Lab - 01:

Cài đặt môi trường chạy Hadoop

Lớp: 19_21

Môn: Nhập môn dữ liệu lớn

Sinh viên thực hiện:

Lê Thành Lộc – 19120562

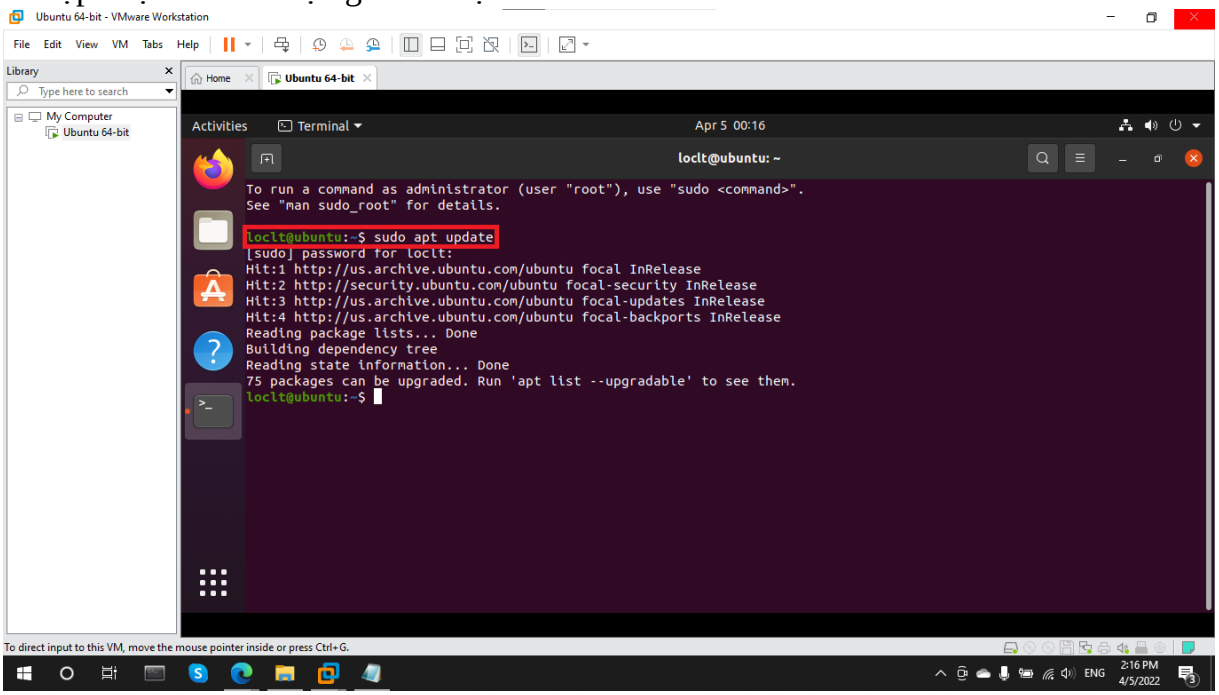
Mục lục

I. Các bước cài đặt và hình ảnh:

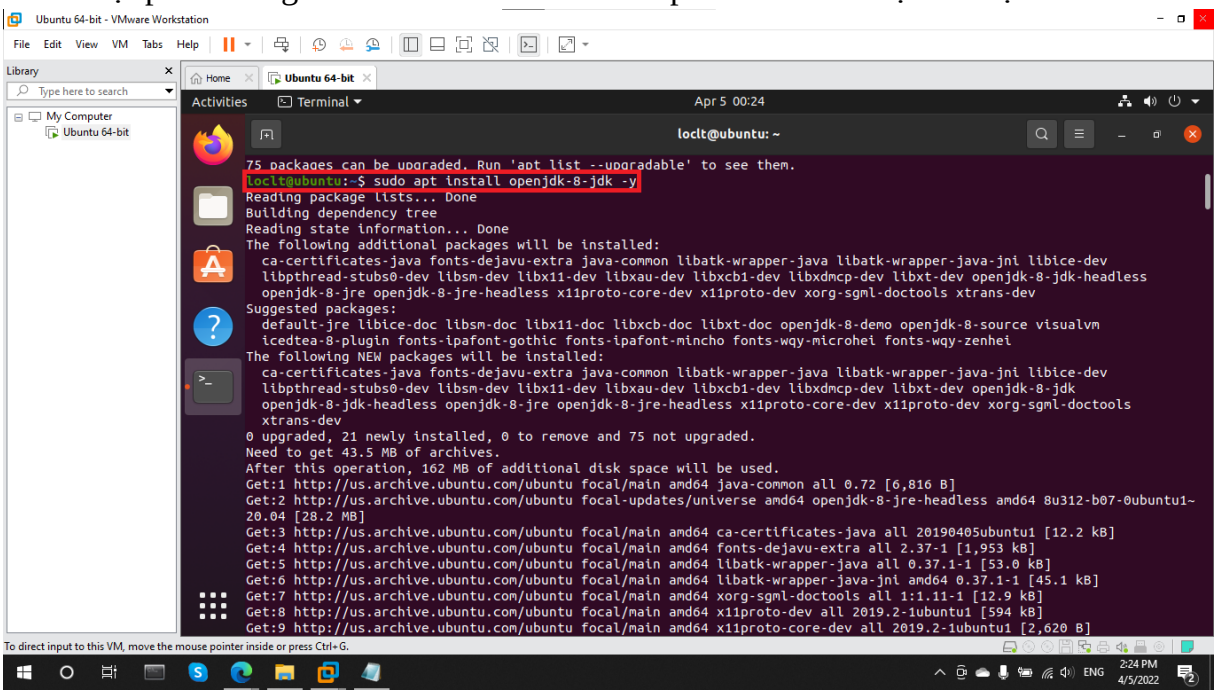
- Hệ điều hành được sử dụng: **Ubuntu 20.04**

