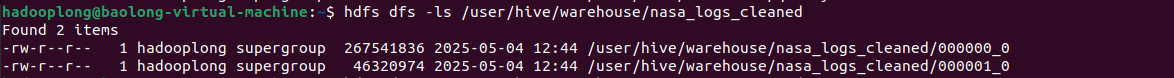
**SAO LƯU VÀ KHÔI PHỤC DỮ LIỆU**

1. **Sao lưu dữ liệu.**

Trong Apache Hive, việc sao lưu (backup) và phục hồi (restore) dữ liệu không được hỗ trợ trực tiếp qua lệnh như trong một số hệ quản trị cơ sở dữ liệu truyền thống. Tuy nhiên, vì Hive lưu trữ dữ liệu trên HDFS (Hadoop Distributed File System), bạn có thể thực hiện sao lưu và phục hồi bằng cách thao tác trực tiếp trên hệ thống tệp HDFS và metastore.

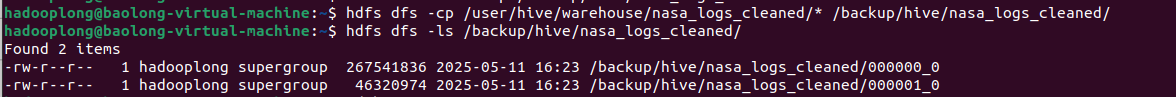
Kiểm tra thư mục lưu trữ warehouse trên hdfs.



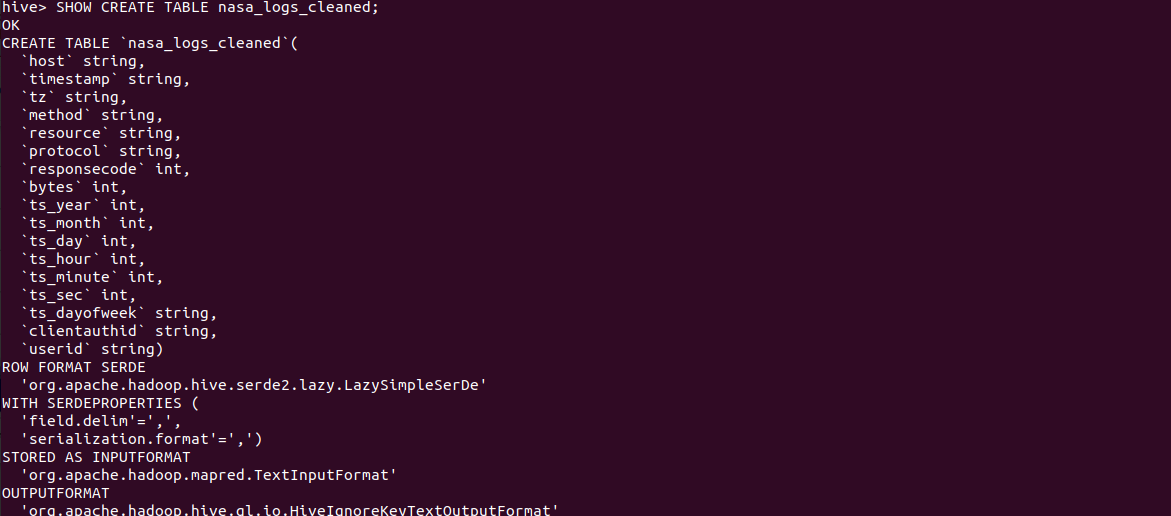
Tạo thư mục mới để lưu trữ dữ liệu sao lưu.



Đưa dữ liệu lưu trữ trên hdfs vào nơi lưu trữ dữ liệu sao lưu.



Vào hive để show lệnh tạo bảng.



Lệnh này chúng em sẽ lưu vào file .sql để sau này khi cần khôi phục sẽ sử dụng để tạo bảng

CREATE TABLE `nasa\_logs\_cleaned`(

`host` string,

`timestamp` string,

`tz` string,

`method` string,

`resource` string,

`protocol` string,

`responsecode` int,

`bytes` int,

`ts\_year` int,

`ts\_month` int,

`ts\_day` int,

`ts\_hour` int,

`ts\_minute` int,

`ts\_sec` int,

`ts\_dayofweek` string,

`clientauthid` string,

`userid` string)

ROW FORMAT SERDE

'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'

WITH SERDEPROPERTIES (

'field.delim'=',',

'serialization.format'=',')

STORED AS INPUTFORMAT

'org.apache.hadoop.mapred.TextInputFormat'

OUTPUTFORMAT

'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'

LOCATION

'hdfs://localhost:9000/user/hive/warehouse/nasa\_logs\_cleaned'

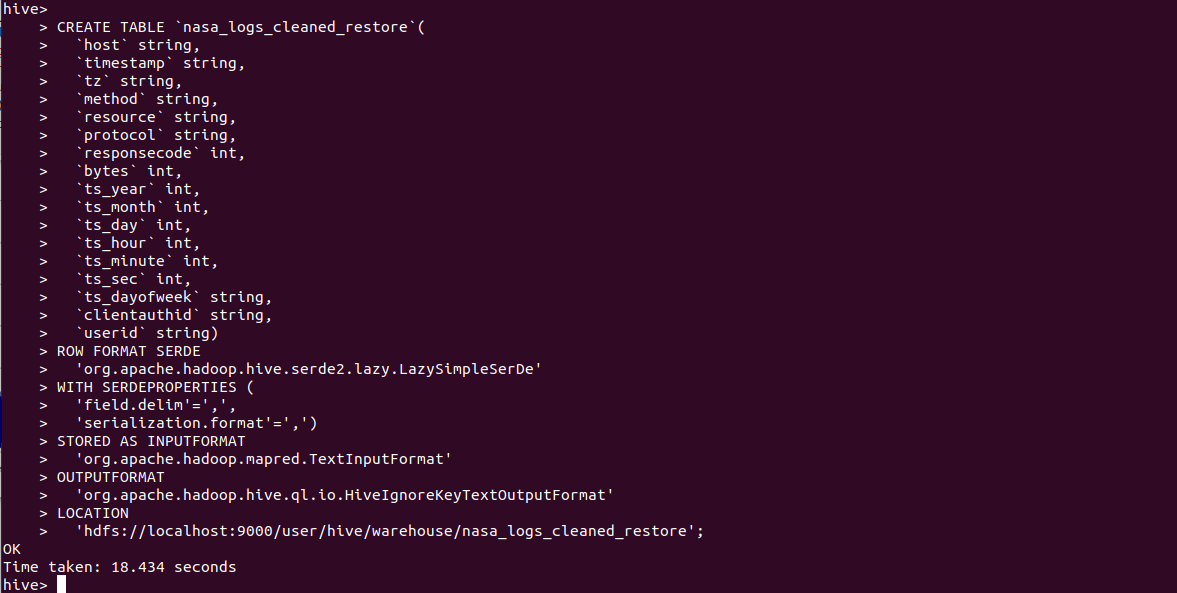
TBLPROPERTIES (

'bucketing\_version'='2',

'transient\_lastDdlTime'='1746337503')

1. **Phục hồi dữ liệu.**

Tạo bảng mới trên hive bằng câu lệnh crud sql ta lưu khi sao lưu



Phục hồi dữ liệu từ backup.



Kiểm tra phục hồi.

