Cao Thị Ngọc Phụng

21110276

**Bài Tập Tuần 5**

# I. Setting

Chọn “Share folders” để cài đặt

Chọn “Always enabled” để cho phép chia sẽ folders

Chọn “Add” để thêm đường dẫn

A screenshot of a computer

Description automatically generated

Chọn đường dẫn mình lưu trên máy windows

A screenshot of a computer

Description automatically generated

“Enable this share” thì cho phép chình sửa, “Read-only” chỉ được phép xem, ở đây mình chọn “Enable this share”

A screenshot of a computer

Description automatically generated

Đường dẫn folders đã được thêm

A screenshot of a computer error

Description automatically generated

# II. Cài đặt tools chia sẽ

**# apt-get update**

Hệ thống sẽ kiểm tra các nguồn gốc của các gói phần mềm trong hệ thống và cập nhật thông tin về các gói phần mềm mới nhất có sẵn

A screenshot of a computer

Description automatically generated

**# apt-get upgrade**

Hệ thống sẽ tải về và cài đặt các bản cập nhật mới nhất cho tất cả các gói phần mềm đã được cài đặt trên hệ thống

A screenshot of a computer program

Description automatically generated

**# apt-get install open-vm-tools**

giúp tương thích và tối ưu hóa hiệu suất khi chạy hệ điều hành Linux trên máy ảo VMware

A screen shot of a computer

Description automatically generated

**# apt-get install open-vm-tools-desktop**



Nếu đã từng cài đặt (hoặc cài đặt thành công nhưng gõ lệnh lại) sẽ hiện ra phiên bản mới nhất mình đã cài

A black screen with white text

Description automatically generated

**# mkdir /mnt/hgfs**

Tạo một thư mục mới

**# 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 Mapreduce**

**# ls**

liệt kê tất cả các tệp tin và thư mục trong thư mục "Mapreduce"

**A screen shot of a computer screen

Description automatically generated**

Kiểm tra xem có phải những phải của folder không

**A screenshot of a computer

Description automatically generated**

# III Coppy file từ folders đã chia sẽ

**# cp sample.txt /home/hadoopngocphung/**

sao chép (copy) tệp tin "sample.txt" vào thư mục "/home/hadoopngocphung/

**# chmod 777 /home/hadoopngocphung/sample.txt**

đặt quyền đọc (read), quyền ghi (write), và quyền thực thi (execute) cho tất cả mọi người trên tệp tin đó

****

**Tương tự với 2 file còn lại**





**# su hadoopngocphung**

**# ls**

Quay về hadoop để kiểm tra xem file đã có chưa

A black screen with white text and green letters

Description automatically generated

# IV Thông dịch và thực thi chương trình

## 4.1 Tạo thư mục units chứa các file sau khi thông dịch ProcessUnits.java

$ mkdir units

A black screen with white text

Description automatically generated

## 4.2 Download hadoop-core-1.2.1.jar

Đã có các file từ folders chia sẽ

## 4.3 Thông dịch file nguồn và tạo file jar

**$ javac -classpath hadoop-core-1.2.1.jar -d units ProcessUnits.java**

hadoop-core-1.2.1.jar chứa các class file cần thiết để thông dịch chương trình, -d xác định thư mục chứa các class file được tạo ra



**$ jar -cvf units.jar -C units/ .**

Tạo file units.jar chứa tất cả (dấu .) file trong thư mục units (-C là lấy trong đường dẫn hiện tại)

A screen shot of a computer

Description automatically generated

**Kiểm tra các dịch vụ giữa 2 máy**

**$ start-all.sh**

Bật tất cả dịch vụ trên máy master



**$ jps (trên máy master)**

A screenshot of a computer

Description automatically generated

**$ hdfs dfsadmin -report**

Truy vấn thông tin về trạng thái của hệ thống tệp HDFS

A screenshot of a computer

Description automatically generated

**$ jps (trên máy slave1)**

**A screenshot of a computer

Description automatically generated**

## 4.4 Tạo thư mục input trong HDFS

$ hdfs dfs -mkdir input\_dir



## 4.5 Đưa dữ liệu sample.txt vào thư mục input\_dir trong HDFS

$ hdfs dfs -put sample.txt input\_dir

A screenshot of a computer

Description automatically generated

## 4.6 Thực thi chương trình Eleunit\_max trong Hadoop

$ hadoop jar units.jar hadoop.ProcessUnits input\_dir output\_dir

Đợi cho chương trình thực thi xong sẽ cho ra kết quả số lượng input được chia tách, số lượng tác vụ Map và Reduce…

A screenshot of a computer screen

Description automatically generated

## 4.7 Xuất kết quả

$ hdfs dfs -cat output\_dir/part-00000

A screen shot of a computer

Description automatically generated

## 4.8 Copy kết quả từ Hadoop ra thư mục bên ngoài

$ hdfs dfs -get output\_dir /home/hadoopminhchau

A screen shot of a computer

Description automatically generated