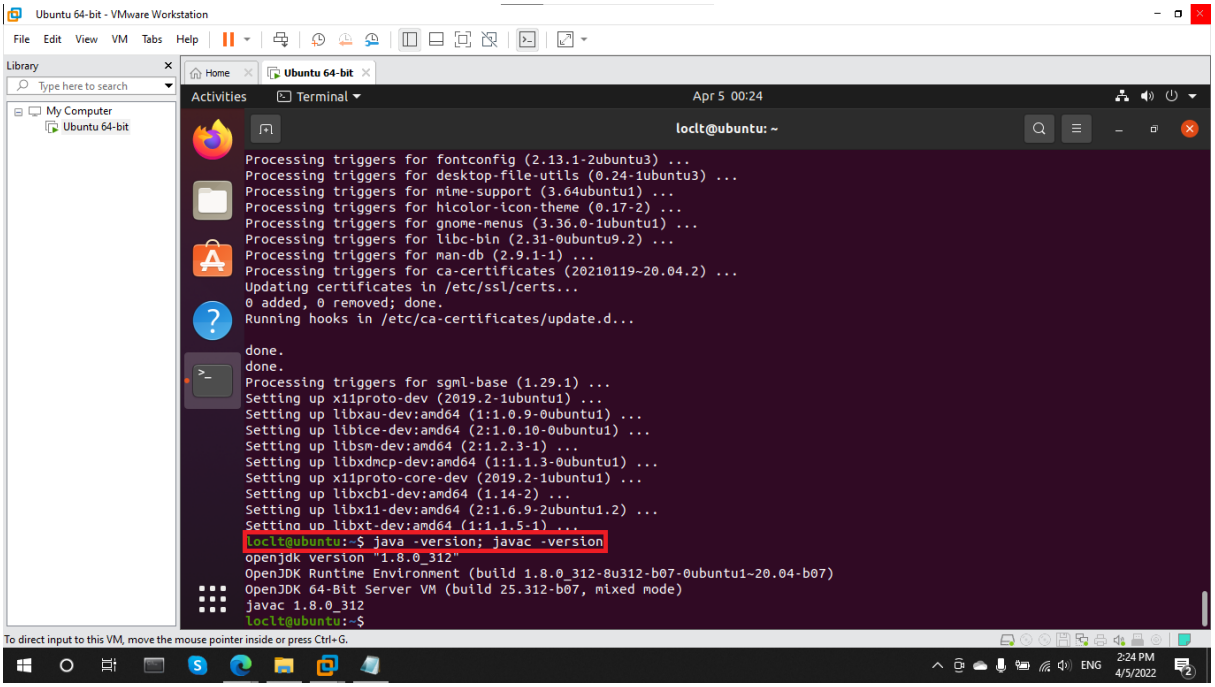
- Lần lượt thực hiện các bước cài đặt Hadoop như sau:

