Lab3 – Sqoop and Pig

1. Create table **Employee** inside database **EmployeeInfo** and load the content from **EmployeeNoheader.csv** to this table. Then query the created database to show the info of employees with **salary >1500**

```
+-------+------+--------+------------+-----+
| id    | name | salary | department | age |
+-------+------+--------+------------+-----+
| 20023 | Tran |   2000 | Product    |  24 |
| 20024 | Vu   |   3000 | Product    |  25 |
| 20025 | Dao  |   2500 | Testing    |  21 |
| 20026 | Nam  |   2500 | Support    |  22 |
| 20027 | Viet |   3000 | Product    |  24 |
+-------+------+--------+------------+-----+
```

2. Use Sqoop import to get id, name, and salary of employees in Employee table **with salary > 2000** and save results in **Employee2000plus** directory. Then show the imported results.

```
20024,Vu,3000.0
20025,Dao,2500.0
20026,Nam,2500.0
20027,Viet,3000.0
```

3. Load data from file **EmployeeNoheader.csv** to a relation named **PigEmployee** and show the result

4. Create a relation named **PigEmployee1500plus** to store employees with **salary > 1500**

5. Store data in **PigEmployee1500plus** to a Hive table name **HiveEmployee1500plus** and show the result

6. Load data in **EmployeeNoheader.csv** to a relation named **PigEmployeeNoschema** without declaring datatype. Then from this relation, create another relation named **BonusSalary** with all data is same as that in **PigEmployeeNoschema** but **salary = salary + 500**

7. Create an internal table name **InternalEmployee** with the schema

ID	int
Name	string
Salary	float
Department	string
Age	int

a. Load data from file **EmployeeInfo.csv** to **HDFS** to table InternalEmp

b. Login to HUE and query EmpName and EmpAge of all employees with salary-$\geq$ 2000

8. Create a dynamic partition table named **EmployeePartition** to store the data from **EmployeeInfo.csv** file. This table is partitioned by **Department**.

a. Load data to EmployeePartition table

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