | Air_Sports            | Illinois   | credit |
|       22 | 4000009 |  19.64 | Water_Sports          | Illinois   | credit |
|       23 | 4000009 |   99.5 | Gymnastics            | Illinois   | credit |
|       24 | 4000003 |  151.2 | Water_Sports          | Arizona    | credit |
|       25 | 4000009 |  144.2 | Indoor_Games          | Arizona    | credit |
|       26 | 4000009 |  31.58 | Combat_Sports         | California | credit |
|       27 | 4000010 |   66.4 | Games                 | California | credit |
|       28 | 4000008 |  79.78 | Team_Sports           | Arizona    | credit |
|       29 | 4000001 |  126.9 | Outdoor_Recreation    | Arizona    | credit |
|       30 | 4000001 |  47.05 | Water_Sports          | Illinois   | credit |
|       32 | 4000008 |  20.13 | Team_Sports           | Illinois   | credit |
|       33 | 4000008 | 154.15 | Outdoor_Recreation    | Tennessee  | credit |
|        0 | 4000001 |  40.33 | Exercise_Fitness      | Tennessee  | credit |
|       46 | 4000001 |  52.29 | Gymnastics            | Ohio       | credit |
|       47 | 4000008 |  100.1 | Outdoor_Play_Equipment | Washington | credit |
|       48 | 4000007 | 157.94 | Exercise_Fitness      | Ohio       | credit |
|       49 | 4000010 | 144.59 | Jumping               | Washington | credit |
|       50 | 4000010 |  55.93 | Jumping               | Washington | credit |
|       51 | 4000002 |  32.65 | Water_Sports          | Arizona    | cash   |
|       52 | 4000005 |  44.82 | Outdoor_Play_Equipment | Arizona    | cash   |
|       53 | 4000004 |  44.46 | Water_Sports          | Tennessee  | cash   |
|       54 | 4000007 | 154.87 | Outdoor_Recreation    | California | credit |
|       55 | 4000006 | 106.11 | Water_Sports          | Illinois   | credit |
|       56 | 4000002 | 176.63 | Outdoor_Recreation    | Washington | credit |
|       57 | 4000003 |  178.2 | Outdoor_Recreation    | California | credit |
|       58 | 4000002 | 194.86 | Water_Sports          | Arizona    | credit |
|       59 | 4000001 |  21.43 | Winter_Sports         | Ohio       | cash   |
|        1 | 4000002 | 198.44 | Exercise_Fitness      | California | credit |
|        2 | 4000002 |   5.58 | Exercise_Fitness      | California | credit |
|        3 | 4000003 | 198.19 | Gymnastics            | Tennessee  | credit |
|        4 | 4000002 |  98.81 | Team_Sports           | Tennessee  | credit |
|       34 | 4000008 |  98.96 | Team_Sports           | Arizona    | credit |
|       35 | 4000008 | 185.26 | Games                 | Washington | credit |
|       36 | 4000007 |  35.66 | Team_Sports           | Illinois   | credit |
|       37 | 4000007 |   20.2 | Outdoor_Recreation    | California | credit |
|       38 | 4000007 |  150.6 | Outdoor_Recreation    | Arizona    | credit |
|       39 | 4000006 | 174.36 | Outdoor_Play_Equipment | Ohio       | credit |
|       40 | 4000005 |  165.1 | Team_Sports           | Ohio       | credit |
|       41 | 4000004 |  28.11 | Indoor_Games          | Washington | cash   |
|       42 | 4000004 |  38.52 | Outdoor_Recreation    | Arizona    | cash   |
|       43 | 4000004 |  32.34 | Water_Sports          | Ohio       | cash   |
|       44 | 4000001 | 135.37 | Water_Sports          | Washington | credit |
```

Figure 2: screenshot of mysql command to show the table content and the result

2. Task name 2: Use Sqoop to get cust_id, amount, game in Transaction table with amount > 100 and save results in VIPplayers directory. Then show the imported results. (1 point)



```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/Assignment02 --username root --password cloud
era --query 'select cust_id, amount, game from Transaction where amount > 100 and $CONDITIONS' --split-by cust_id --
target-dir /user/cloudera/VIPplayers
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/04/23 05:34:30 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/04/23 05:34:30 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P
instead.
22/04/23 05:34:31 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/04/23 05:34:31 INFO tool.CodeGenTool: Beginning code generation
22/04/23 05:34:33 INFO manager.SqlManager: Executing SQL statement: select cust_id, amount, game from Transaction wh
ere amount > 100 and  (1 = 0)
22/04/23 05:34:33 INFO manager.SqlManager: Executing SQL statement: select cust_id, amount, game from Transaction wh
ere amount > 100 and  (1 = 0)
22/04/23 05:34:33 INFO manager.SqlManager: Executing SQL statement: select cust_id, amount, game from Transaction wh
ere amount > 100 and  (1 = 0)
22/04/23 05:34:33 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/6337e2bfeb4b2ef429b05e2c495f83b4/QueryResult.java uses or overrides a deprecated A
PI.
Note: Recompile with -Xlint:deprecation for details.
