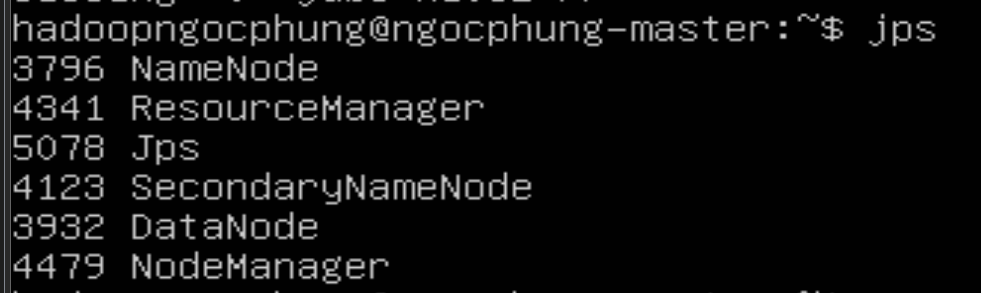
**APACHE HIVE 4.0.1 TUTORIAL**

# 1. Setup

## 1.1 Hadoop clusters

Hệ thống hadoop cluster đã được cài từ trước và đảm bảo chạy tốt

Máy Master



Máy Slave

A black background with white text

Description automatically generated

## 1.2 Chuẩn bị dữ liệu

### 1.2.1 HiveSample.txt

|  |
| --- |
| 1201 Gopal 45000 Technical manager  1202 Manisha 45000 Proof reader  1203 Masthanvali 40000 Technical writer  1204 Kiran 40000 Hr Admin  1205 Kranthi 30000 Op Admin |

### 1.2.2 Partition\_1.txt

|  |
| --- |
| 1, gopal, TP, 2012  2, kiran, HR, 2012  3, kaleel, SC, 2013  4, Prasanth, SC, 2013 |

### 1.2.3 Partition\_2.txt

|  |
| --- |
| 1, gopal, TP, 2012  2, kiran, HR, 2012 |

### 1.2.4 Partition\_3.txt

|  |
| --- |
| 3, kaleel, SC, 2013  4, Prasanth, SC, 2013 |

### 1.2.5 Join\_1.txt

|  |
| --- |
| 1,Ramesh,32,Ahmedabad,2000.00  2,Khilan,25,Delhi,1500.00  3,Kaushik,23,Kota,2000.00  4,Chaitali,25,Mumbai,6500.00  5,Hardik,27,Bhopal,8500.00  6,Komal,22,MP,4500.00  7,Muffy,24,Indore,10000.00 |

### 1.2.6 Join\_2.txt

|  |
| --- |
| 102,2009-10-08 00:00:00,3,3000  100,2009-10-08 00:00:00,3,1500  101,2009-11-20 00:00:00,2,1560  103,2008-05-20 00:00:00,4,2060 |

## 1.3 Share file vào máy ảo

Vào quyền root

|  |
| --- |
| mkdir /mnt/hgfs |
| /usr/bin/vmhgfs-fuse .host:/ /mnt/hgfs/ -o subtype=vmhgfs-fuse,allow\_other |

