Assignment 2: Sqoop Hive Pig

1. Login to **mysql** and create table **Transaction** with the following schema: (**2** points)

trans_id	cust_id	amount	game	state	method
integer	integer	float	varchar	varchar	varchar

Load the content from **trans.csv** to this table using **Sqoop** and show the content of the table.

Required Figure 1a: screenshot of Sqoop command

Required Figure 1b: screenshot of mysql command to show the table content and the result

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)

Required Figure 2a: screenshot of **sqoop** command

Required Figure 2b: screenshot of hdfs dfs -cat command to show the imported results

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

Required Figure 3: screenshot of command to show the Pig relation and the result

4. Create a **Pig** relation named **PigVIPplayers** to store info of transactions with **amount > 100** (**1** point)

Required Figure 4: screenshot of command to show results the relation and the result

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

Required Figure 5a: screenshot of command to load data to **hive** table

Required Figure 5b: screenshot of command to show result and the result

6. Create an internal table name **InternalTrans** in **Hive** with the schema (**1** point)

trans_id	cust_id	amount	game	state	method
int	int	float	string	string	string

Load data from file **trans.csv** to table **InternalTrans** then show cust_id, amount, game with amount > 100

Required Figure 6: **s**creenshot of Hive command and result

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

a. Load data to **TransDyPartition** table

Required Figure 7a: the command to insert data to this table

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

Required Figure 7b: the screenshot of **HUE** with the query and the result