22/04/23 05:34:39 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/6337e2bfeb4b2ef429b05e2
c495f83b4/QueryResult.jar
22/04/23 05:34:40 INFO mapreduce.ImportJobBase: Beginning query import.
22/04/23 05:34:40 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracke
r.address
22/04/23 05:34:40 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/04/23 05:34:43 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/04/23 05:34:43 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/04/23 05:34:45 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:34:46 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:34:46 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/04/23 05:34:46 WARN hdfs.DFSClient: Caught exception
```

Figure 3: screenshot of sqoop command

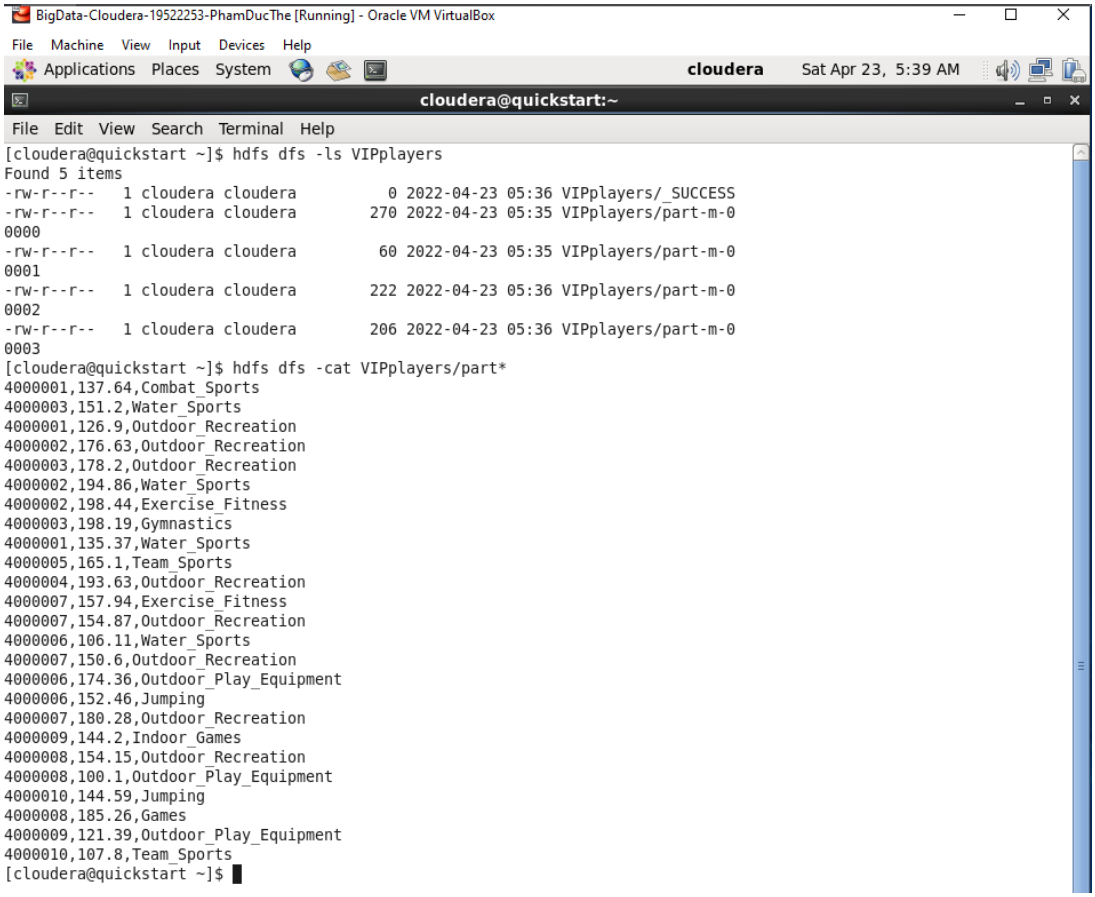


Figure 4: screenshot of hdfs dfs -cat command to show the imported results

3. Task name 3: Load data from file trans.csv to a Pig relation named PigTrans and show the result (1 point)

- Command: PigTrans = load 'trans.csv' using PigStorage(',') as (trans_id:int, cust_id:int, amount:float, game:chararray, state: chararray, method: chararray);
- Command: dump PigTrans

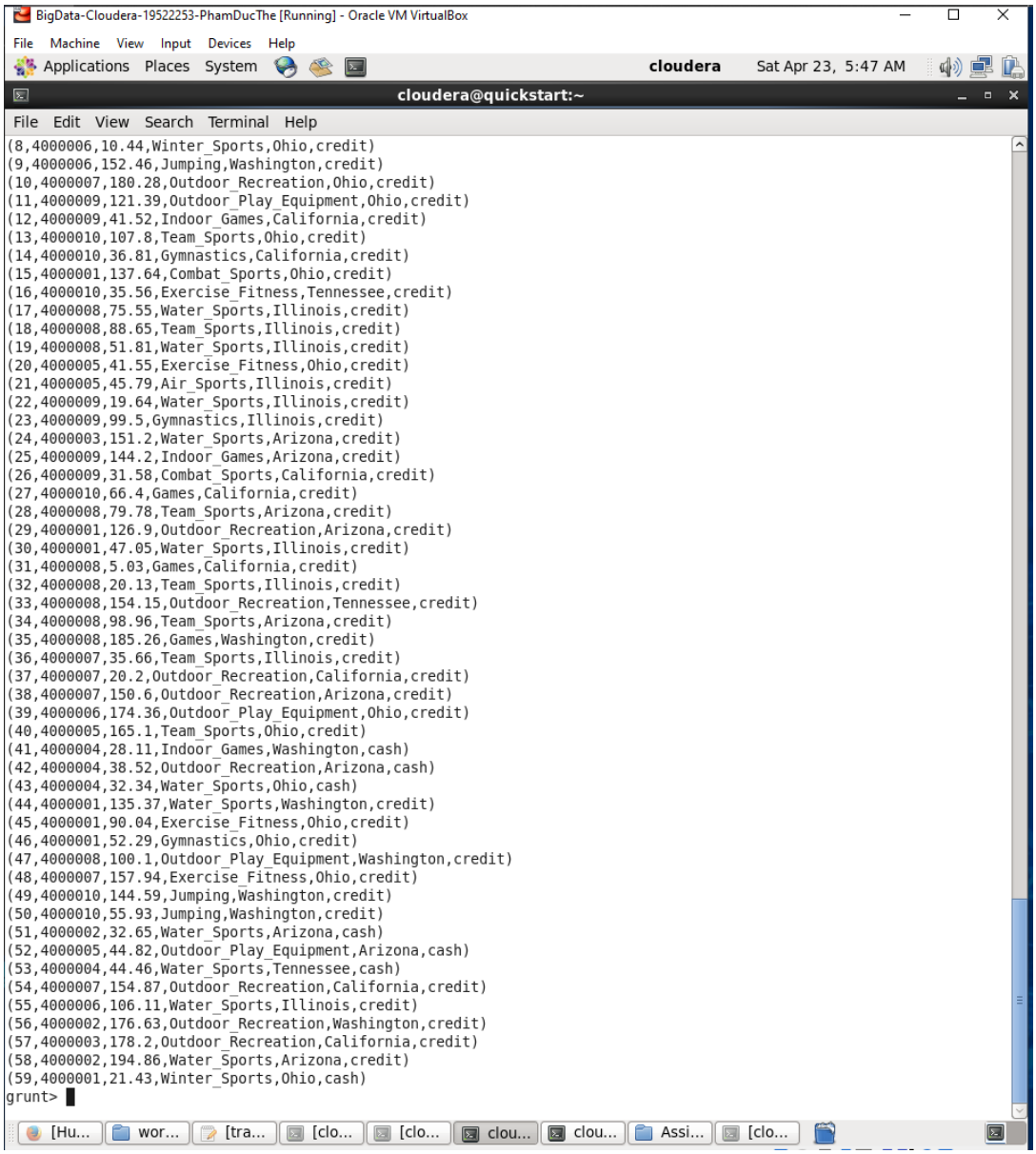


Figure 5: screenshot of command to show the Pig relation and the result

4. Task name 4

