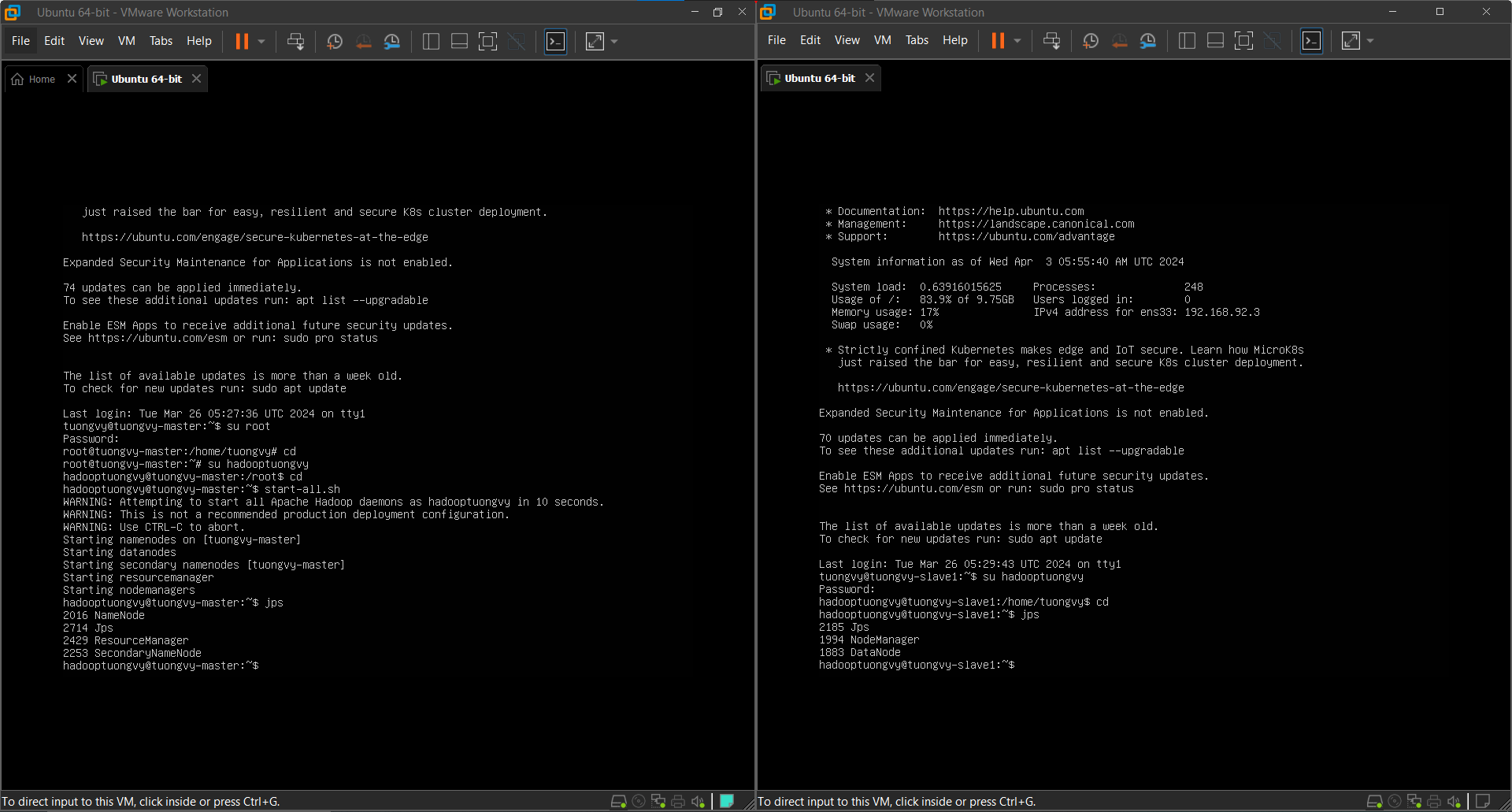
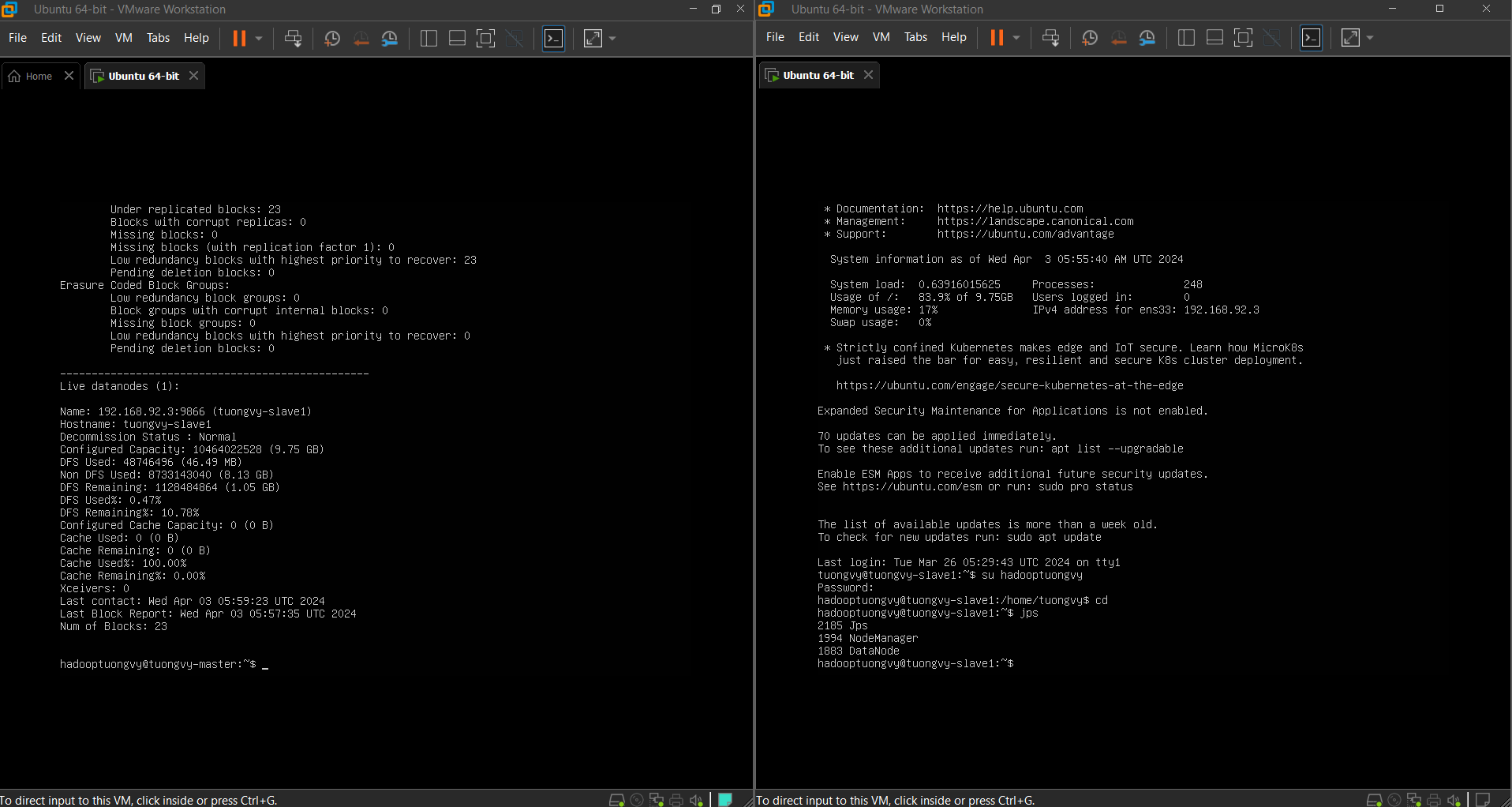