1. Cập nhật các chỉ mục gói của hệ điều hành Linux

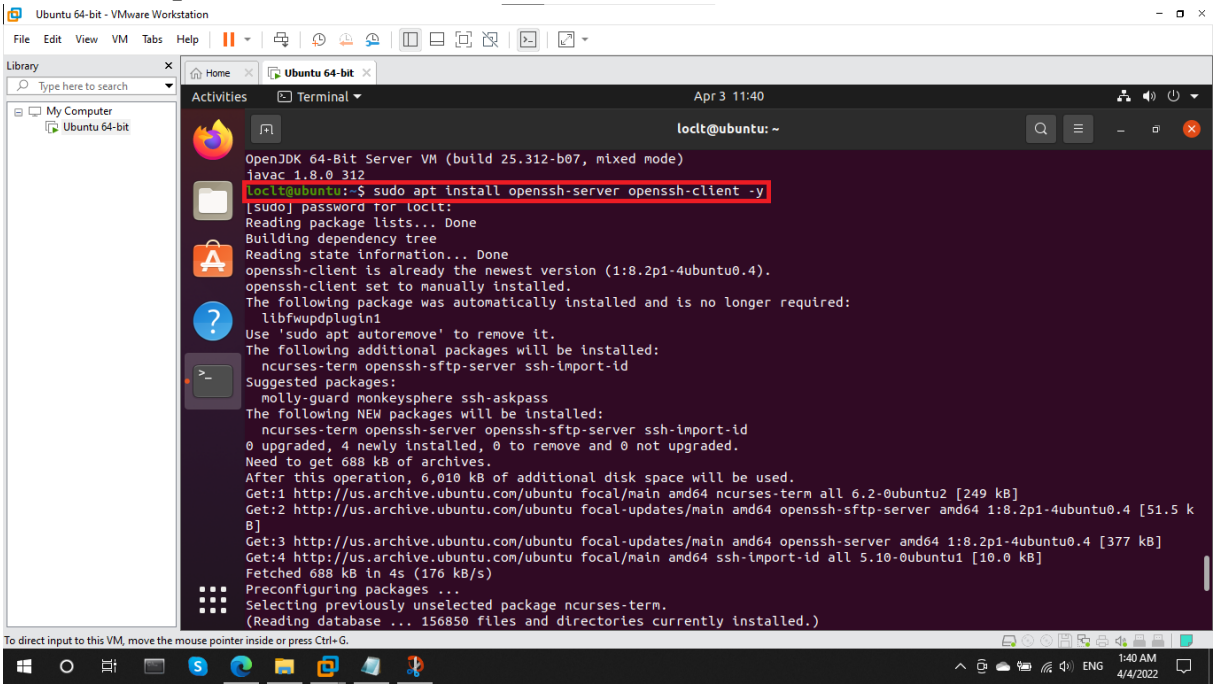


2. Cài đặt phiên bảng **Java 1.8** và kiểm tra xem phiên bản đã được cài đặt chưa

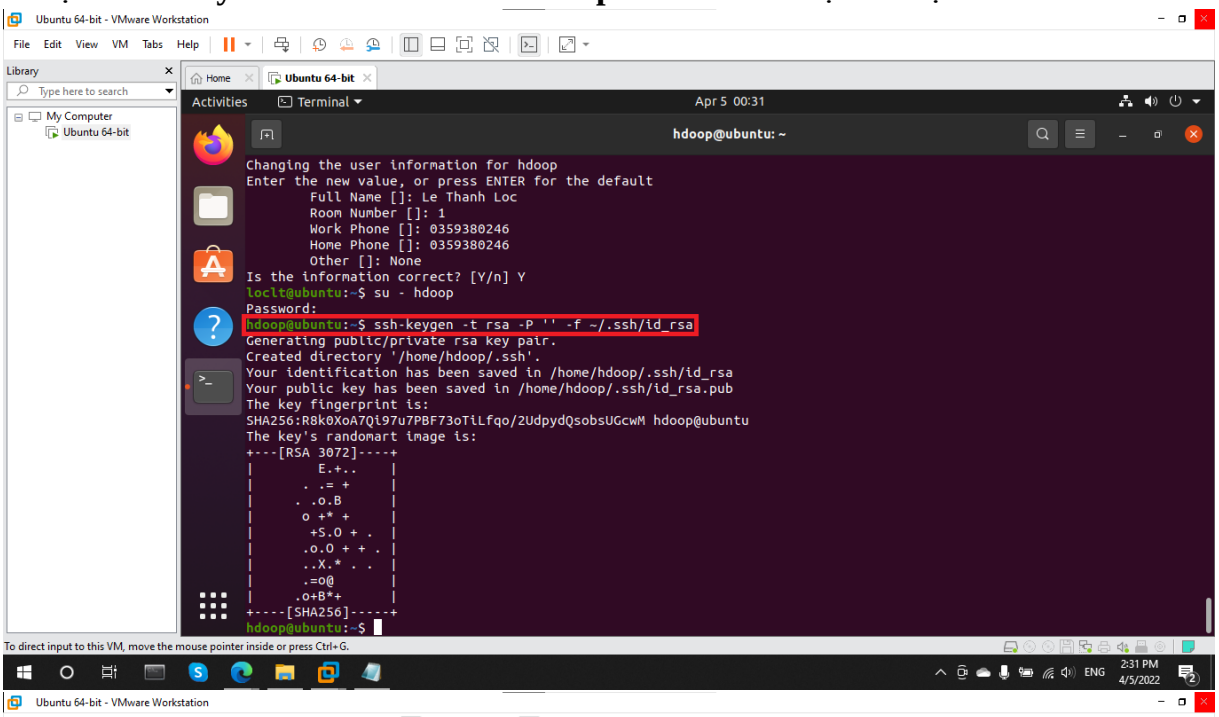


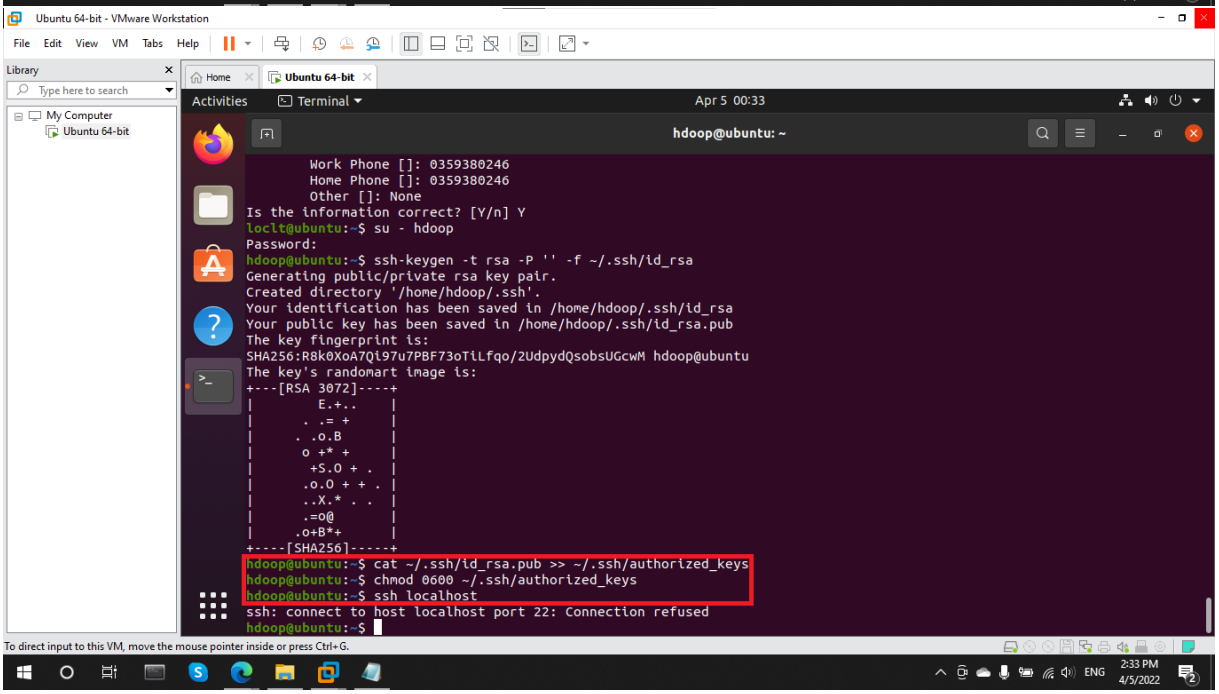


3. Cài đặt **OpenSSH Server** trên Linux

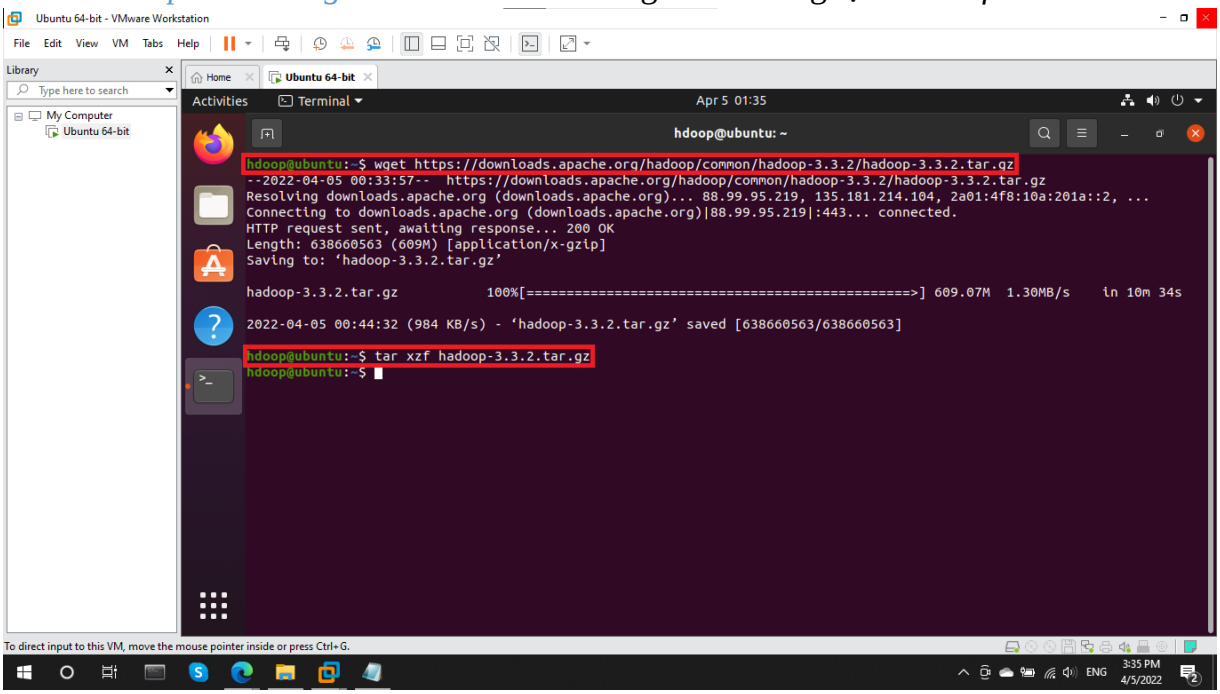


4. Tạo **SSH Key** và **SSH certificate** với **OpenSSH** vừa được cài đặt





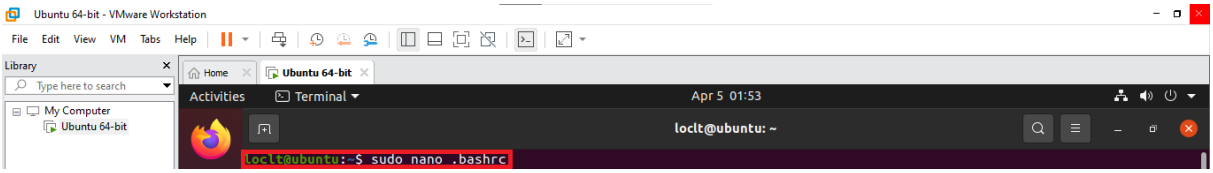
5. Cài đặt hadoop từ link *https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz* và sau đó tiến hành giải nén bằng lệnh *tar xzf*

