# **KIỂM TRA GIỮA KỲ LẦN 1**

**Yêu cầu:**

1. SV tạo file Đỗ Hoàng Thịnh\_20133122\_KTGiuaKy1.docx lưu minh chứng quá trình làm bài và nộp lại trên hệ thống UTEX.
2. Chụp hình minh chứng thấy rõ thông tin sinh viên (HoTen-master, HoTen-slave), hình ảnh mờ, không thấy rõ không được chấm điểm.
3. Chạy Apache Pig ở chế độ Map Reduce.
4. 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:

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

**1. DoHoangThinh** = LOAD ‘.Pig\_Data/greeneggsandham.txt’ USING TextLoader () as (myword: chararray);

-- Câu lệnh này giúp cho pig đọc dữ liệu từ file greeneggsandham.txt từ hgfs

2. words = FOREACH **DoHoangThinhInput** GENERATE FLATTEN (TOKENIZE (LOWER(\*)));

--

3. grouped = GROUP words BY $0;

-- Toán tử GROUP được sử dụng để nhóm dữ liệu theo một hoặc nhiều quan hệ. Nó thu thập dữ liệu có cùng khóa là 0.

4. counts = FOREACH grouped GENERATE group, COUNT (words);

-- câu lệnh này đang sử dụng các điều khoản "foreach" và " GENERATE " để lặp lại mối quan hệ " grouped " và tạo mối quan hệ mới có chứa trường "group" và số lượng từ trong mỗi nhóm.

5. counts = ORDER counts BY $1;

-- chọn ra nhóm count mới với khóa là 1

6. dump counts;

-- chạy ra kết quả counts