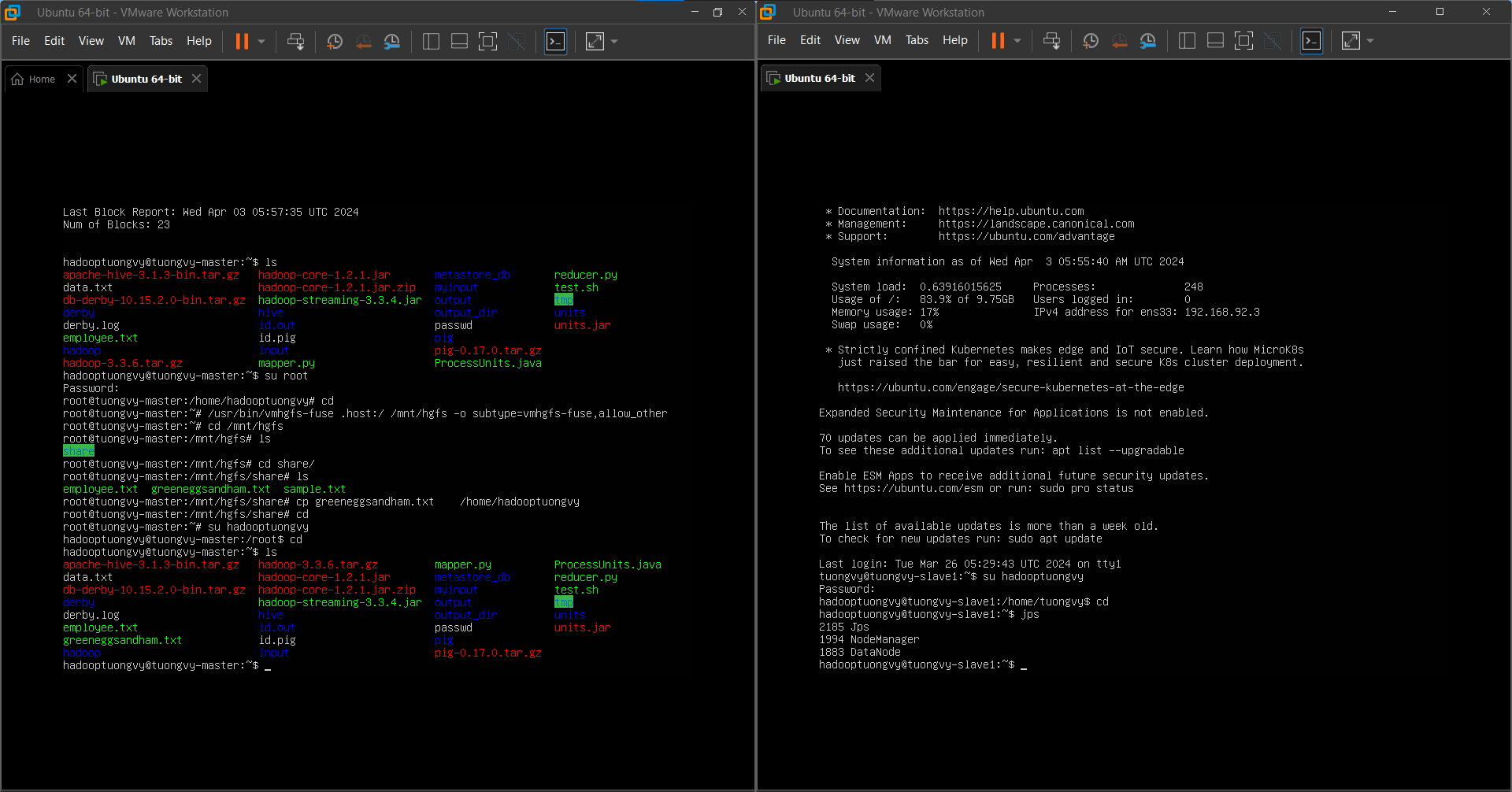
**Lương Tường Vy – 21133093 – KTGiuaKy**