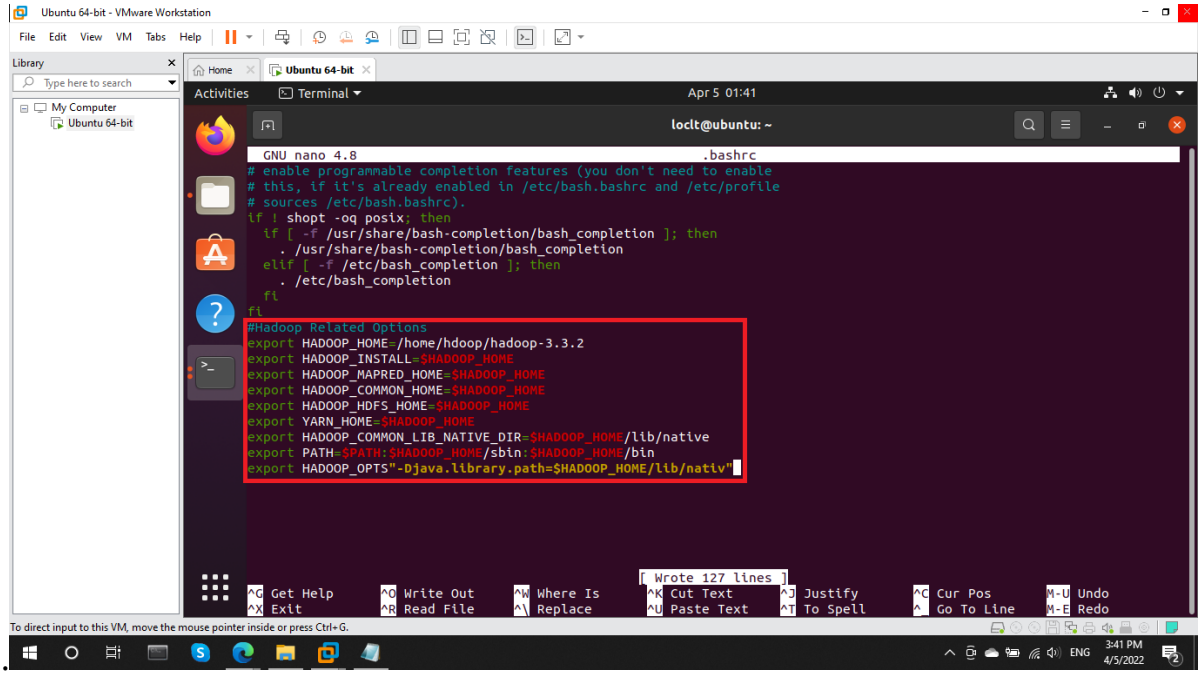


6. Thực hiện chỉnh sửa file **.bashrc**

- Nhập lệnh *sudo nano .bashrc* để mở file **.bashrc**.

- Thêm các dòng lệnh *export* như hình vào cuối file.

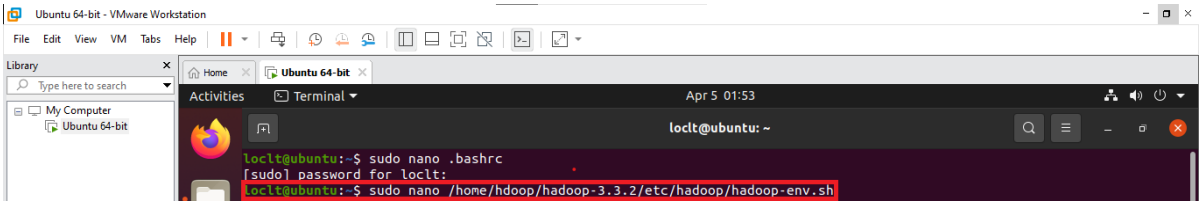


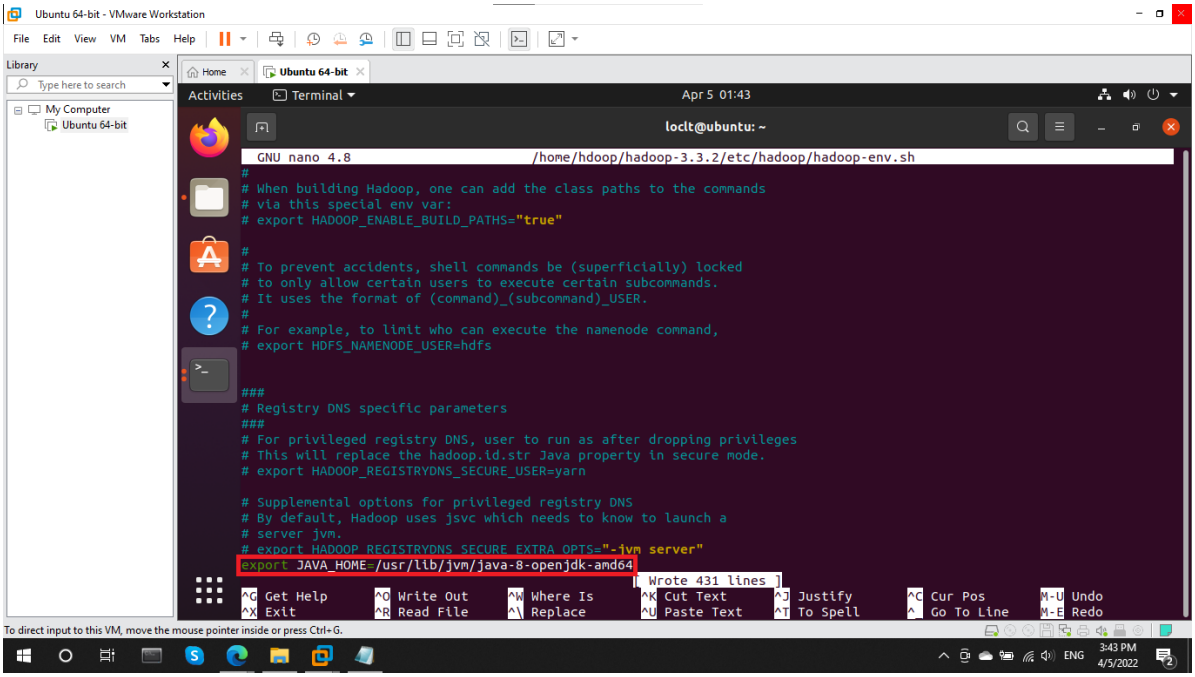


7. Thực hiện chỉnh sửa **hadoop-env.sh** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh *sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hadoop-env.sh* để mở file.