- Command: PigVIPplayers = filter PigTrans by amount > 100;
- Command: dump PigVIPplayers;

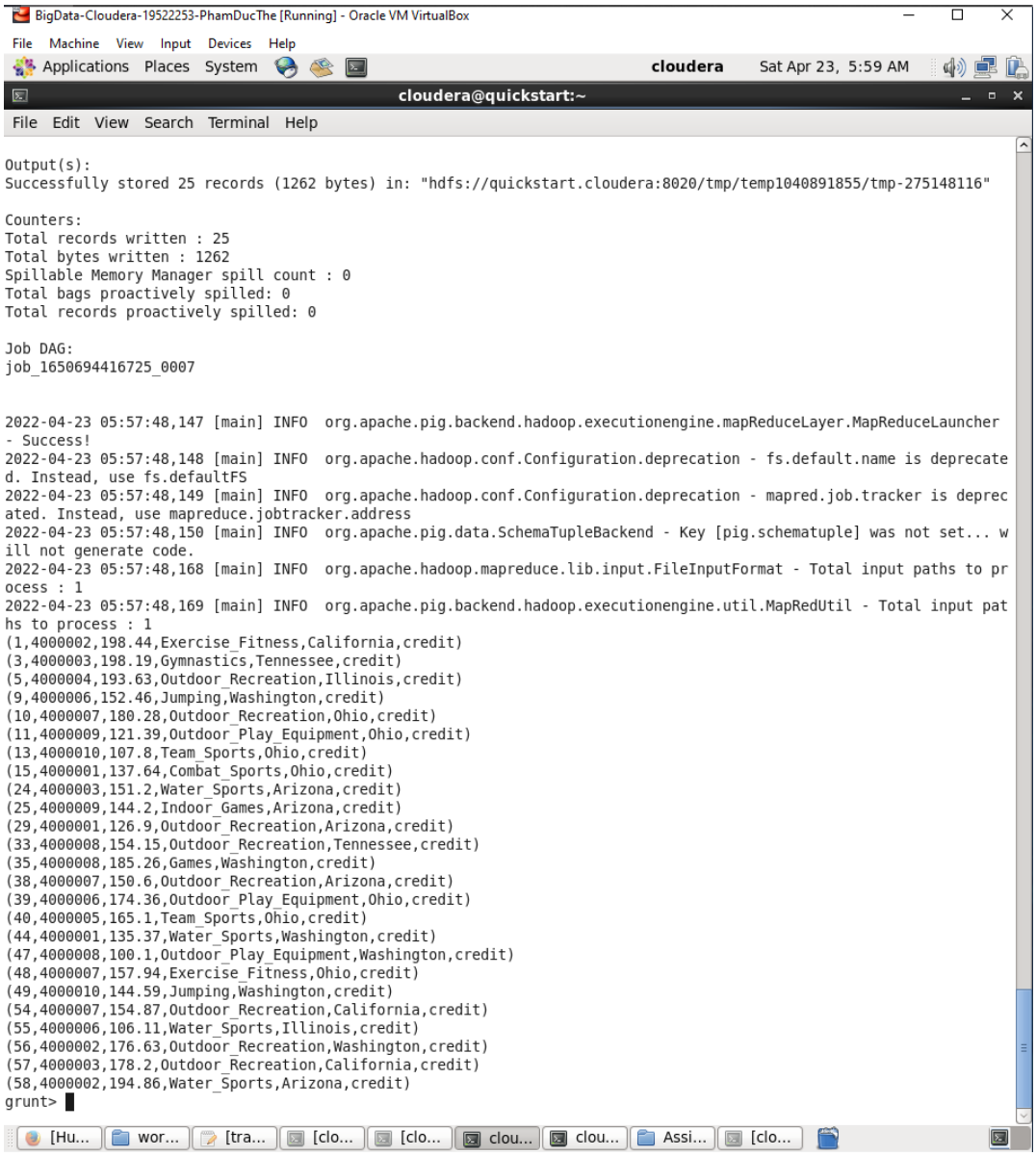
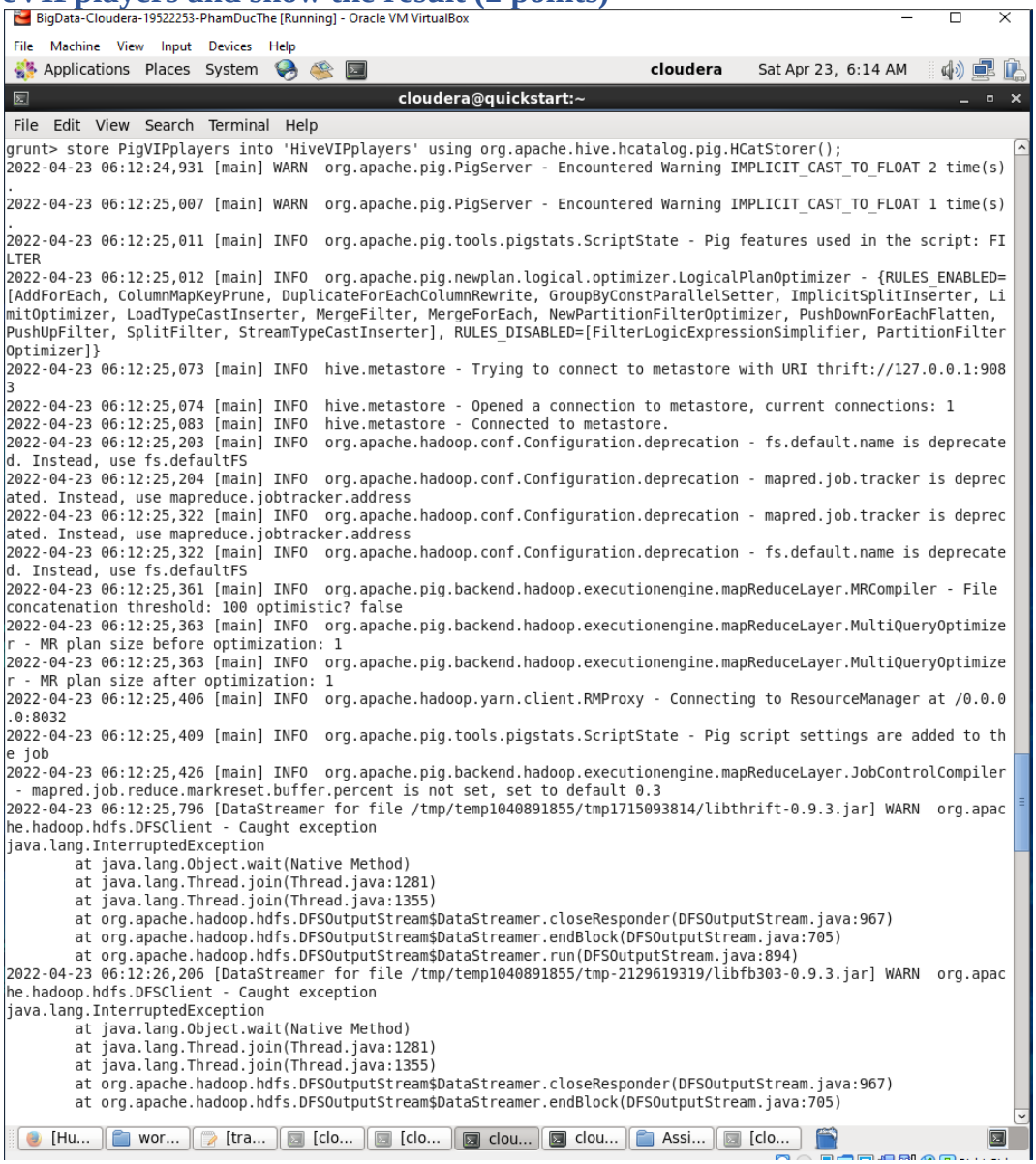


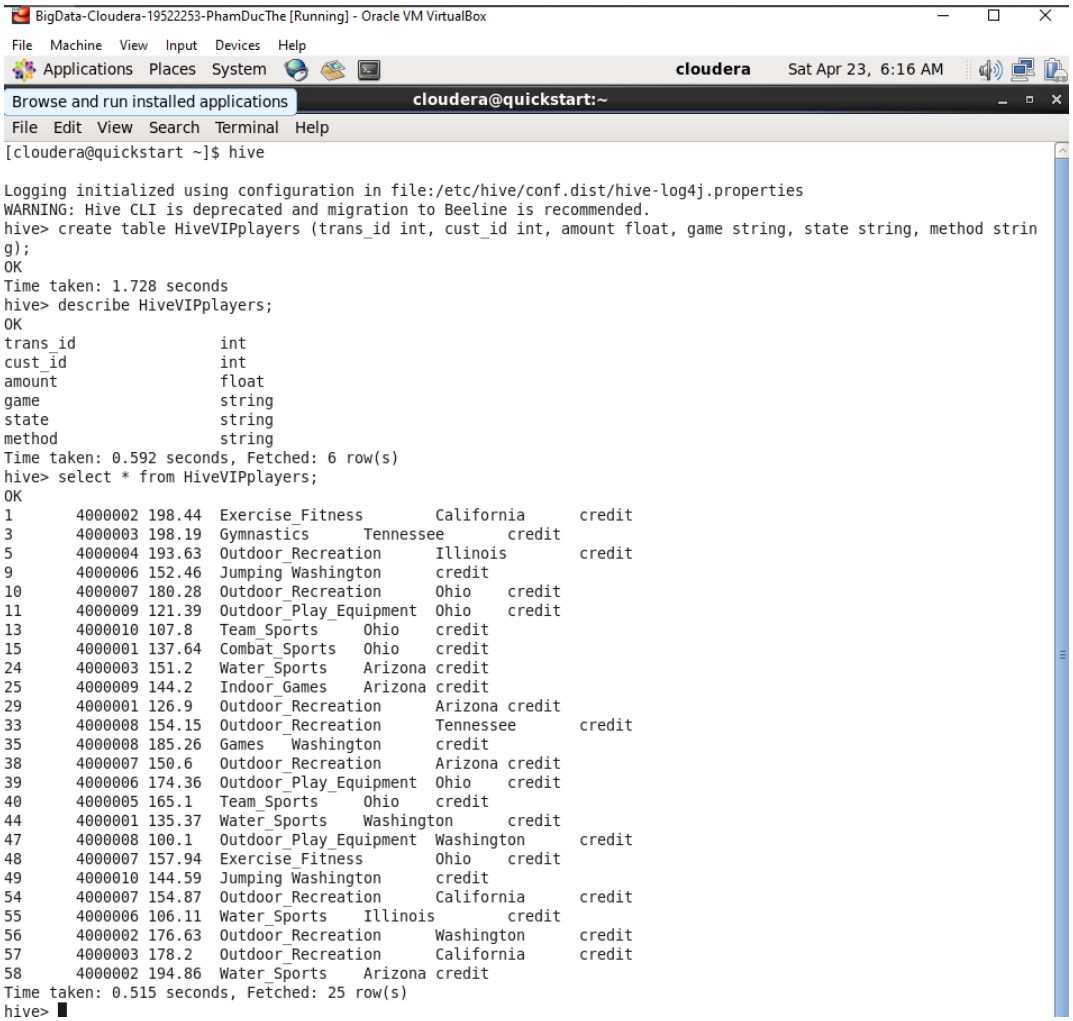
Figure 6: screenshot of command to show results the relation and the result

5. Task name 5: Store data in PigVIPplayers to a Hive table name HiveVIPplayers and show the result (2 points)



```
grunt> store PigVIPplayers into 'HiveVIPplayers' using org.apache.hive.hcatalog.pig.HCatStorer();
2022-04-23 06:12:24,931 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 2 time(s)
.
2022-04-23 06:12:25,007 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 1 time(s)
.
2022-04-23 06:12:25,011 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FI
LTER
2022-04-23 06:12:25,012 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=
[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, Li
mitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten,
PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilter
Optimizer]}
2022-04-23 06:12:25,073 [main] INFO  hive.metastore - Trying to connect to metastore with URI thrift://127.0.0.1:908
3
2022-04-23 06:12:25,074 [main] INFO  hive.metastore - Opened a connection to metastore, current connections: 1
2022-04-23 06:12:25,083 [main] INFO  hive.metastore - Connected to metastore.