$ hdfs dfsadmin –report

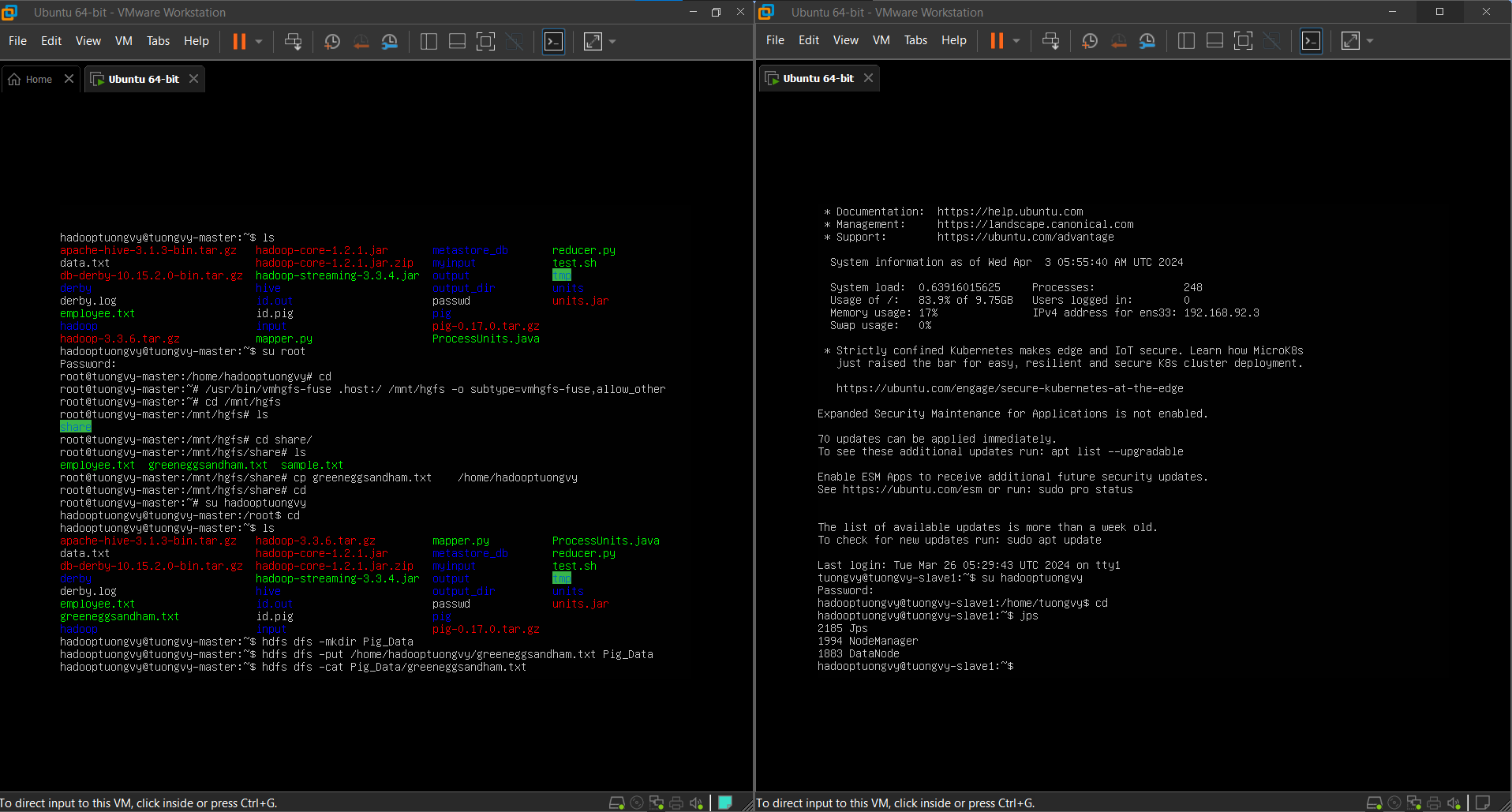


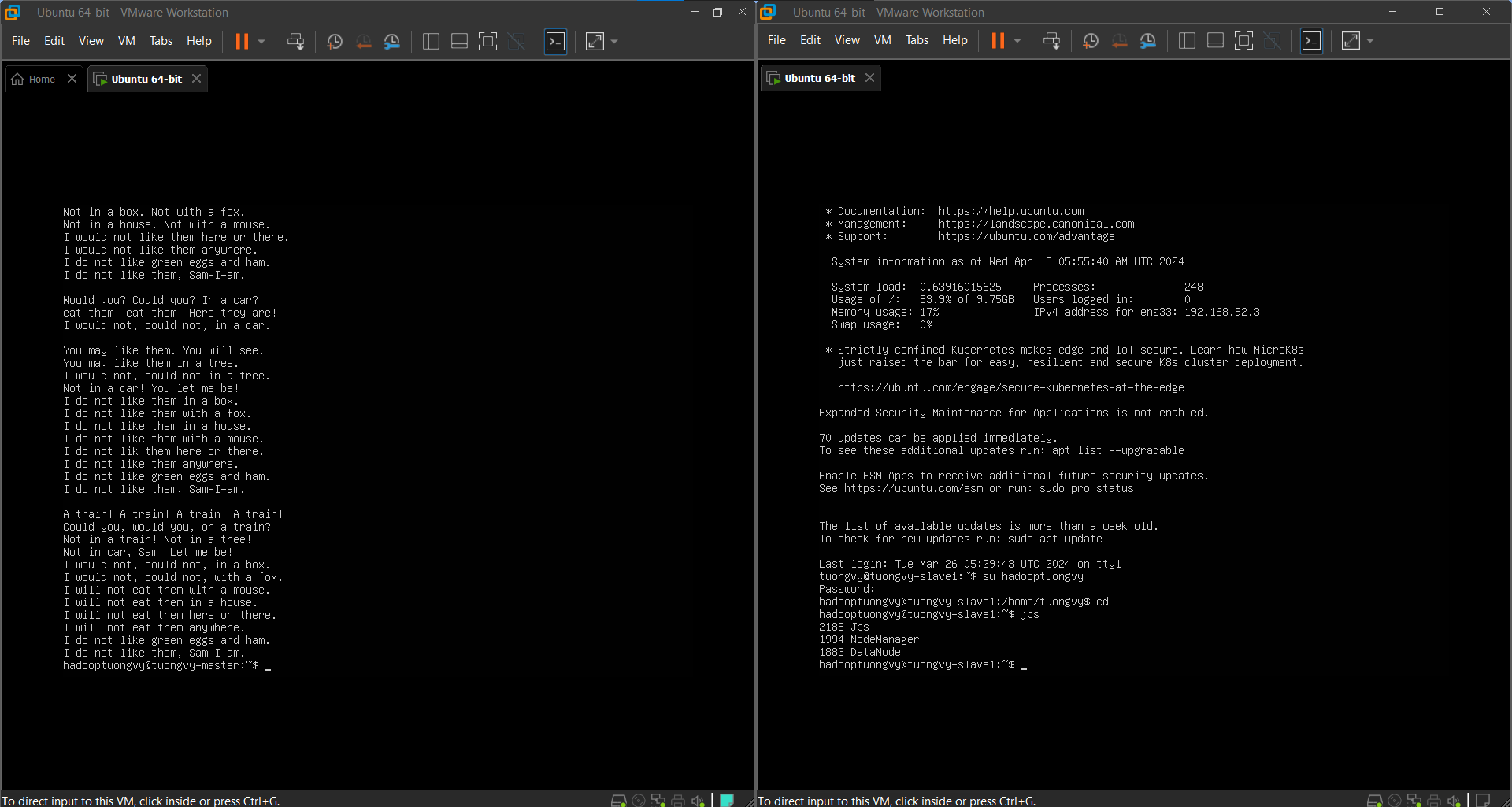
Về root cp file greeneggsandham.txt



