Cao Thị Ngọc Phụng

21110276

# BÀI TẬP TUẦN 6

## 1. Install Python

**# apt update && apt upgrade -y**

Cập nhật danh sách các gói phần mềm có sẵn trong kho lưu trữ của hệ thống



**A screenshot of a computer

Description automatically generated**

**# apt install software-properties-common -y**

Cài đặt gói phần mềm có tên "software-properties-common" trên hệ thống của bạn

A screenshot of a computer

Description automatically generated

**# add-apt-repository ppa:deadsnakes/ppa -y**

Thêm một kho lưu trữ (PPA - Personal Package Archive) mới vào danh sách các nguồn phần mềm của hệ thống

A screenshot of a computer screen

Description automatically generated

**# add-apt-repository ppa:deadsnakes/nightly -y**

**A screenshot of a computer

Description automatically generated**

**# apt update**

**A screenshot of a computer program

Description automatically generated**

**# apt install python3.11**

Cài đặt python3.11

**A screenshot of a computer screen

Description automatically generated**

**# python3.11 –version**

Xem phiên bản

****

## 2. Example Using Python

### 2.1 Setting

**Tạo thư mục my\_input chứa các file như sau:**

A screenshot of a computer

Description automatically generated

**Tiến hành share thư mục vào máy ảo Ubuntu**

Setting đường dẫn

A screenshot of a computer

Description automatically generated

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

Để gắn (mount) các chia sẻ từ máy chủ VMware (host) vào hệ thống tập tin của máy ảo Linux (guest)

****

**# cd /mnt/hgfs**

**# ls**

danh sách các tệp tin và thư mục trong các chia sẻ từ máy chủ Vmware

A screen shot of a computer

Description automatically generated

**# cd my\_input**

**# ls**

**A screen shot of a computer

Description automatically generated**

### 2.2 Copy hết các file vào /home/hadoopngocphung

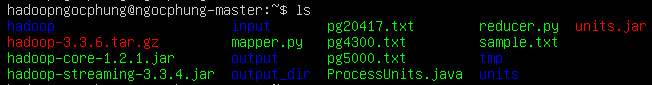
**A screenshot of a computer program

Description automatically generated**

**# su hadoopngocphung**

**# ls**

Quay về hadoop để kiểm tra xem file đã có chưa



### 2.3. Mapper Phase Code

# vim mapper.py

A screen shot of a computer program

Description automatically generated

# vim reducer.py

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

## 3. Thực thi chương trình WordCount trên thư mục cục bộ

$ echo "foo foo quux labs foo bar quux" | /home/hadoopngocphung/mapper.py

A computer screen with white text

Description automatically generated

Sau khi chạy lệnh thì em nhận đc thông báo lỗi ký tự ^M thường là kết quả của việc sử dụng tập tin có định dạng Windows trong một môi trường Unix và em có thử tải dos2unix nhưng không sử được và đây là cách khác em tìm được trên web stack overflow

# sed -i 's/\r$//' /home/hadoopngocphung/mapper.py

A computer screen with white text

Description automatically generated

$ echo "foo foo quux labs foo bar quux" | /home/hadoopphuthanh/mapper.py | sort -k1,1 | /home/hadoopphuthanh/reducer.py

A black screen with white text

Description automatically generated



A black background with white text

Description automatically generated

$ cat ./data.txt | ./mapper.py:

A screen shot of a computer

Description automatically generated

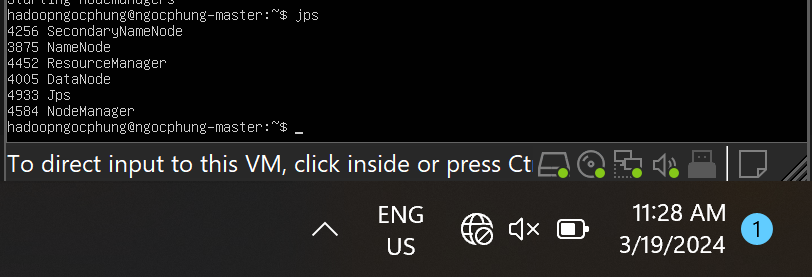
cat ./data.txt | ./mapper.py | sort -k1,1 | ./reducer.py

