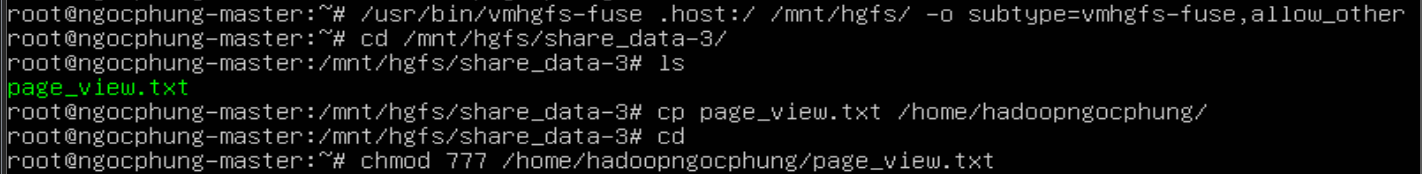
# 1. Chuẩn bị dữ liệu

File ‘page\_view.txt’ chứa nội dung sau

|  |
| --- |
| 1391102617,1234,http://quicloud.com/,http://google.com/,198.211.110.9  1391101111,1234,http://somewhere.com/,http://google.com/,197.211.110.9  1391002637,1234,http://somwhere-else.com/,http://yahoo.com/,196.211.110.9  1391002617,1234,http://quicloud.com/,http://bing.com/,198.211.110.9 |

Share file vào máy ảo

|  |
| --- |
| mkdir /mnt/hgfs |
| /usr/bin/vmhgfs-fuse .host:/ /mnt/hgfs/ -o subtype=vmhgfs-fuse,allow\_other |
| cd /mnt/hgfs/share\_data-3/ |



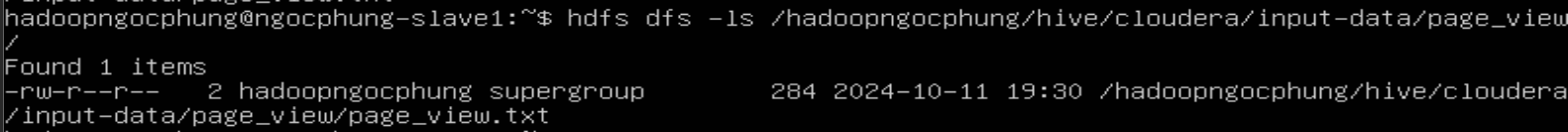
A computer screen with blue text

Description automatically generated

Đặt file vào hdfs

|  |
| --- |
| hdfs dfs -put /Apache\_Hive/Page\_View/page\_view.txt /hadoopngocphung/hive/cloudera/input-data/page\_view |





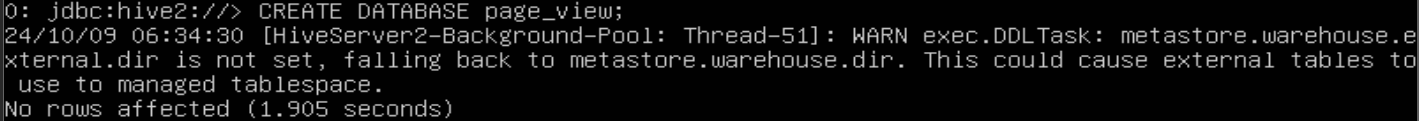
# 2. Thực hành

Kết nối với hive

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

Tạo bảng page\_view

|  |
| --- |
| CREATE DATABASE page\_view; |



A screen shot of a computer

Description automatically generated

Tạo một bảng ngoại vi trong Hive để lưu trữ dữ liệu

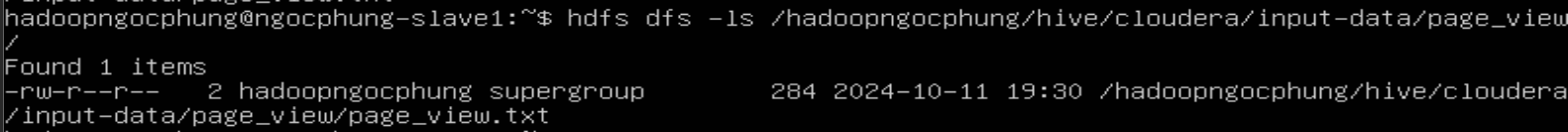
|  |
| --- |
| CREATE EXTERNAL TABLE page\_view\_external (viewTime INT, userid BIGINT, page\_url STRING, referrer\_url STRING, ip STRING) COMMENT 'This is the EXTERNAL page view table' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LOCATION '/hadoopngocphung/hive/cloudera/input-data/page\_view/'; |
| DESC page\_view\_external; |