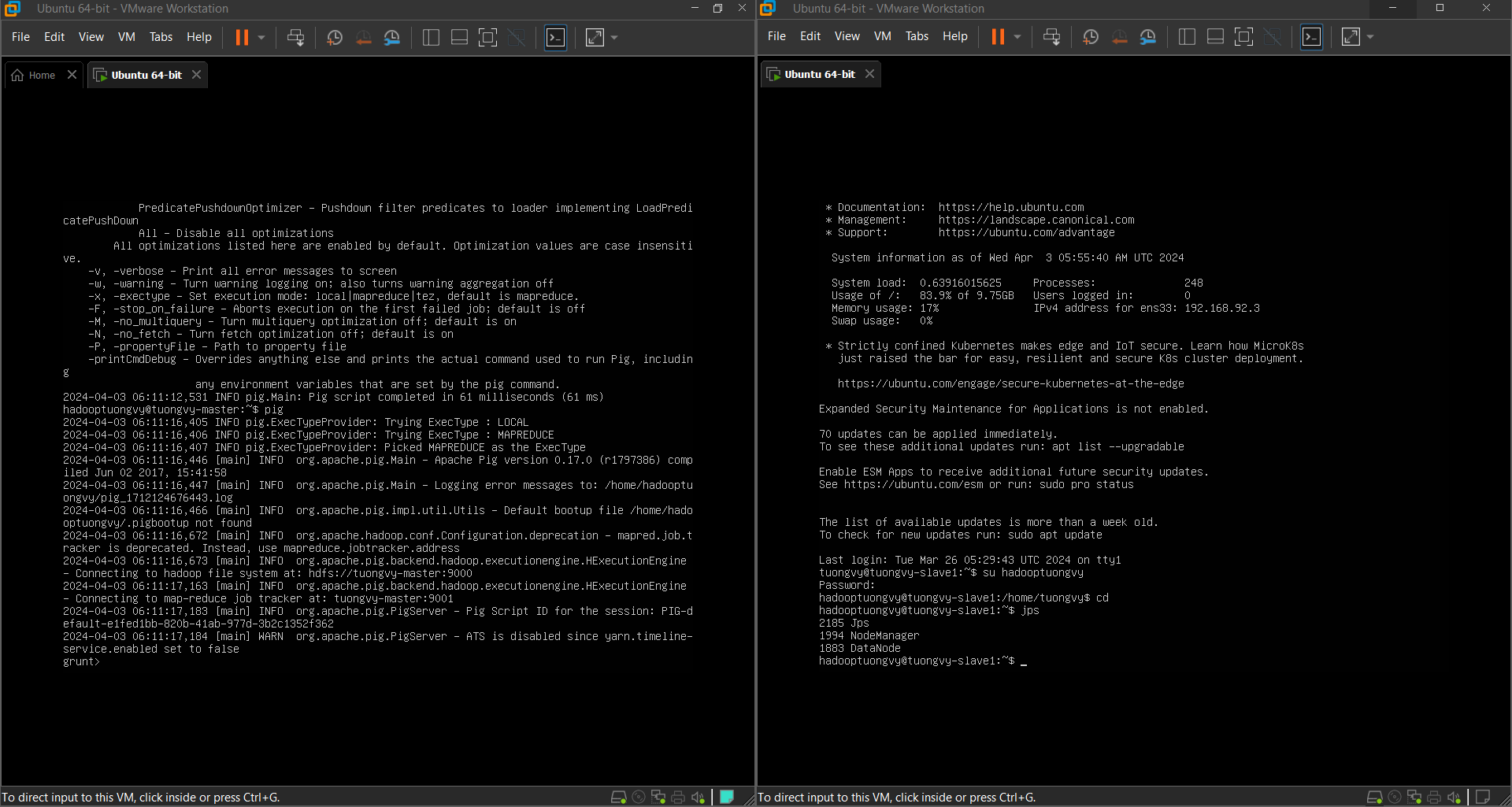
Tạo thư mục Pig\_Data

$ hdfs dfs –mkdir Pig\_Data



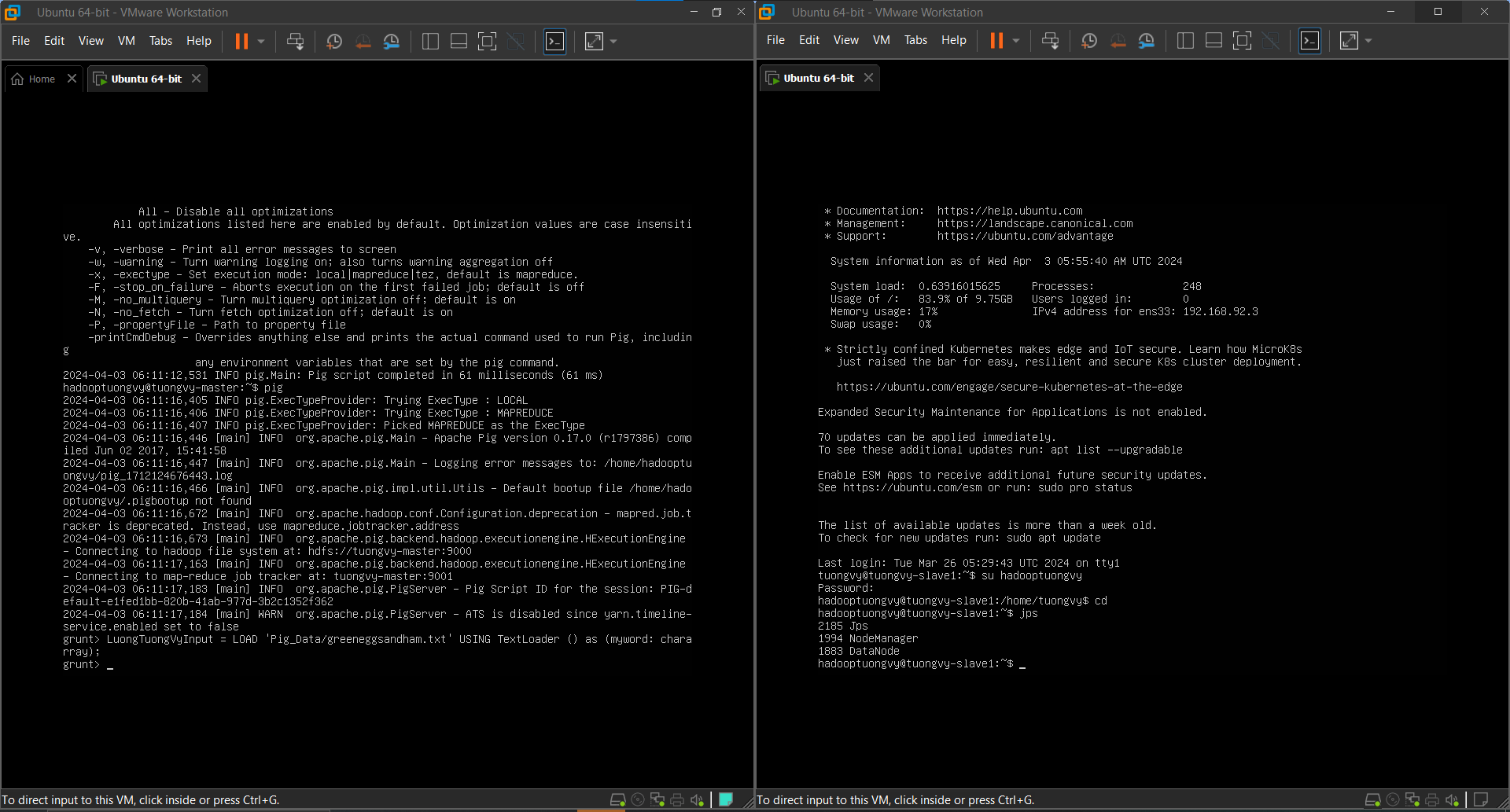


Chạy pig



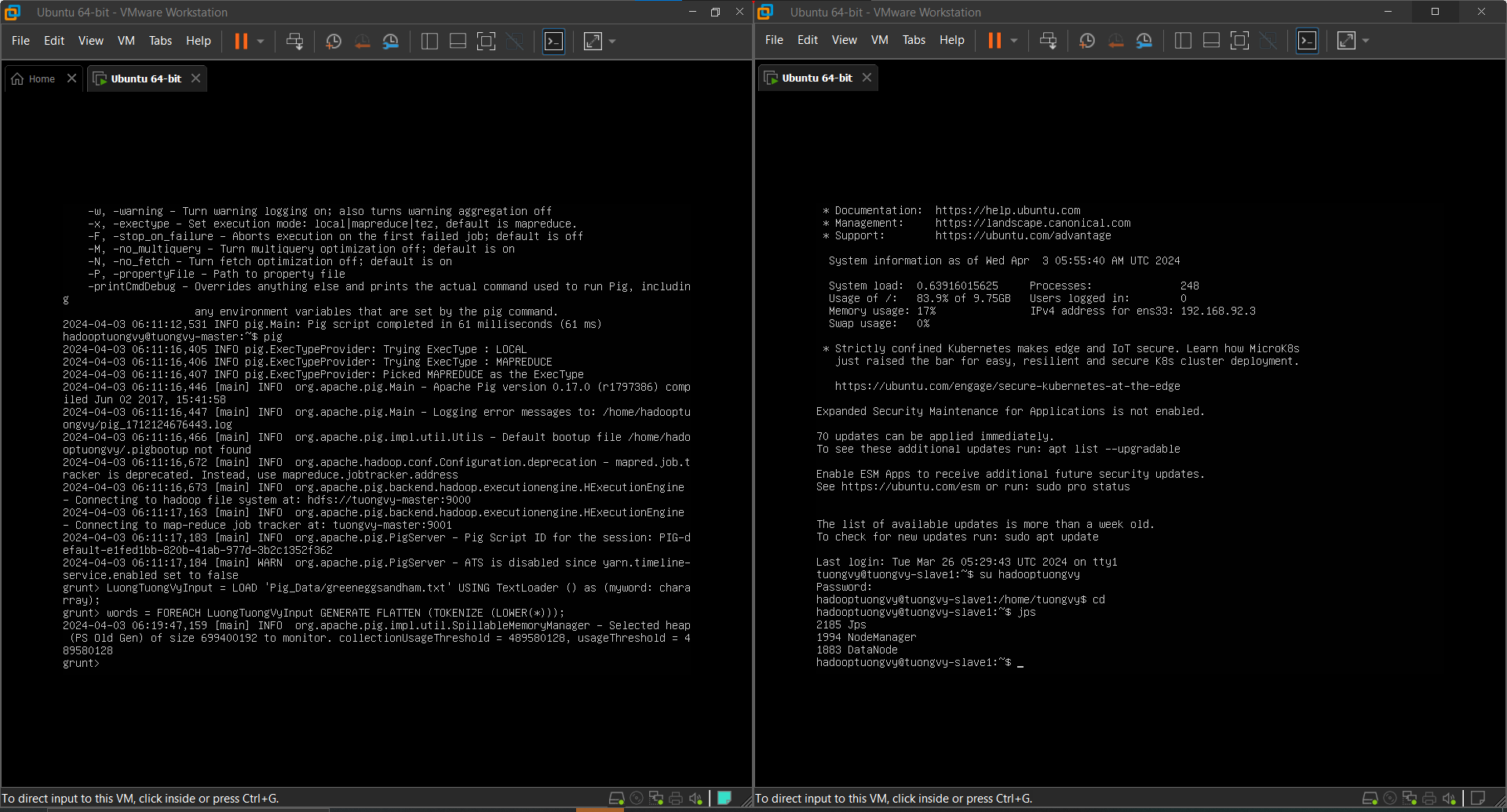
Load dữ liệu greeneggsandham.txt vào Pig và thực hiện các lệnh sau trong cửa sổ Grunt Shell, giải thích câu lệnh và chụp minh chứng kết quả thực thi:

