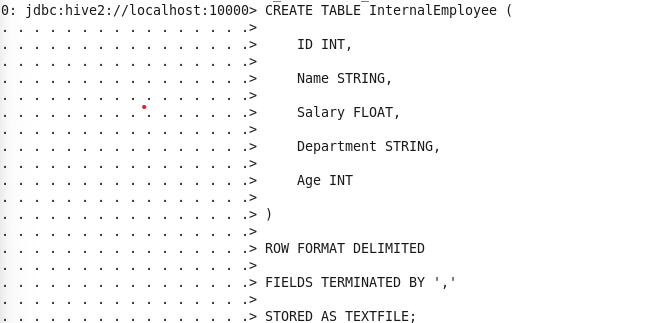
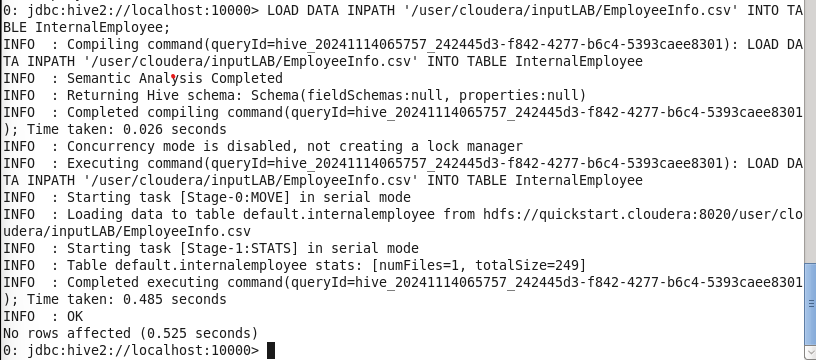
Big data analysis - Assignment 4 – Hive

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

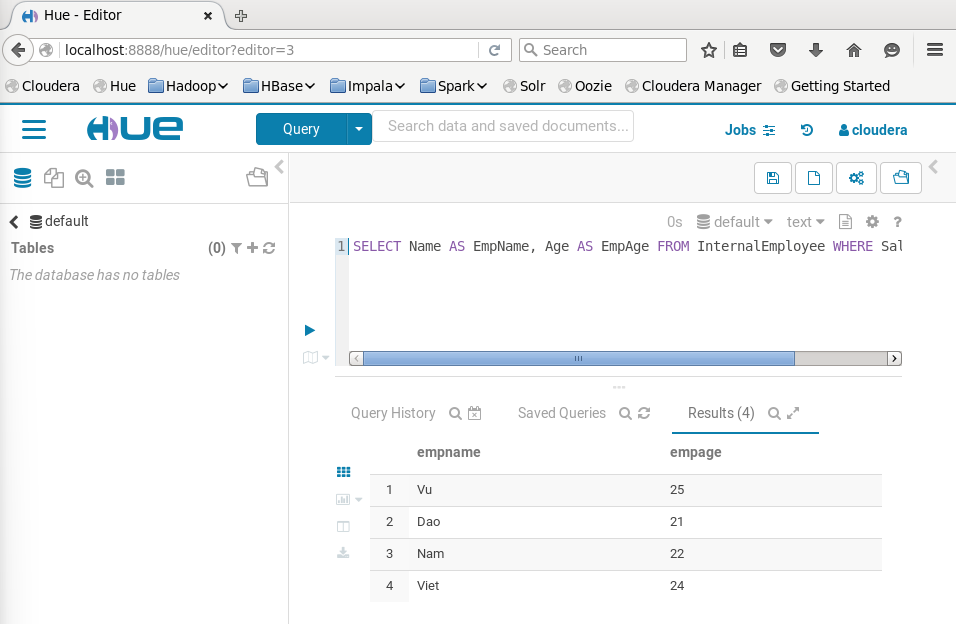
|  |  |
| --- | --- |
| ID | int |
| Name | string |
| Salary | float |
| Department | string |
| Age | int |



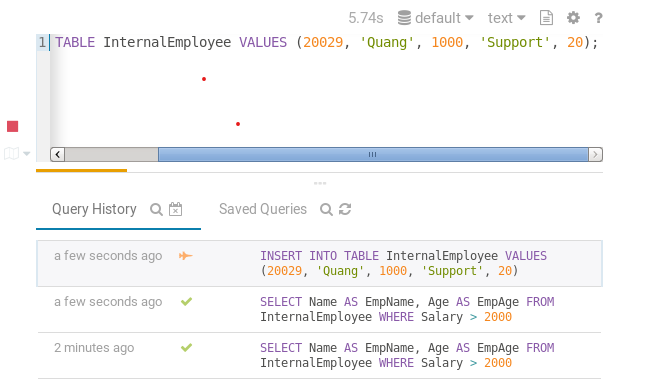
a. Put file **EmployeeInfo.csv** to **HDFS**, then load data from this file to table InternalEmp (**Figure 1a** should show the screenshot of loading command)



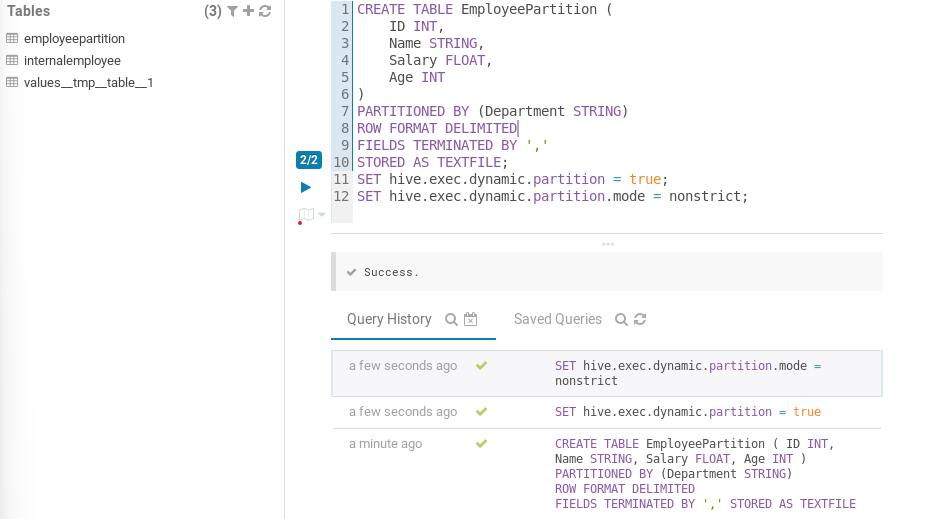
b. Login to HUE and query EmpName and EmpAge of all employees with salary age 2000 (**Figure 1b** should show the screenshot of the HUE page with both query and result)



c. In HUE, insert a row with the content [20029,Quang,1000,Support,20] to the table and then show the content of the newly created file which stores the data of the new row (**Figure 1c** should show the screenshot of the HUE page with the content of the newly created file)



2. 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 (**Figure 2a** should should show the command to insert data to this table)

A screenshot of a computer

Description automatically generated

b. Open HUE and query to show all the rows of the table (**Figure 2b** should show the screenshot of HUE with the query and the result)