# 2. Khởi động hive

|  |
| --- |
| hive |

A screen shot of a computer program

Description automatically generated

|  |
| --- |
| !connect jdbc:hive2:// |

Sau đó nhập tên username và password

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

# 3. Apache Hive Tutorials

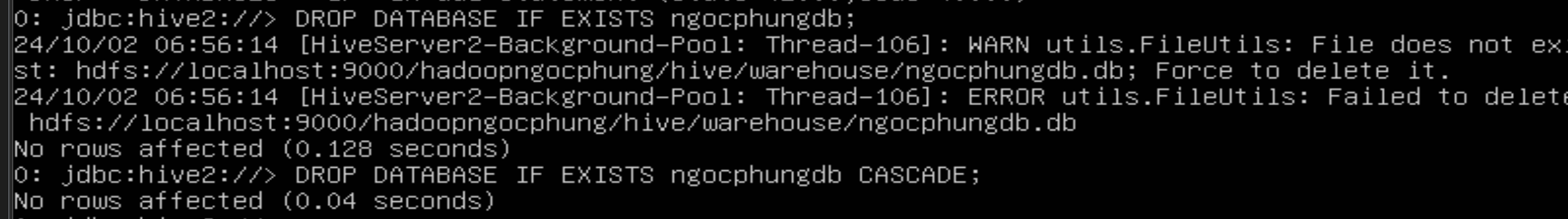
## 3.1 Create Database

|  |
| --- |
| CREATE SCHEMA ngocphungdb; |



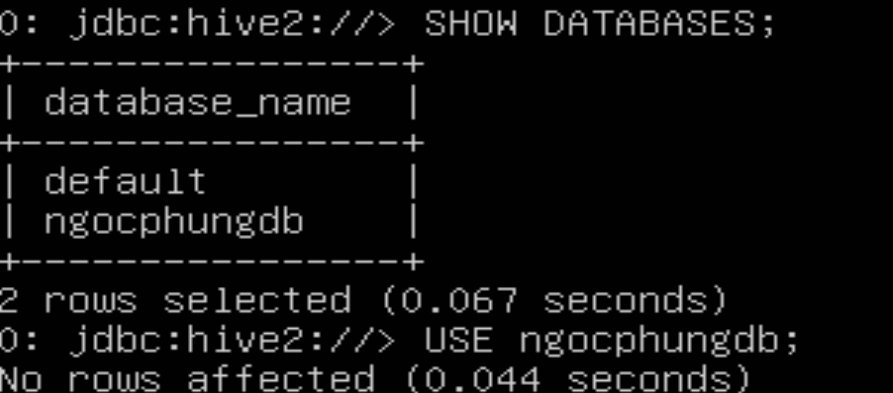
## 3.2 Drop Database

|  |
| --- |
| DROP DATABASE IF EXISTS ngocphungdb; |
| DROP DATABASE IF EXISTS ngocphungdb CASCADE; |

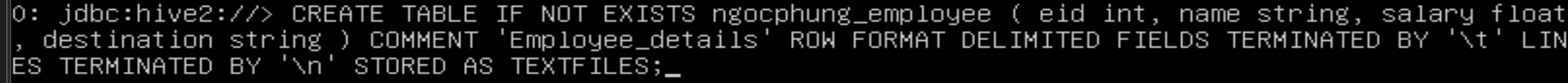


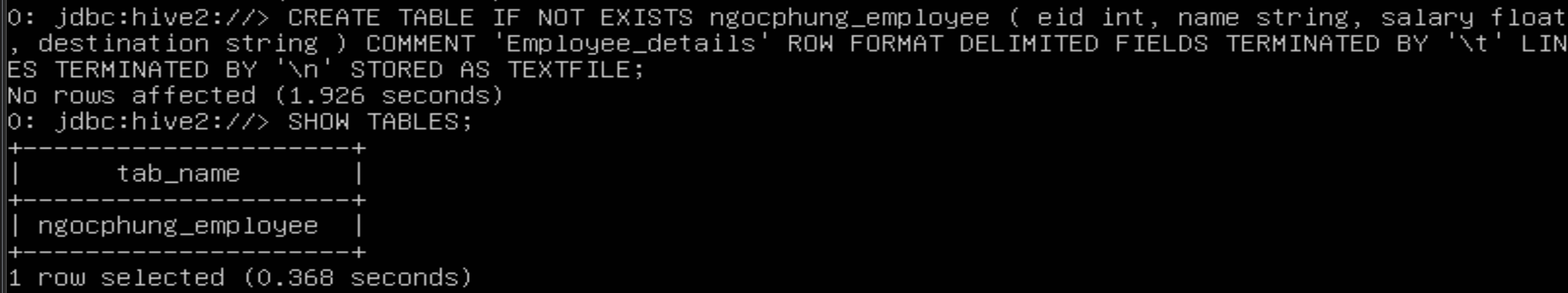
## 3.3 Create Table

|  |
| --- |
| SHOW DATABASES; |
| USE ngocphungdb; |



|  |
| --- |
| CREATE TABLE IF NOT EXISTS ngocphung\_employee ( eid int, name String, salary Float, destination String) COMMENT 'Employee details' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE; |