1. grunt > LuongTuongVyInput = LOAD 'Pig\_Data/greeneggsandham.txt' USING TextLoader () as (myword: chararray);



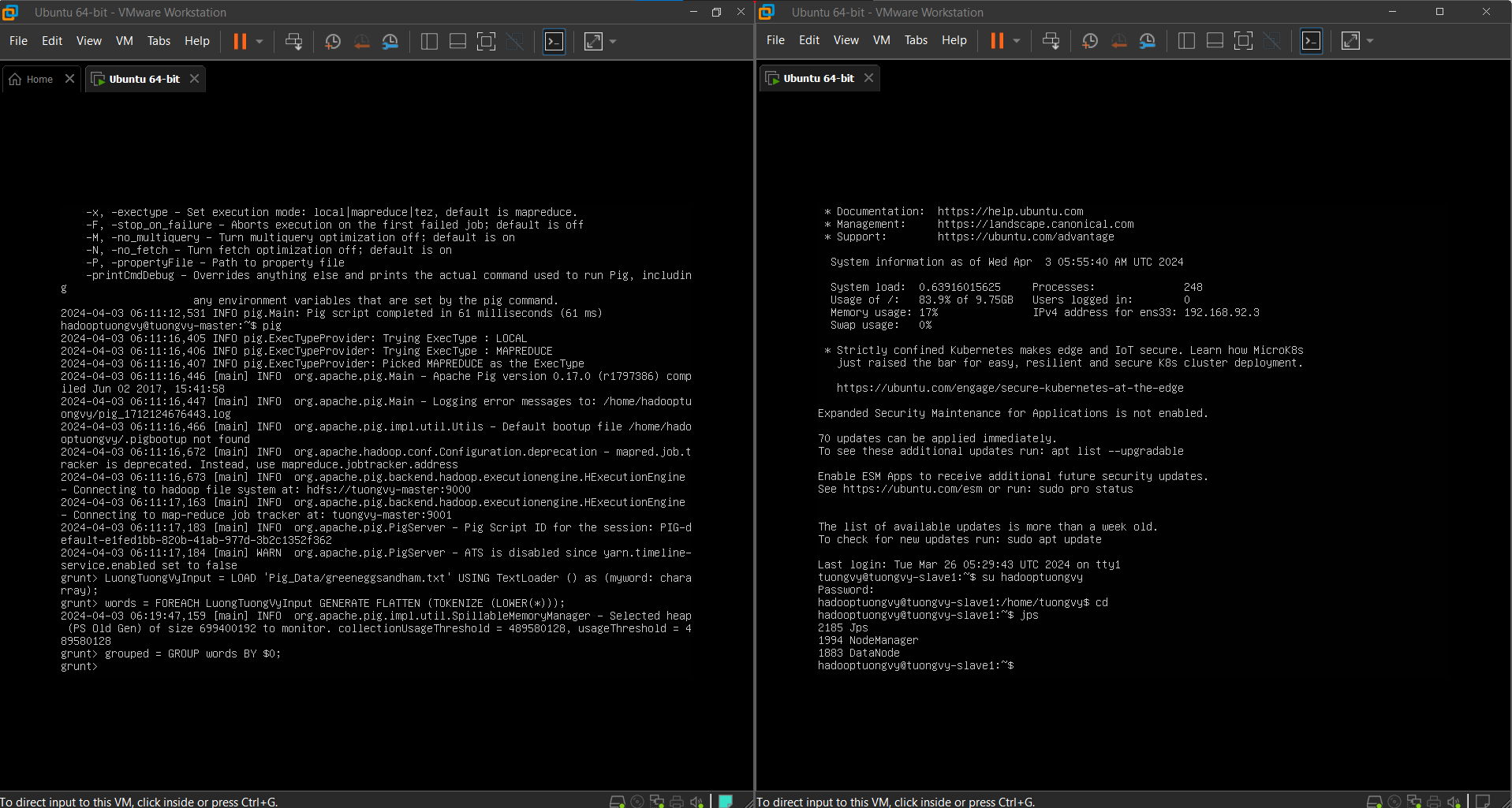
**Giải thích câu lệnh:**  Câu lệnh trên tải dữ liệu từ tập tin văn bản 'greeneggsandham.txt' và lưu trữ vào biến 'LuongTuongVyInput'. Dữ liệu trong tập tin được hiểu là chuỗi ký tự và được đặt tên cho cột dữ liệu là 'myword'.

1. grunt > words = FOREACH LuongTuongVyInput GENERATE FLATTEN (TOKENIZE (LOWER(\*)));