2022-04-23 06:12:25,203 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecate
d. Instead, use fs.defaultFS
2022-04-23 06:12:25,204 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprec
ated. Instead, use mapreduce.jobtracker.address
2022-04-23 06:12:25,322 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprec
ated. Instead, use mapreduce.jobtracker.address
2022-04-23 06:12:25,322 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecate
d. Instead, use fs.defaultFS
2022-04-23 06:12:25,361 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File
concatenation threshold: 100 optimistic? false
2022-04-23 06:12:25,363 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimize
r - MR plan size before optimization: 1
2022-04-23 06:12:25,363 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimize
r - MR plan size after optimization: 1
2022-04-23 06:12:25,406 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0
.0:8032
2022-04-23 06:12:25,409 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to th
e job
2022-04-23 06:12:25,426 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler
 - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-23 06:12:25,796 [DataStreamer for file /tmp/temp1040891855/tmp1715093814/libthrift-0.9.3.jar] WARN  org.apac
he.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
2022-04-23 06:12:26,206 [DataStreamer for file /tmp/temp1040891855/tmp-2129619319/libfb303-0.9.3.jar] WARN  org.apac
he.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
```

Figure 7: screenshot of command to load data to hive table



```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table HiveVIPplayers (trans_id int, cust_id int, amount float, game string, state string, method strin
g);
OK
Time taken: 1.728 seconds
hive> describe HiveVIPplayers;
OK
trans_id                int
cust_id                 int
amount                  float
game                    string
state                   string
method                  string
Time taken: 0.592 seconds, Fetched: 6 row(s)
hive> select * from HiveVIPplayers;
OK
1       4000002 198.44  Exercise_Fitness        California      credit
3       4000003 198.19  Gymnastics      Tennessee       credit
5       4000004 193.63  Outdoor_Recreation      Illinois        credit
9       4000006 152.46  Jumping Washington      credit
10      4000007 180.28  Outdoor_Recreation      Ohio    credit