- Thêm dòng lệnh **export** vào cuối file như bên dưới và lưu lại.

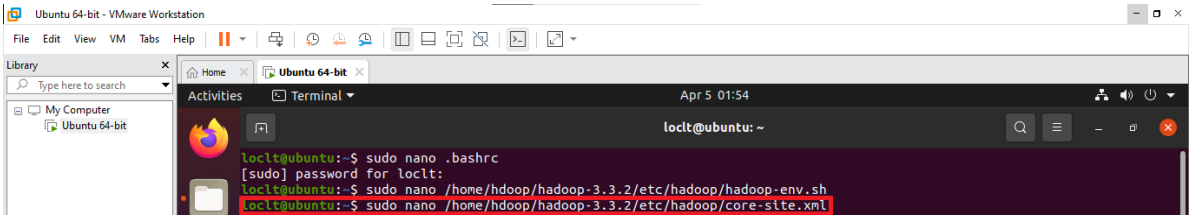


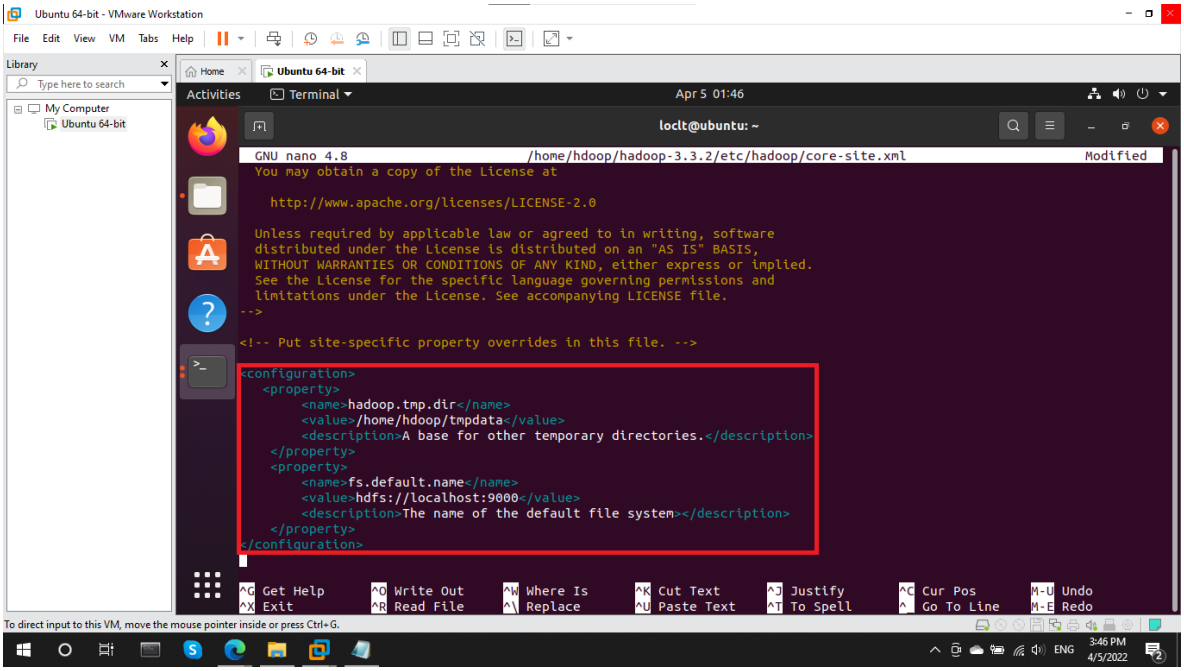


8. Thực hiện chỉnh sửa **core-site.xml** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh *sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/core-site.xml* để mở file.

- Thêm các dòng lệnh vào giữa các tag **configuration** như hình vẽ bên dưới.

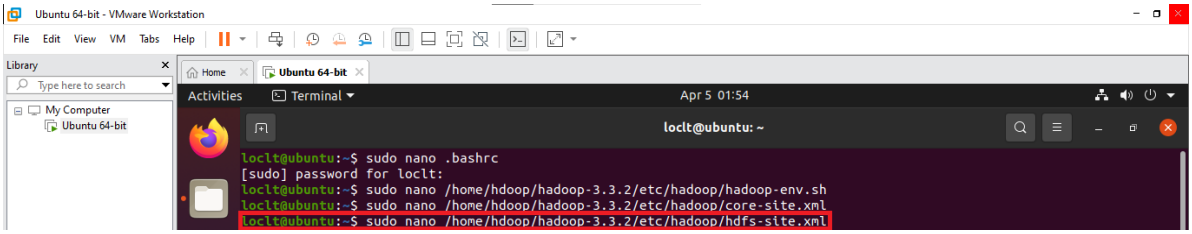


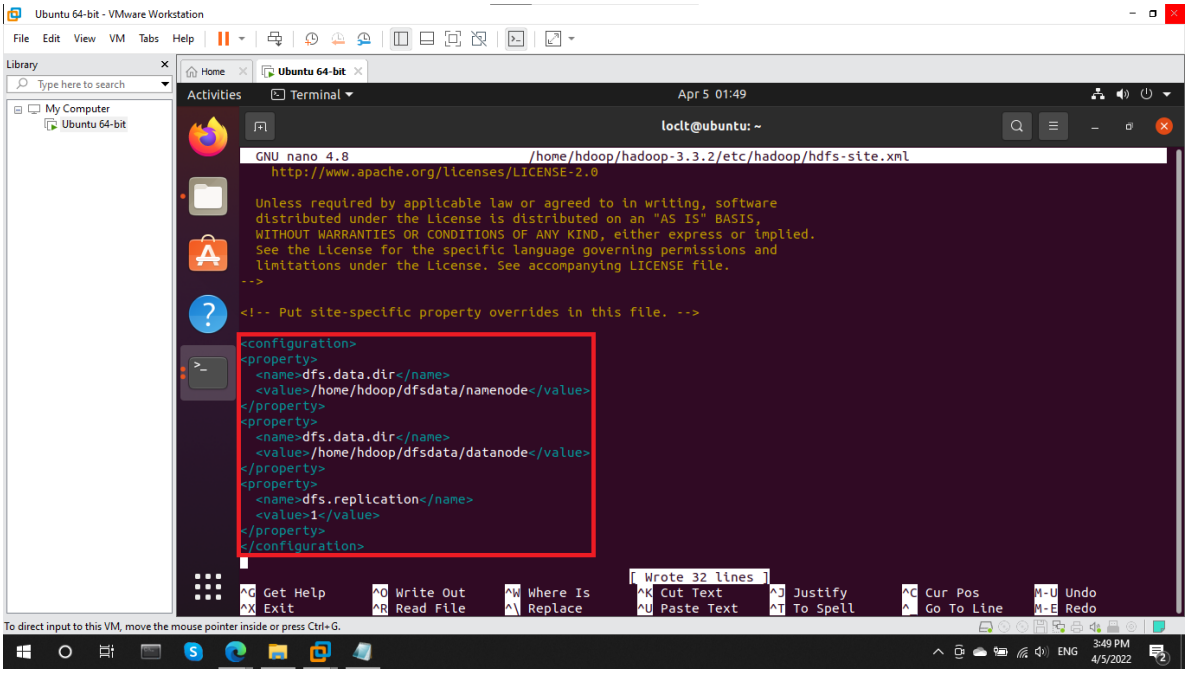


9. Thực hiện chỉnh sửa **hdfs-site.xml** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh *sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/hdfs-site.xml* để mở file.

- Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.

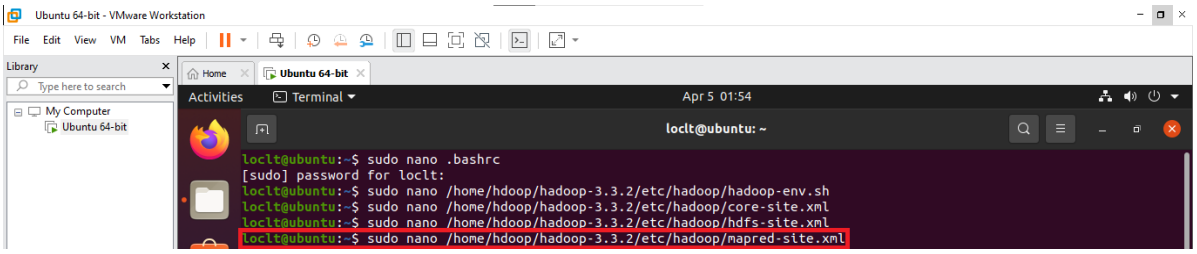


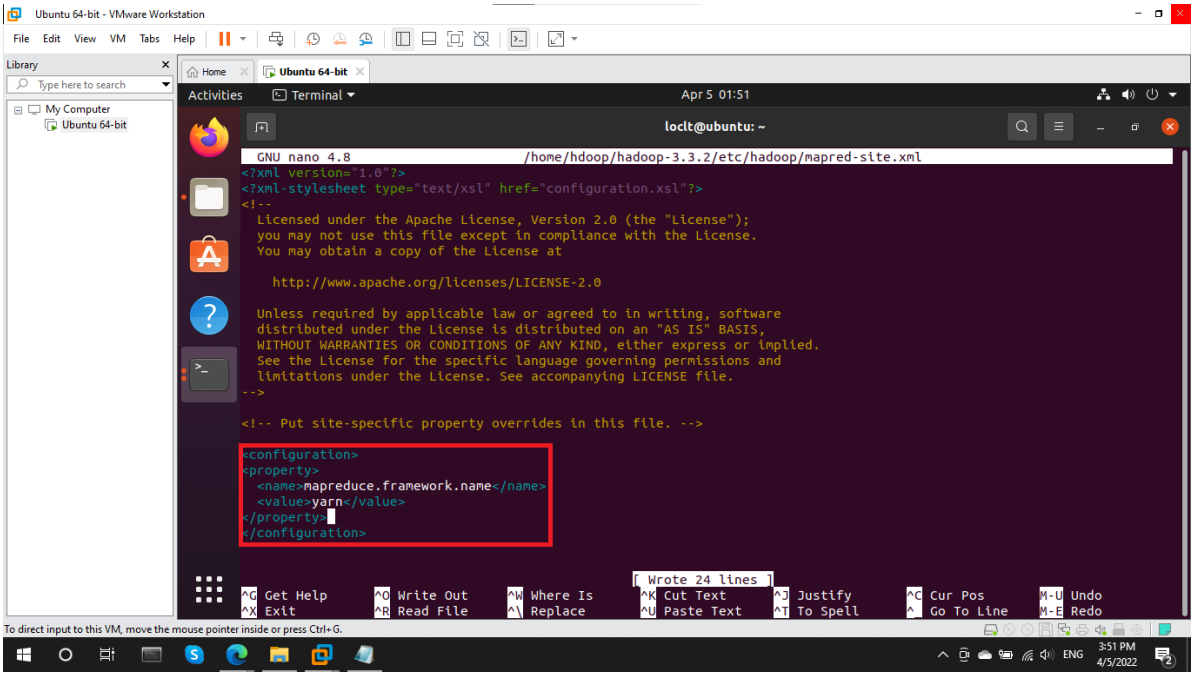


10. Thực hiện chỉnh sửa **mapred-site.xml** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh *sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/mapred-site.xml* để mở file.

- Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.