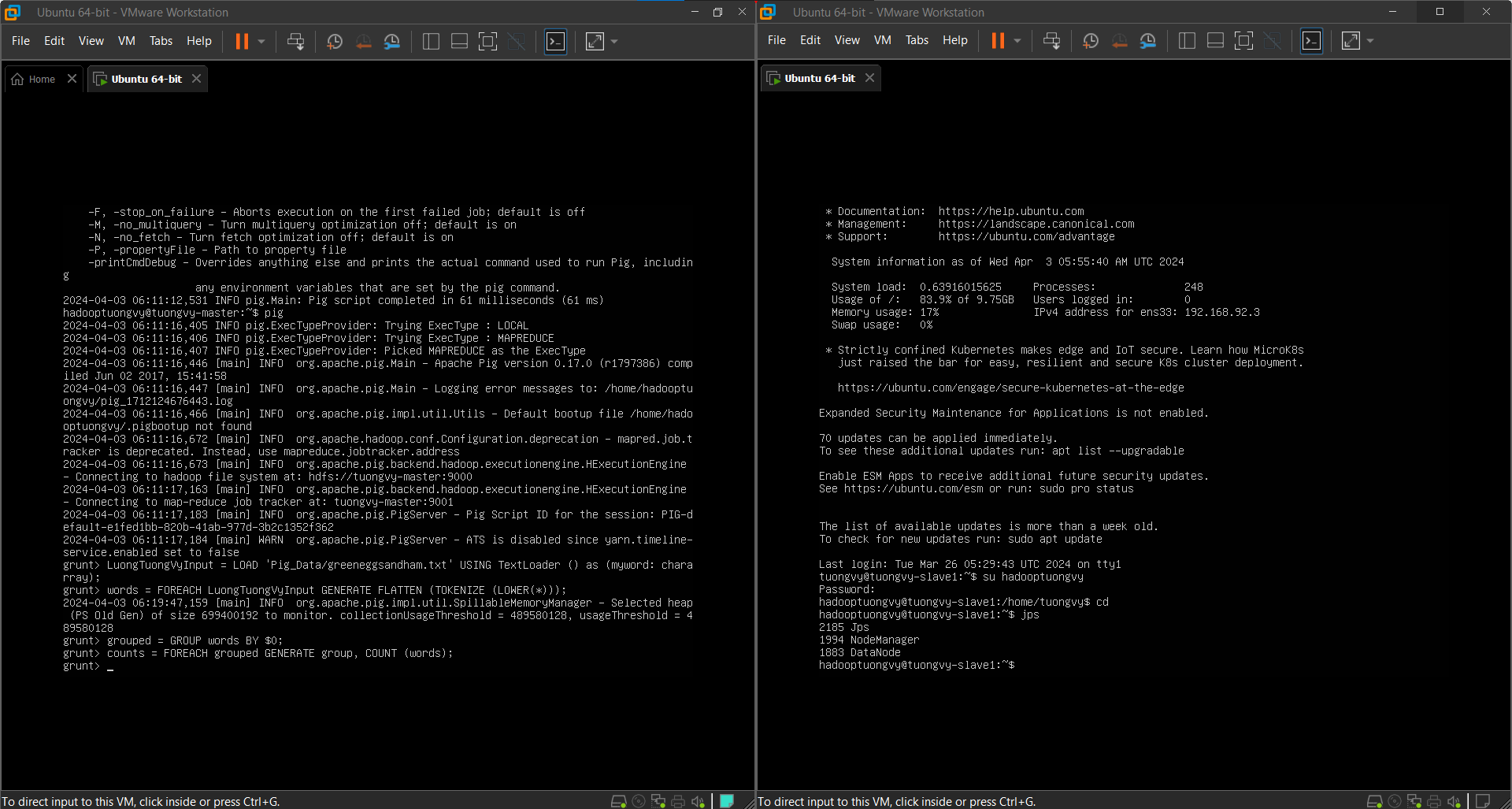
**Giải thích câu lệnh:**  Câu lệnh trên tạo ra một biến mới có tên là 'words' từ dữ liệu đầu vào 'LuongTuongVyInput'. Mỗi bản ghi trong 'LuongTuongVyInput' được chuyển thành một chuỗi chữ thường và sau đó tách thành các từ riêng lẻ. Các từ này được lưu trữ trong biến 'words' dưới dạng danh sách phẳng.

1. grunt > grouped = GROUP words BY $0;



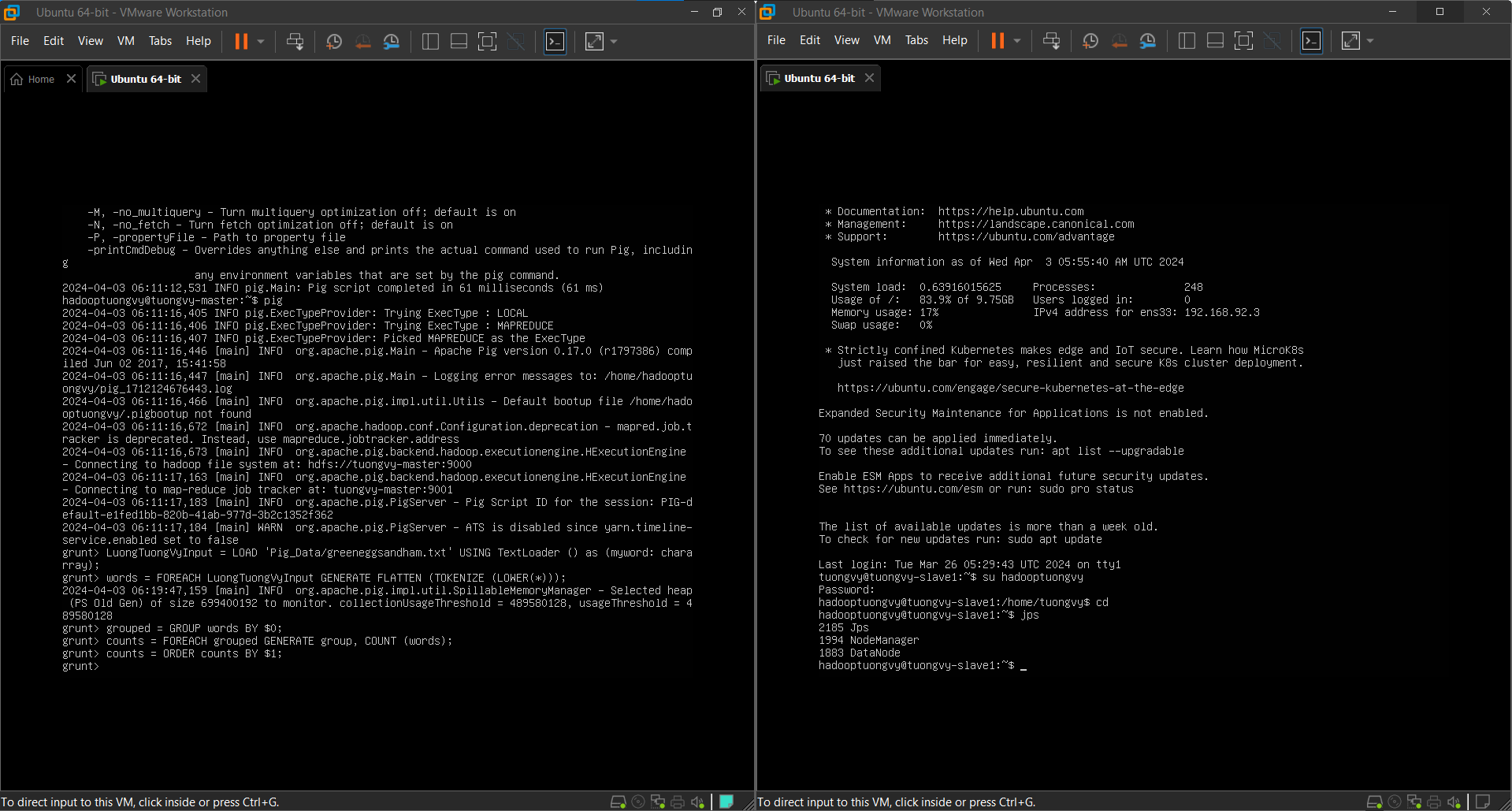
**Giải thích câu lệnh:**  Câu lệnh trên nhóm các từ trong biến 'words' dựa trên giá trị của trường đầu tiên trong danh sách các từ. Kết quả của việc nhóm này được lưu trữ trong biến 'grouped'.