Load data

|  |
| --- |
| LOAD DATA LOCAL INPATH '/home/hadoopngocphung/Apache\_Hive/Hive\_Tutorial/HiveSample.txt' OVERWRITE INTO TABLE ngocphung\_employee; |

A screen shot of a computer

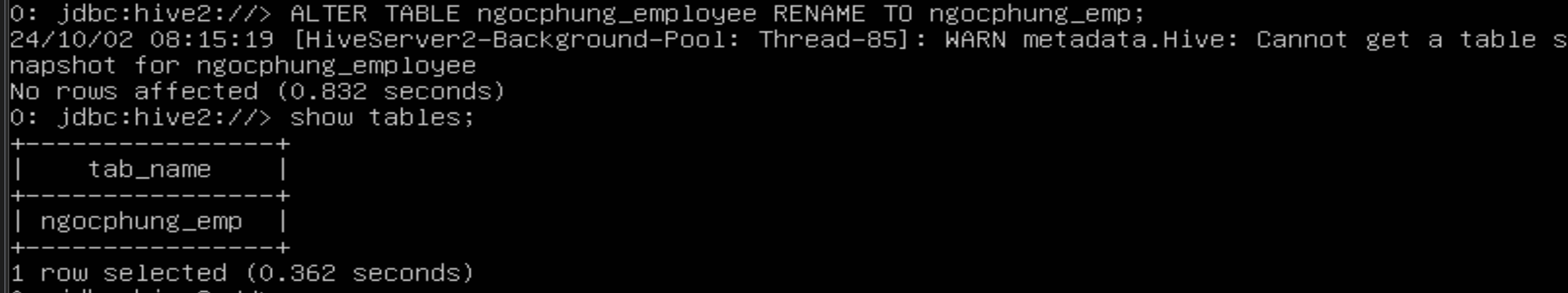
Description automatically generated

A screenshot of a computer

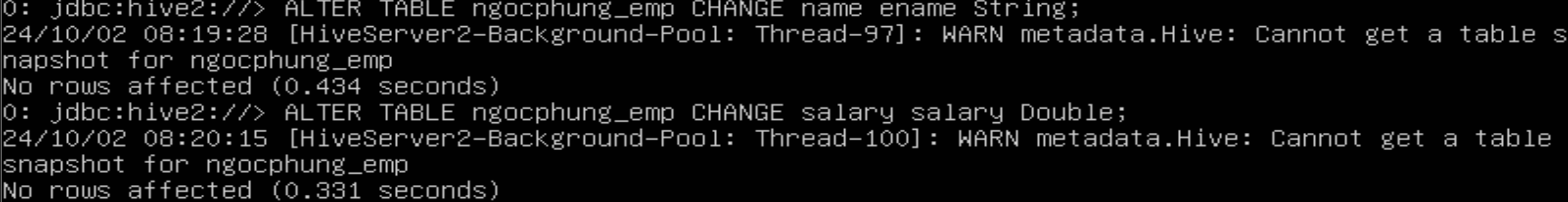
Description automatically generated

## 3.4 Alter Table

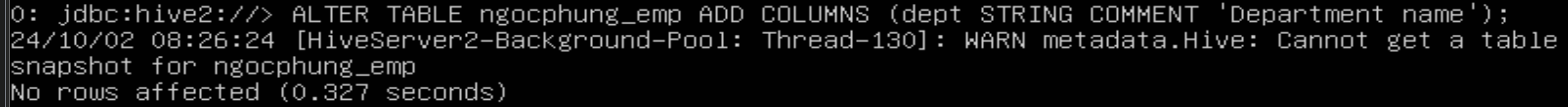
|  |
| --- |
| ALTER TABLE ngocphung\_employee RENAME TO ngocphung\_emp; |



