

## Học phần Tích hợp và phân tích dữ liệu lớn: Bài thực hành số 1

*Phạm Tiến Lâm, Đặng Văn Báu*

### 1. Ôn tập một số lệnh linux cơ bản

- Các lệnh với thư mục: mkdir, cp -r, mv, ls, rm
- Xem nội dung file: more, less, head, tail, cat ...

#### *Activity 1.*

- Login vào hệ thống: ssh your\_ID@103.143.206.51 -p 21122  
Password mặc định: 12345
- Tạo một thư mục có tên : “**your\_ID**”
- Tạo một thư mục có tên là **big\_data** trên thư mục **your\_ID** của mình
- Di chuyển vào thư mục **big\_data**, thực hiện lệnh "ls -l" để xem nội dung của thư mục /usr, và lưu output bằng lệnh "ls -l" vào file a.txt
- Trong thư mục **big\_data** tạo các thư mục có cấu trúc case như sau



- Copy file cpufreq trong thư mục /proc vào các thư mục b, b11, b12, b1
- Sử dụng tổ hợp lệnh cat, grep, và wc để đếm số cpu của hệ thống

#### *Activity 2.*

- Download movielens 100K data set
- Copy data lên server  
scp -P 21122 ml-100k.zip your\_ID@103.143.206.51:/home/your\_ID/big\_data
- Unzip file ml-100k.zip trên hệ thống sandbox-hdp

## 2. Xử lý file và thư mục trên Hadoop HDFS

Một số lệnh xem nội dung, tạo thư mục, copy file lên HDFS

`hadoop fs -ls`

`hadoop fs -mkdir`

`hadoop fs -copyFromLocal`

`hadoop fs -rm`

`hadoop fs -rmdir`

### Activity 3.

- Login vào server
- Tạo thư mục **your\_ID** trên hệ thống HDFS: `hadoop fs -mkdir /user/your_ID`
- Trong thư mục **your\_ID** tạo các thư mục có cấu trúc case như sau trên hệ thống HDFS



- Copy file `cpuinfo` trong thư mục `/proc` vào các thư mục **b**, **b11**, **b12**, **b1** trên hệ thống HDFS
- Liệt kê nội dung của thư mục **b** trên hệ thống HDFS
- Copy file `u.data` vào thư mục **b**, **c**, **c1**, **cc** trên hệ thống HDFS

### 3. Ôn tập Java và vim

#### Activity 4

- Tạo một thư mục có tên là **day1** trong thư mục **big\_data**
- Dùng vim viết chương trình Java sau trong thư mục **day1**

```
public class HelloWorld {  
  
    /**  
     * @param args the command line arguments  
     */  
    public static void main(String[] args) {  
        // TODO code application logic here  
        System.out.println("Hello there!");  
        System.out.println(args[1]);  
  
        // type variableName = value;  
        int a = 5;  
        a = 5 * 2;  
  
        System.out.println(a);  
  
        //final float b = 2.5f;  
        float c = 4.5f;  
  
        System.out.println(c);  
  
    }  
}
```

- Dịch và chạy chương trình java trên