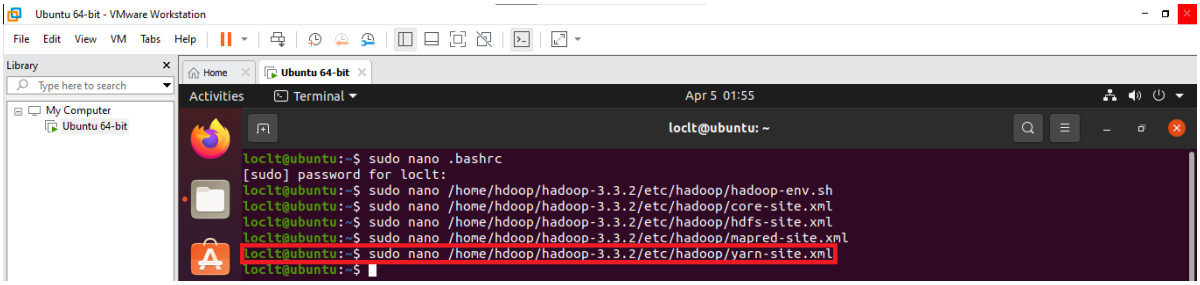


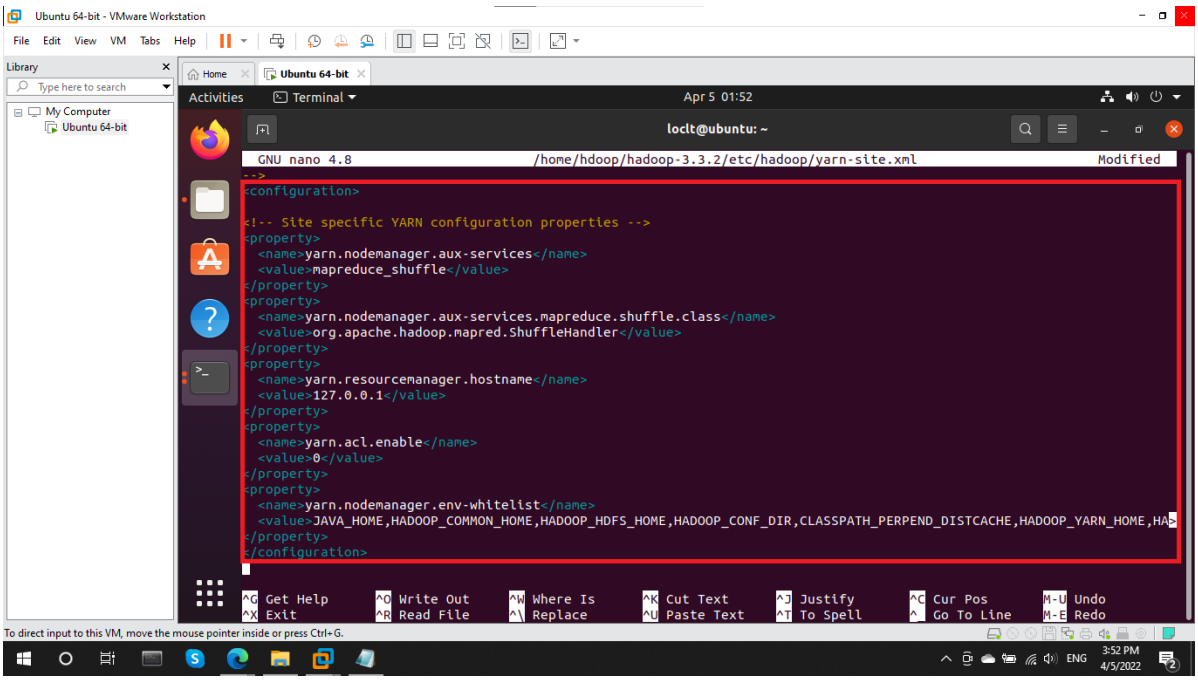


11. Thực hiện chỉnh sửa **yarn-site.xml** trong thư mục **hadoop-3.3.2** được tải về

- Nhập lệnh *sudo nano /home/hdoop/hadoop-3.3.2/etc/hadoop/yarn-site.xml* để mở file.

- Thêm các dòng lệnh vào giữa các tag **configuration** như hình bên dưới.