A black screen with white text

Description automatically generated

Các dữ liệu lấy từ các tệp trong thư mục /hadoopngocphung/hive/cloudera/input-data/page\_view/ trên HDFS. Dữ liệu trong thư mục này sẽ không bị xóa khi bảng page\_view\_external bị xóa.

Trước đó chúng ta đã chuẩn bị tệp page\_view.txt lưu ở địa chỉ /hadoopngocphung/hive/cloudera/input-data/page\_view/ vì vậy database sẽ đọc tệp này và tự nạp dữ liệu vào bảng external



|  |
| --- |
| DESC EXTENDED page\_view\_external; |

A screenshot of a computer

Description automatically generated

Khi liệt kê tất cả dữ liệu của bảng page\_view\_external ra, ta có thể thấy tất cả dữ liệu trong file page\_view.txt mặc dù trước đó không có bất kì thao tác load dữ liệu nào

|  |
| --- |
| SELECT \* FROM page\_view\_external; |

Liệt kê những user đã truy cập page có kí tự quicloud

|  |
| --- |
| SELECT \* FROM page\_view\_external WHERE page\_url LIKE '%quicloud%'; |

A black screen with white text

Description automatically generated

Chỉ liệt kê page\_url nếu có người truy cập page có kí tự quicloud

|  |
| --- |
| SELECT DISTINCT(page\_url) FROM page\_view\_external WHERE page\_url LIKE '%quicloud%'; |

A black screen with white text

Description automatically generated

Tạo bảng nội trong hive

|  |
| --- |
| CREATE TABLE page\_view\_internal(viewTime INT, userid BIGINT,page\_url STRING, referrer\_url STRING,ip STRING) COMMENT 'This is the INTERNAL page view table' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','; |
| DESC page\_view\_internal; |

A black screen with white text

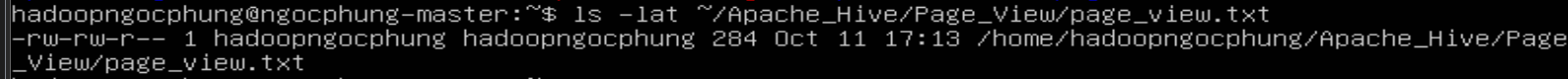
Description automatically generated

|  |
| --- |
| LOAD DATA LOCAL INPATH '/home/hadoopngocphung/Apache\_Hive/Page\_View/page\_view.txt' OVERWRITE INTO TABLE page\_view\_internal; |
| SELECT \* FROM page\_view\_internal; |