11      4000009 121.39  Outdoor_Play_Equipment  Ohio    credit
13      4000010 107.8   Team_Sports     Ohio    credit
15      4000001 137.64  Combat_Sports   Ohio    credit
24      4000003 151.2   Water_Sports    Arizona credit
25      4000009 144.2   Indoor_Games    Arizona credit
29      4000001 126.9   Outdoor_Recreation      Arizona credit
33      4000008 154.15  Outdoor_Recreation      Tennessee       credit
35      4000008 185.26  Games   Washington      credit
38      4000007 150.6   Outdoor_Recreation      Arizona credit
39      4000006 174.36  Outdoor_Play_Equipment  Ohio    credit
40      4000005 165.1   Team_Sports     Ohio    credit
44      4000001 135.37  Water_Sports    Washington      credit
47      4000008 100.1   Outdoor_Play_Equipment  Washington      credit
48      4000007 157.94  Exercise_Fitness        Ohio    credit
49      4000010 144.59  Jumping Washington      credit
54      4000007 154.87  Outdoor_Recreation      California      credit
55      4000006 106.11  Water_Sports    Illinois        credit
56      4000002 176.63  Outdoor_Recreation      Washington      credit
57      4000003 178.2   Outdoor_Recreation      California      credit
58      4000002 194.86  Water_Sports    Arizona credit
Time taken: 0.515 seconds, Fetched: 25 row(s)
hive>
```

Figure 8: screenshot of command to show result and the result

6. Task name 6: Load data from file trans.csv to table InternalTrans then show cust_id, amount, game with amount > 100.

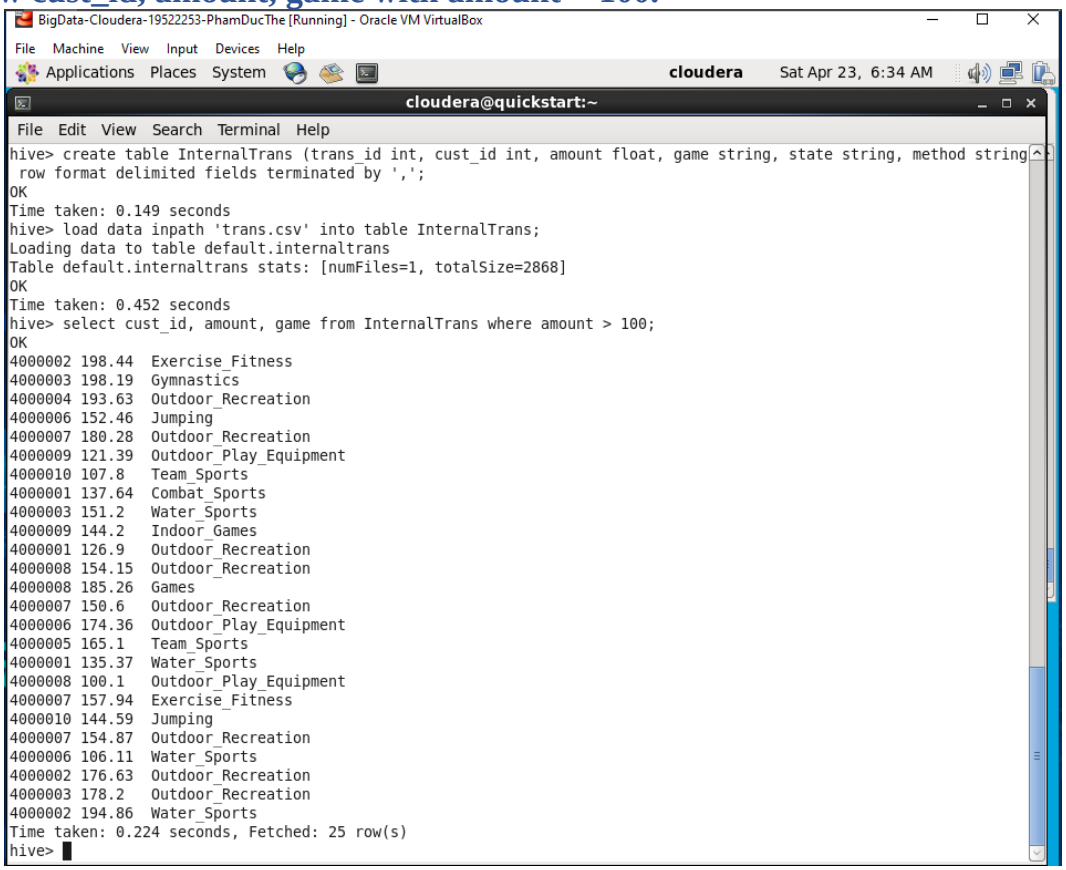
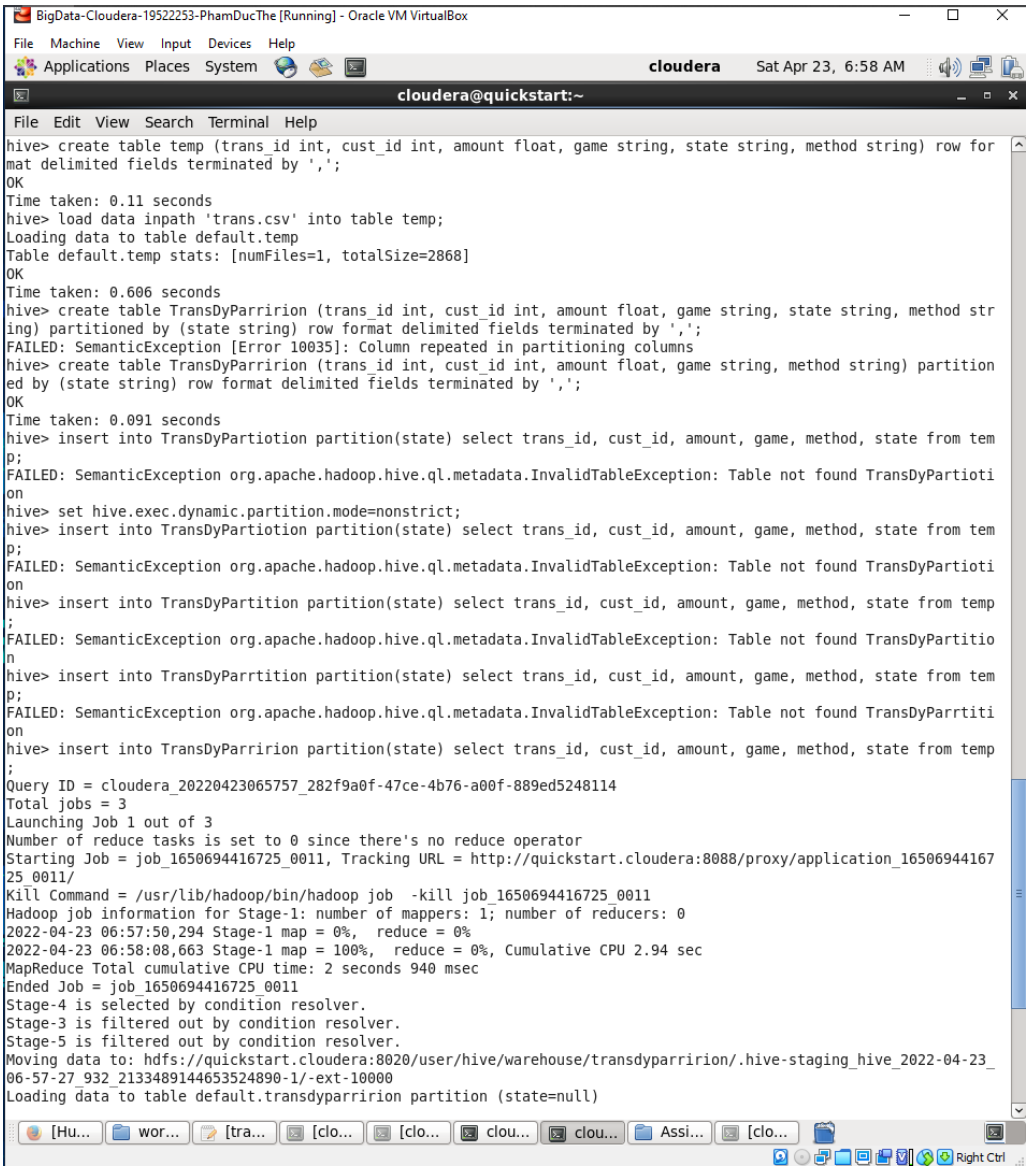


Figure 9: screenshot of Hive command and result

7. **Task name 7: Create a dynamic partition table named TransDyPartition to store the data from trans.csv file. This table is partitioned by state. (2 points)**
   - Load data to TransDyPartition table



```
hive> create table temp (trans_id int, cust_id int, amount float, game string, state string, method string) row for
mat delimited fields terminated by ',';
OK
Time taken: 0.11 seconds
hive> load data inpath 'trans.csv' into table temp;
Loading data to table default.temp
Table default.temp stats: [numFiles=1, totalSize=2868]
OK
Time taken: 0.606 seconds
hive> create table TransDyParririon (trans_id int, cust_id int, amount float, game string, state string, method str
ing) partitioned by (state string) row format delimited fields terminated by ',';