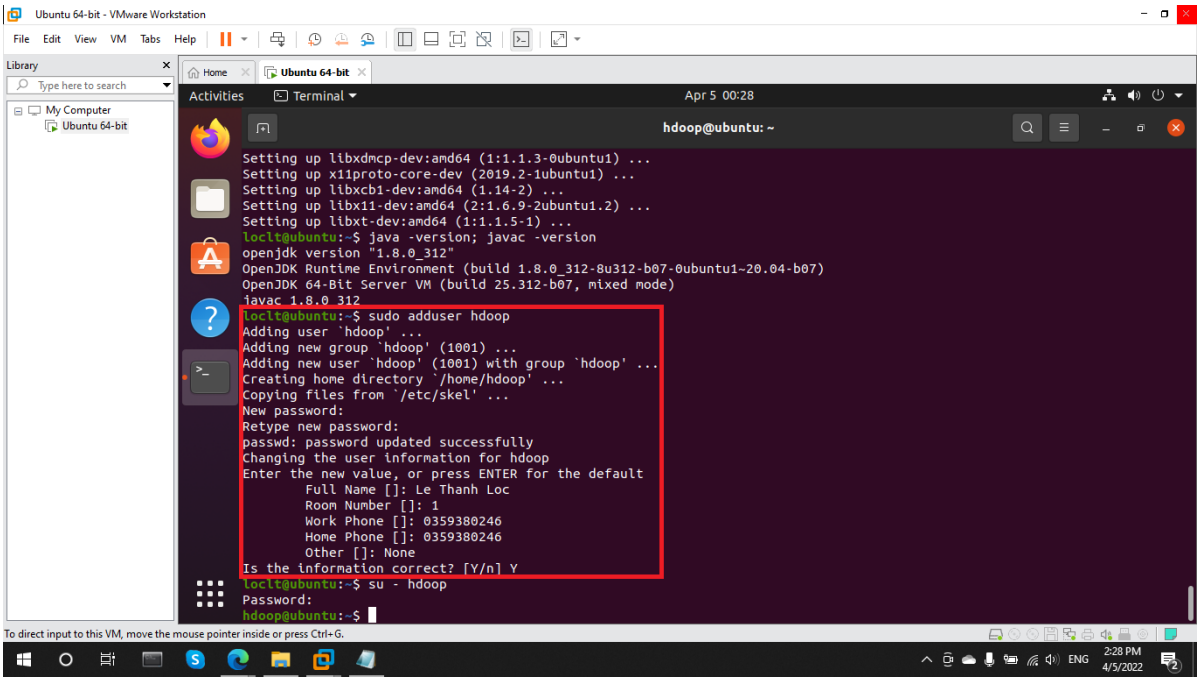


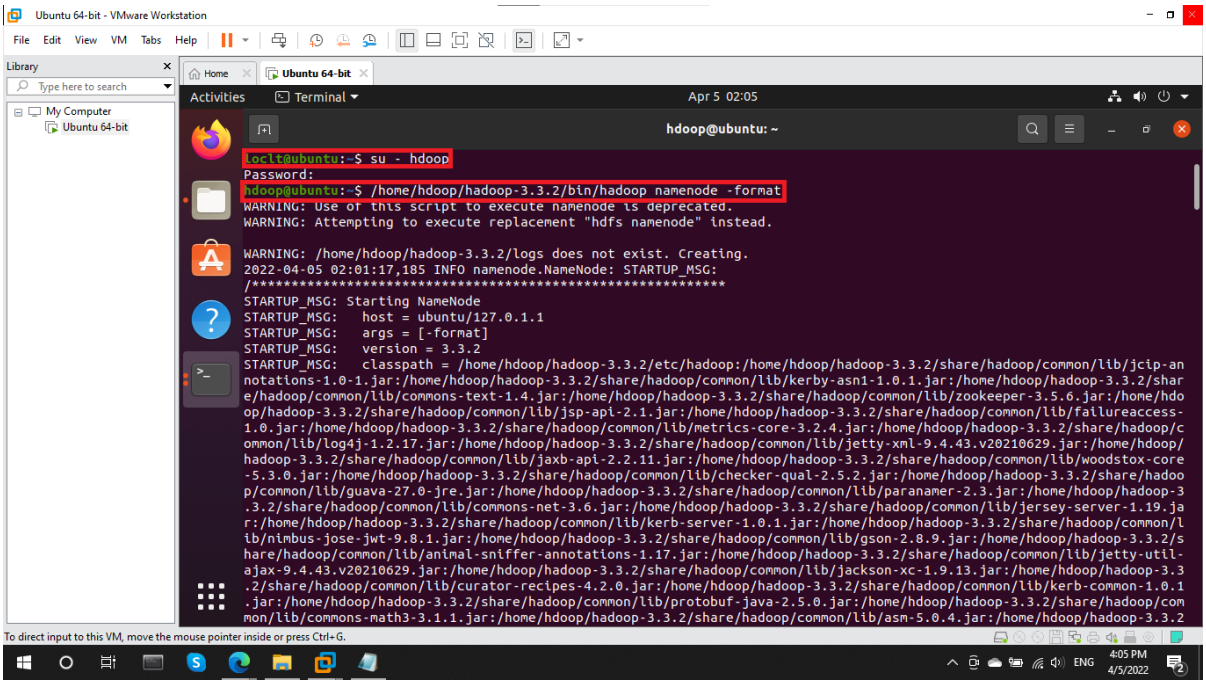


⇒ Hoàn tất các bước cài đặt Hadoop

II. Chạy thử Hadoop

1. Tạo user, chuyển quyền làm việc hiện tại sang tên user vừa tạo và khởi động **namenode**

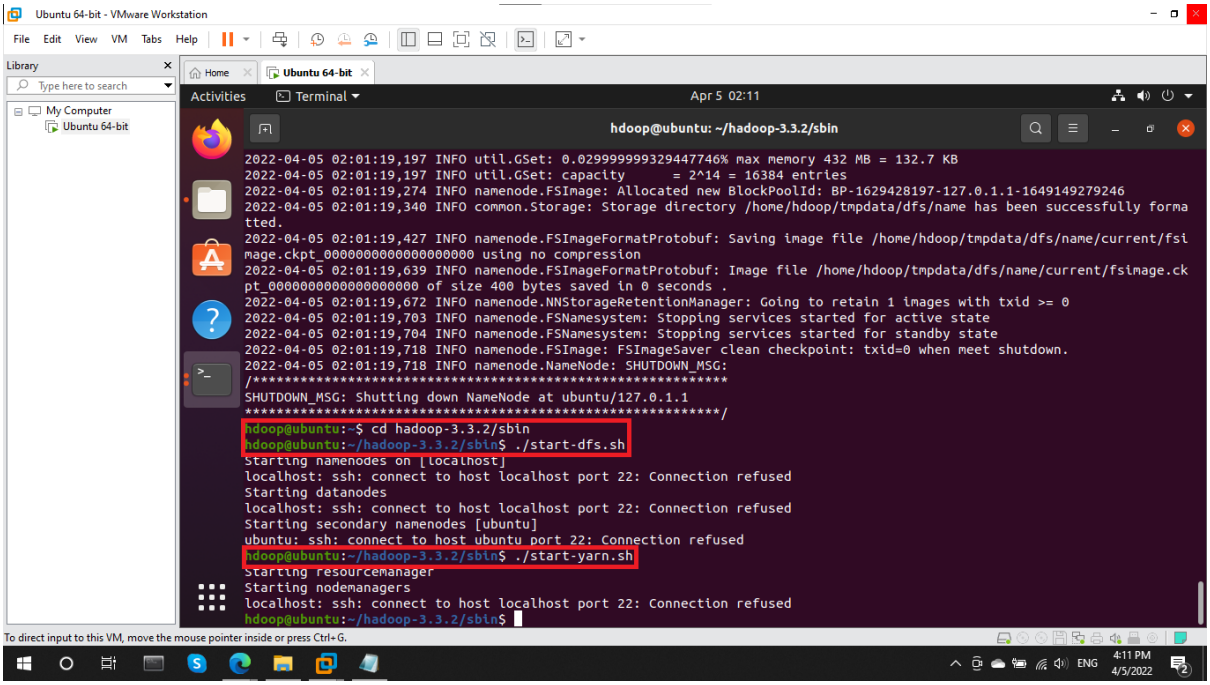




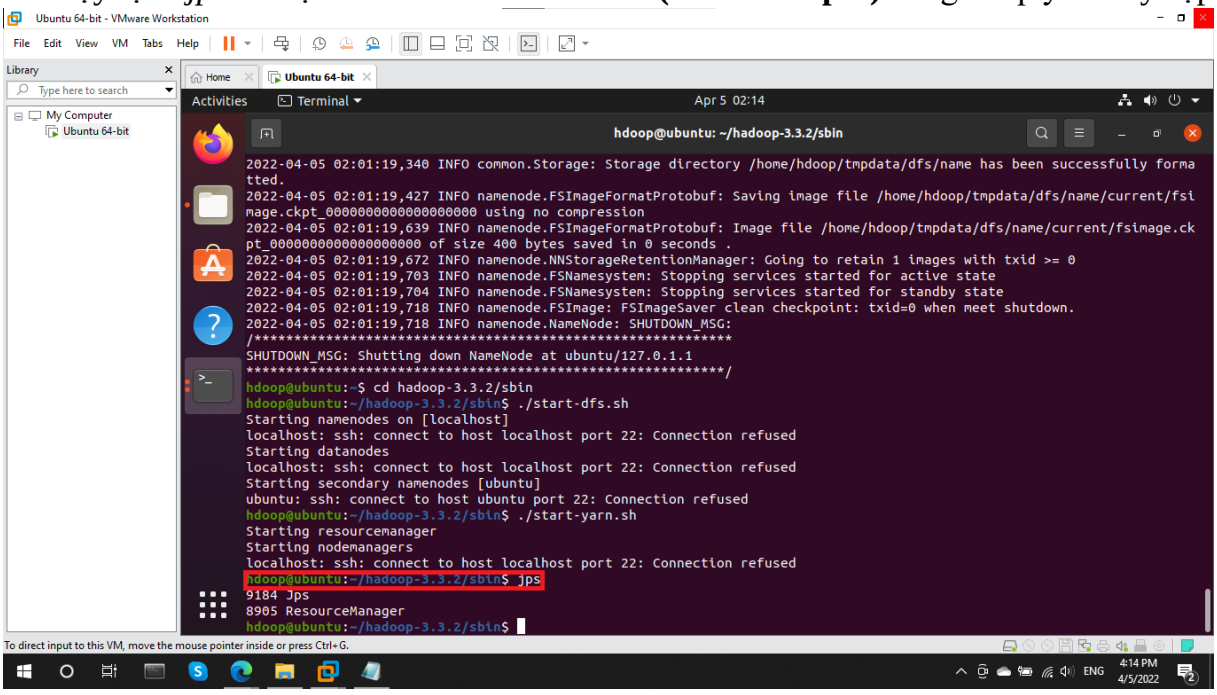
2. Khởi động **Hadoop Service**

- Chuyển đến thư mục **sbin** trong **hadoop-3.3.2**.

- Chạy 2 lệnh *./start-dfs.sh* và *./start-yarn.sh* để khởi động **NameNode**, **DataNode**, **ResourceManager** và **NodeManagers**



3. Chạy lệnh *jps* để liệt kê danh sách các **JVM (Java HotSpot)** đang có quyền truy cập



III. Tài liệu tham khảo:

[1] Powerpoint Hướng dẫn lab 1: Môi trường Hadoop.

[2] Hadoop Installation Guide Step by Step:
https://www.youtube.com/watch?v=Ih5cuJYYz6Y

[3] Guide file from video above:
https://drive.google.com/file/d/17B8ptBee2YCWsqXpTm1UASW5Tezua68O/view