A screenshot of a computer program

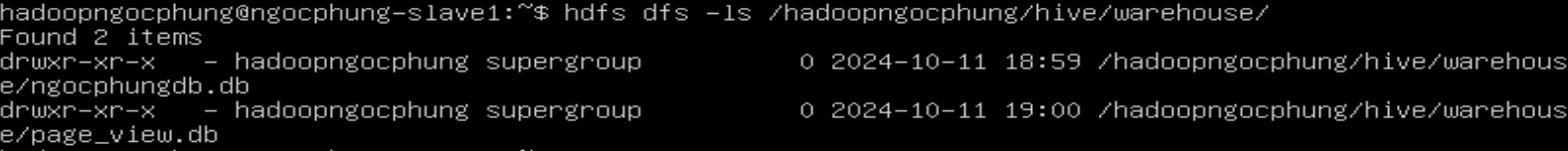
Description automatically generated

|  |
| --- |
| ls -lat ~/Apache\_Hive/Page\_View/page\_view.txt |

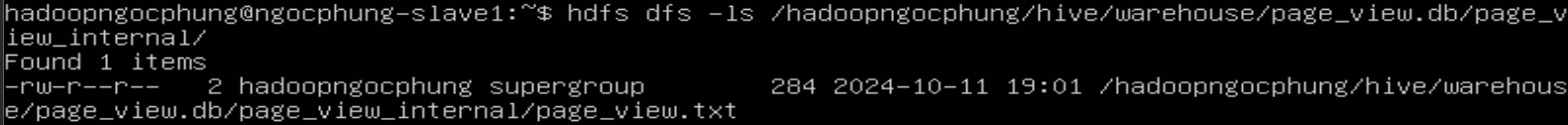


Dữ liệu lưu trong file page\_view.txt lưu ở 2 bảng nội và ngoại ở 2 địa chỉ khác nhau, bảng nội được lưu trong data warehouse của hive [/hadoopngocphung/hive/warehouse/page\_view.db/page\_view\_internal/] trong khi bảng ngoại thì lưu ở chổ tùy chỉnh [/hadoopngocphung/hive/cloudera/input-data/page\_view/]

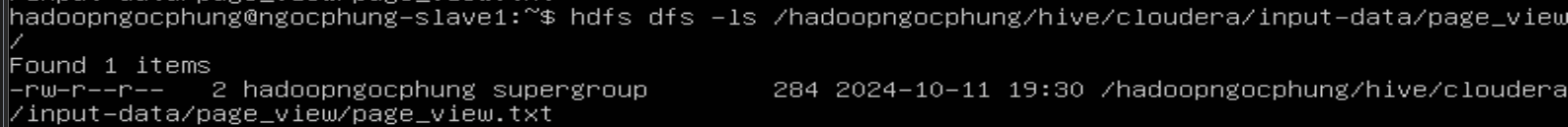
|  |
| --- |
| hdfs dfs -ls /hadoopngocphung/hive/warehouse |



|  |
| --- |
| hdfs dfs -ls /hadoopngocphung/hive/warehouse/page\_view.db/page\_view\_internal/ |



|  |
| --- |
| hdfs dfs -ls /hadoopngocphung/hive/cloudera/input-data/page\_view/ |

****

|  |
| --- |
| DESC EXTENDED page\_view\_internal; |

A computer screen shot of a computer screen

Description automatically generated

Xóa bảng và so sánh dữ liệu có bị mất k

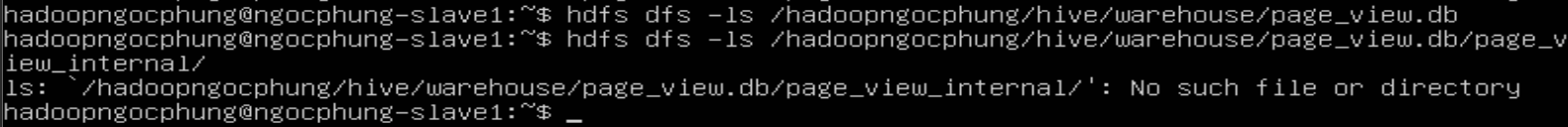
|  |
| --- |
| SHOW TABLES; |
| DROP TABLE page\_view\_internal; |
| DROP TABLE page\_view\_external; |
| SHOW TABLES; |

A screenshot of a computer

Description automatically generated

Tìm đến địa chỉ lưu bảng page\_view\_internal

|  |
| --- |
| hdfs dfs -ls /hadoopngocphung/hive/warehouse/page\_view.db/ |
| hdfs dfs -ls /hadoopngocphung/hive/warehouse/page\_view.db/page\_view\_internal/ |



Do bảng đã bị hive xóa nên ta không thể thấy địa chỉ 2 bản và vào đúng địa chỉ của bảng để xem bất kì dữ liệu nào

Tuy nhiên, dù bảng đã bị hive xóa nhưng địa chỉ lấy dữ liệu không được quản lý bởi hive nên dữ liệu không bị xóa

|  |
| --- |
| hdfs dfs -ls /hadoopngocphung/hive/cloudera/input-data/page\_view/ |