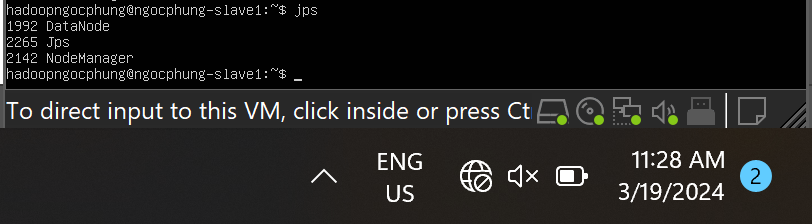
A screen shot of a computer

Description automatically generated

## 4. Thực thi chương trình WordCount trên HDFS

Nhập lệnh **start-all.sh** để chạy các dịch vụ Hadoop trong Ubuntu



****

Tạo thư mục myinput

A screenshot of a computer

Description automatically generated

Di chuyển file vào thư mục myinput

A screen shot of a computer

Description automatically generated

Xem các file trong thư mục

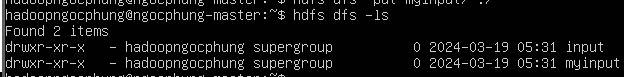


Copy thư mục myinput vào hdfs

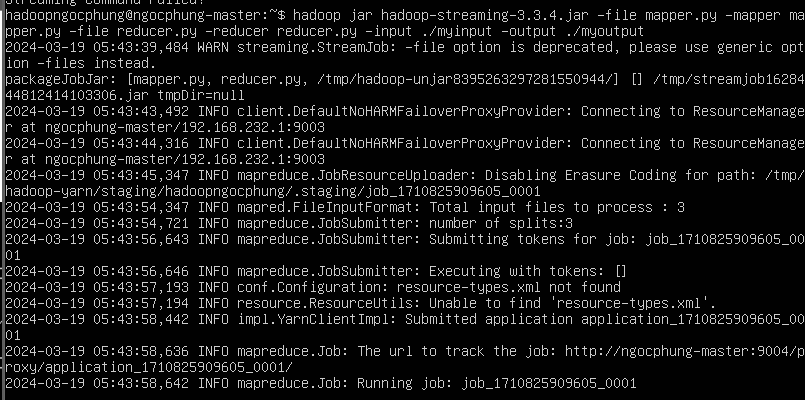


$ hdfs dfs -put myinput/ ./

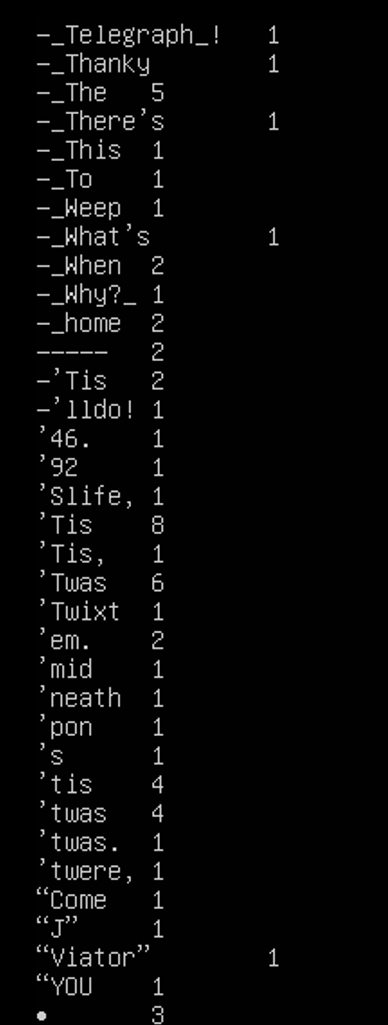
$ hdfds dfs -ls



$ hadoop jar hadoop-streaming-3.3.4.jar -file mapper.py -mapper mapper.py -file reducer.py -reducer reducer.py -input ./myinput -output ./myoutput



$ hdfs dfs -cat ./myoutput/part-00000

**