|  |
| --- |
| 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

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

Text

Description automatically generated

Figure 1: screenshot of sqoop command

Graphical user interface, table

Description automatically generated

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

## 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)

Text

Description automatically generated

Figure 3: screenshot of sqoop command

Graphical user interface, text

Description automatically generated

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

## 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

Text

Description automatically generated

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

## Task name 4

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

Graphical user interface, text, application

Description automatically generated

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

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

Text

Description automatically generated

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

Text

Description automatically generated

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

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

Graphical user interface, text, application

Description automatically generated

Figure 9: screenshot of Hive command and result

## 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

Text

Description automatically generated

Figure 10: the command to insert data to this table

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

Graphical user interface

Description automatically generated

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

* …