FAILED: SemanticException [Error 10035]: Column repeated in partitioning columns
hive> create table TransDyParririon (trans_id int, cust_id int, amount float, game string, method string) partition
ed by (state string) row format delimited fields terminated by ',';
OK
Time taken: 0.091 seconds
hive> insert into TransDyPartiotion partition(state) select trans_id, cust_id, amount, game, method, state from tem
p;
FAILED: SemanticException org.apache.hadoop.hive.ql.metadata.InvalidTableException: Table not found TransDyPartioti
on
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into TransDyPartiotion partition(state) select trans_id, cust_id, amount, game, method, state from tem
p;
FAILED: SemanticException org.apache.hadoop.hive.ql.metadata.InvalidTableException: Table not found TransDyPartioti
on
hive> insert into TransDyPartition partition(state) select trans_id, cust_id, amount, game, method, state from temp
;
FAILED: SemanticException org.apache.hadoop.hive.ql.metadata.InvalidTableException: Table not found TransDyPartitio
n
hive> insert into TransDyParrtition partition(state) select trans_id, cust_id, amount, game, method, state from tem
p;
FAILED: SemanticException org.apache.hadoop.hive.ql.metadata.InvalidTableException: Table not found TransDyParrtiti
on
hive> insert into TransDyParririon partition(state) select trans_id, cust_id, amount, game, method, state from temp
;
Query ID = cloudera_20220423065757_282f9a0f-47ce-4b76-a00f-889ed5248114
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1650694416725_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_16506944167
25_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1650694416725_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-04-23 06:57:50,294 Stage-1 map = 0%,  reduce = 0%
2022-04-23 06:58:08,663 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.94 sec
MapReduce Total cumulative CPU time: 2 seconds 940 msec
Ended Job = job_1650694416725_0011
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/transdyparririon/.hive-staging_hive_2022-04-23_
06-57-27_932_2133489144653524890-1/-ext-10000
Loading data to table default.transdyparririon partition (state=null)
```

Figure 10: the command to insert data to this table

– Open HUE and query to show all the rows of the table

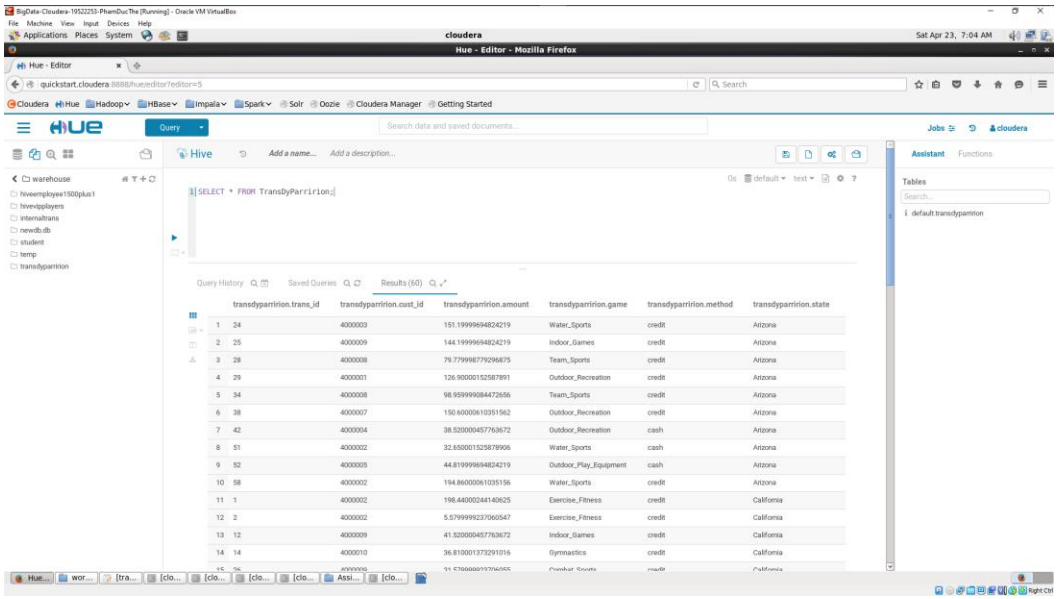


Figure 11: the screenshot of HUE with the query and the result

– …