1. grunt > counts = FOREACH grouped GENERATE group, COUNT (words);



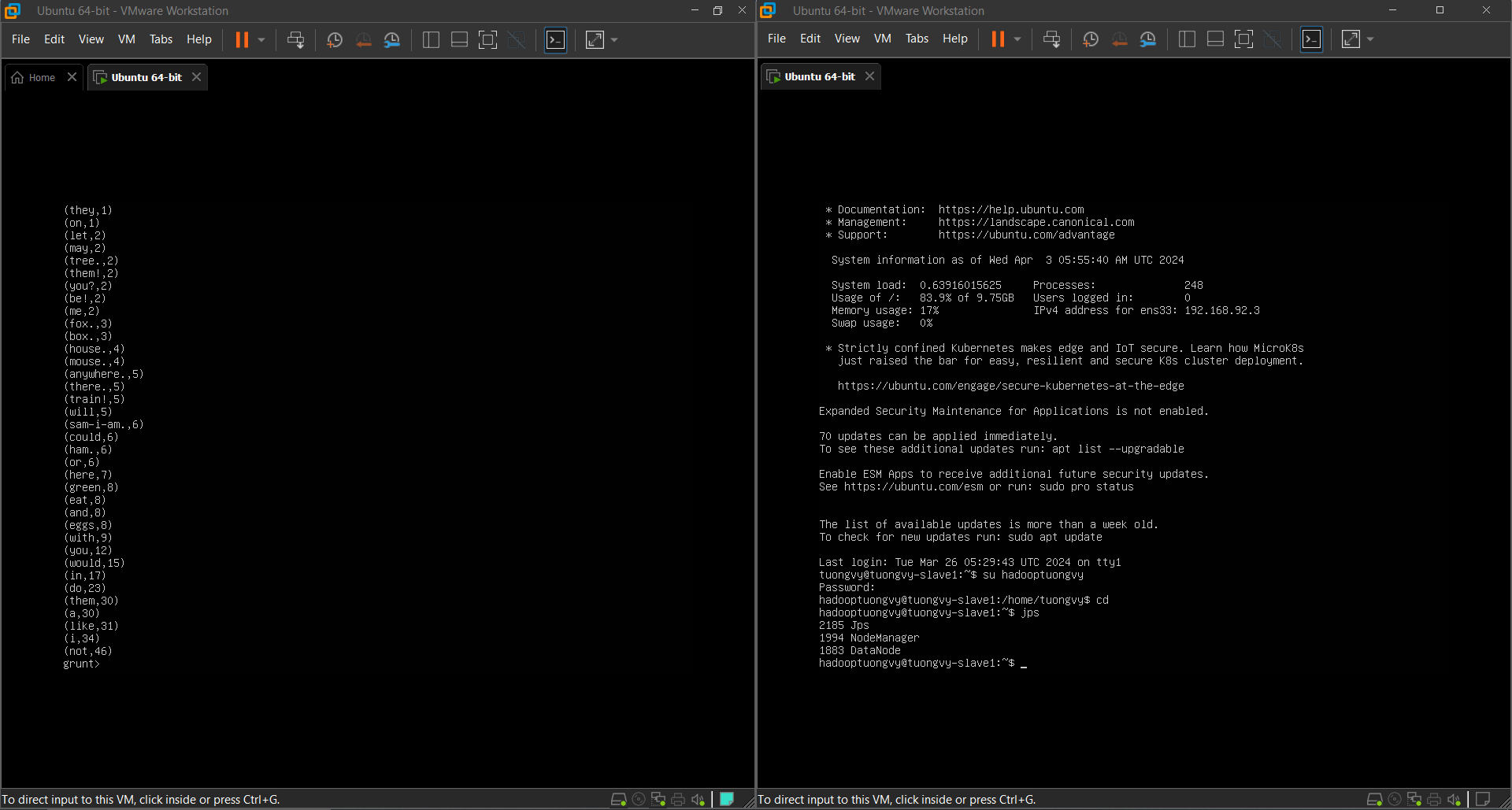
**Giải thích câu lệnh:**  Câu lệnh trên tạo ra một biến mới có tên là 'counts' từ dữ liệu đầu vào 'grouped'. Mỗi bản ghi trong 'grouped' được trích xuất trường 'group' và số lượng từ trong trường 'words' được đếm. Kết quả này được lưu trữ trong biến 'counts'.

1. grunt > counts = ORDER counts BY $1;



**Giải thích câu lệnh:**  Câu lệnh trên sắp xếp dữ liệu trong biến 'counts' dựa trên số lượng từ trong mỗi nhóm. Kết quả sắp xếp này được lưu trữ lại trong biến 'counts'.

1. grunt > dump counts;



**Giải thích câu lệnh:**  Câu lệnh trên được sử dụng để hiển thị nội dung của biến 'counts' lên màn hình.