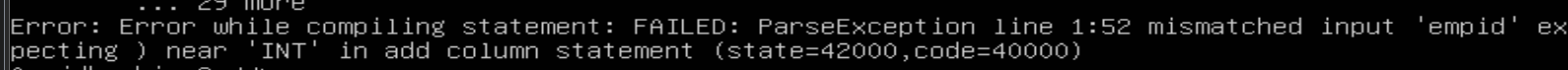
|  |
| --- |
| ALTER TABLE ngocphung\_emp CHANGE name ename String; |
| ALTER TABLE ngocphung\_emp CHANGE salary salary Double; |



|  |
| --- |
| ALTER TABLE ngocphung\_emp ADD COLUMNS (dept STRING COMMENT 'Department name'); |



|  |
| --- |
| ALTER TABLE ngocphung\_emp REPLACE COLUMNS ( eid INT empid Int, ename STRING name String); |



Hive không hỗ trợ thay đổi kiểu dữ liệu trực tiếp qua lệnh REPLACE COLUMNS nếu muốn có thể thực hiện các bước như thêm cột mới, sao chép dữ liệu, và đổi tên cột nếu cần.

## 3.5 Drop Table

|  |
| --- |
| DROP TABLE IF EXISTS ngocphung\_emp; |



A black screen with white text

Description automatically generated

## 3.6 Partitioning

Tạo file

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Tạo bảng

|  |
| --- |
| CREATE TABLE employee ( id INT, name STRING, dept STRING, yoj DOUBLE ) PARTITIONED BY (year INT); |

A screenshot of a computer

Description automatically generated

Thêm partition

|  |
| --- |
| ALTER TABLE employee ADD PARTITION (year='2012') location 'hive/tab1/employeedata/part2012'; |

A black screen with white text

Description automatically generated

Chỉnh sửa partition

|  |
| --- |
| ALTER TABLE employee PARTITION (year=2012) RENAME TO PARTITION (year=2013); |

A screen shot of a computer

Description automatically generated

Xóa partition

|  |
| --- |
| ALTER TABLE employee DROP PARTITION (year=2013); |

A screen shot of a computer

Description automatically generated

## 3.7 Built-in Operators

Tại lại và load dữ liệu trong bảng ngocphung\_employee

|  |
| --- |
| CREATE TABLE IF NOT EXISTS ngocphung\_employee ( eid INT, name STRING, salary FLOAT, destination STRING, dept STRING) COMMENT 'Employee details' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE; |
| LOAD DATA LOCAL INPATH '/home/hadoopngocphung/Apache\_Hive/Hive\_Tutorial/HiveSample.txt' OVERWRITE INTO TABLE ngocphung\_employee; |
| SELECT \* FROM ngocphung\_employee; |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| SELECT \* FROM ngocphung\_employee WHERE eid=1202; |

A screenshot of a computer program

Description automatically generated

|  |
| --- |
| SELECT \* FROM ngocphung\_employee WHERE salary>40000; |

A black screen with white text

Description automatically generated

Arithmetic Operators

|  |
| --- |
| SELECT 20+30; |

A screenshot of a computer

Description automatically generated

Logical Operators

|  |
| --- |
| SELECT \* FROM ngocphung\_employee WHERE salary>40000 AND dept='manager'; |

A screenshot of a computer

Description automatically generated

## 3.8 Built-in Function

|  |
| --- |
| SELECT round(2.6); |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| SELECT floor(2.6); |

A screenshot of a computer

Description automatically generated

|  |
| --- |
| SELECT ceil(2.6); |

A screen shot of a computer

Description automatically generated

## 3.9 View and Indexes

Tạo view chỉ hiện ra các nhân viên có lương trên 30000;

|  |
| --- |
| CREATE VIEW emp\_30000 AS SELECT \* FROM ngocphung\_employee WHERE salary>30000; |
| SELECT \* FROM emp\_30000; |

A black screen with white text

Description automatically generated

|  |
| --- |
| CREATE INDEX inedx\_salary ON TABLE ngocphung\_employee(salary) AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'; |

A black and white screen

Description automatically generated

Trong phiên bản Hive 4.0.1, tính năng indexing (tạo chỉ số) đã không còn được khuyến khích và không được sử dụng rộng rãi do hiệu suất không cao

# 4. HiveQL

## 4.1 Where

|  |
| --- |
| SELECT \* FROM ngocphung\_employee WHERE salary>30000 |

A black screen with white text

Description automatically generated

## 4.2 Order By

|  |
| --- |
| SELECT eid, name, dept FROM ngocphung\_employee ORDER BY dept; |

A computer screen with white text

Description automatically generated

## 4.3 Group By

|  |
| --- |
| SELECT dept,count(\*) FROM ngocphung\_employee GROUP BY dept; |

A black screen with white text

Description automatically generated

## 2.4 Joins

Nội dung file Join\_1.txt và Join\_2.txt đã chuẩn bị

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

- Tạo 2 bảng customers và orders

|  |
| --- |
| CREATE TABLE ngocphung\_customers (id INT, name STRING, age INT, address STRING, salary DOUBLE) COMMENT 'Customers Details' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE; |
| CREATE TABLE ngocphung\_orders (oid INT, o\_date TIMESTAMP, customer\_id INT, amount DOUBLE) COMMENT 'Orders Details' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE; |
| DESCRIBE ngocphung\_customers; |
| DESCRIBE ngocphung\_orders; |

A computer screen shot of a computer

Description automatically generated

Load dữ liệu

|  |
| --- |
| LOAD DATA LOCAL INPATH '/home/hadoopngocphung/Apache\_Hive/Hive\_Tutorial/Join\_1.txt' OVERWRITE INTO TABLE ngocphung\_customers; |
| LOAD DATA LOCAL INPATH '/home/hadoopngocphung/Apache\_Hive/Hive\_Tutorial/Join\_2.txt' OVERWRITE INTO TABLE ngocphung\_orders; |
| DESCRIBE ngocphung\_customers; |
| DESCRIBE ngocphung\_orders; |

A black screen with white text

Description automatically generated

A screenshot of a computer

Description automatically generated

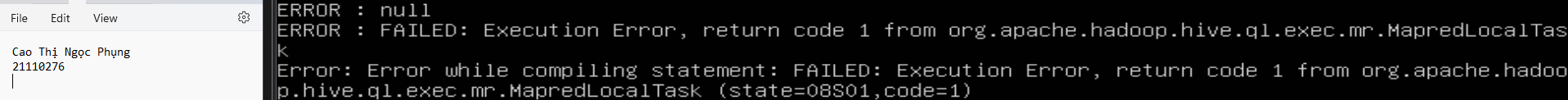
A screenshot of a computer

Description automatically generated

|  |
| --- |
| SELECT c.id, c.name, c.age, o.amount FROM ngocphung\_customers c JOIN ngocphung\_orders o ON (c.id = o.customer\_id); |

A black and white text

Description automatically generated

đã xảy ra lỗi 

thử tăng giới hạn bộ nhớ cho các tác vụ của Hive bằng cách cấu hình lại tham số trong Hive:

|  |
| --- |
| SET hive.auto.convert.join=false; |



Thử chạy lệnh join trên

A screen shot of a computer

Description automatically generated

LEFT OUTER JOIN:

|  |
| --- |
| SELECT c.id, c.name, c.age, o.amount FROM ngocphung\_customers c LEFT OUTER JOIN ngocphung\_orders o ON (c.id = o.customer\_id); |

A screenshot of a computer

Description automatically generated

RIGHT OUTER JOIN:

|  |
| --- |
| SELECT c.id, c.name, c.age, o.amount FROM ngocphung\_customers c RIGHT OUTER JOIN ngocphung\_orders o ON (c.id = o.customer\_id); |

A black screen with white text

Description automatically generated

FULL OUTER JOIN

|  |
| --- |
| SELECT c.id, c.name, c.age, o.amount FROM ngocphung\_customers c FULL OUTER JOIN ngocphung\_orders o ON (c.id = o.customer\_id); |

A screenshot of a computer

Description automatically generated