BÀI TẬP TRÊN LỚP

MÔN HỌC: LƯU TRỮ VÀ XỬ LÝ DỮ LIỆU LỚN

LAB 1: HDFS

NHÓM: GIỮA CHÚNG TA

1. Kết quả cài đặt

1.1. Configure the System

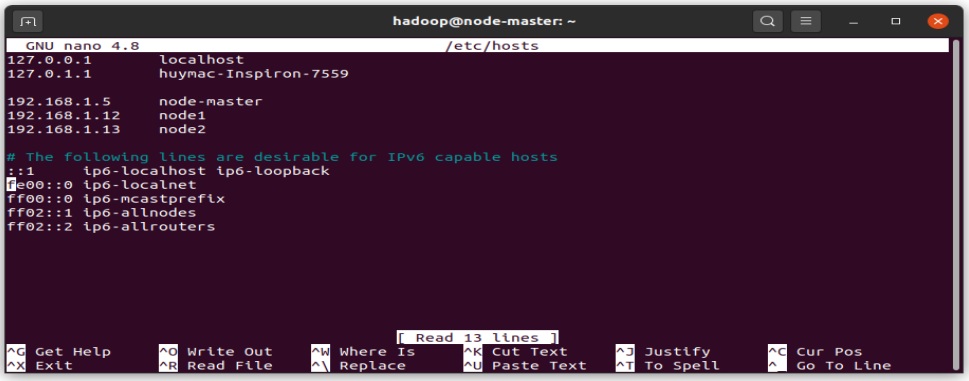

Figure 1:Hosts file

1.2. Set Evironment Variables

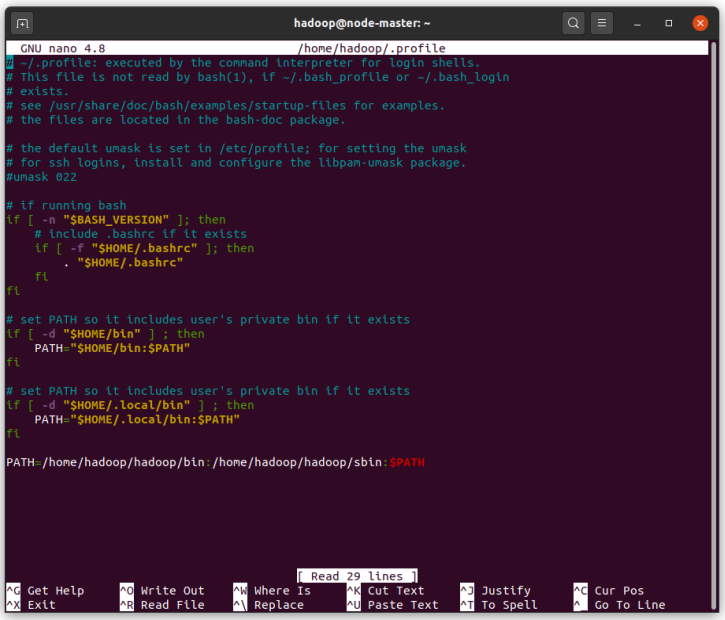

Figure 2:Add Hadoop binaries to PATH

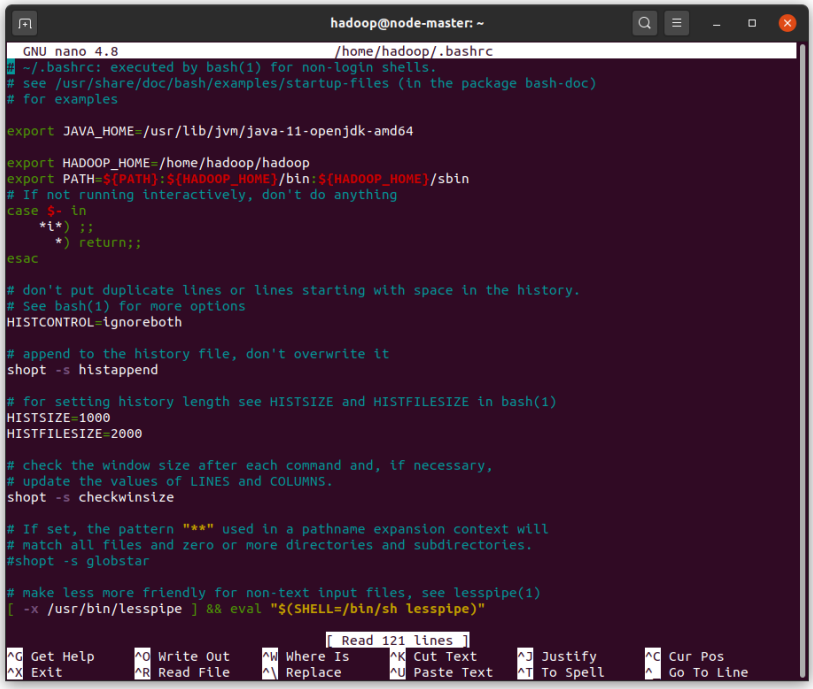

Figure 3: Add Hadoop to your PATH for the shell

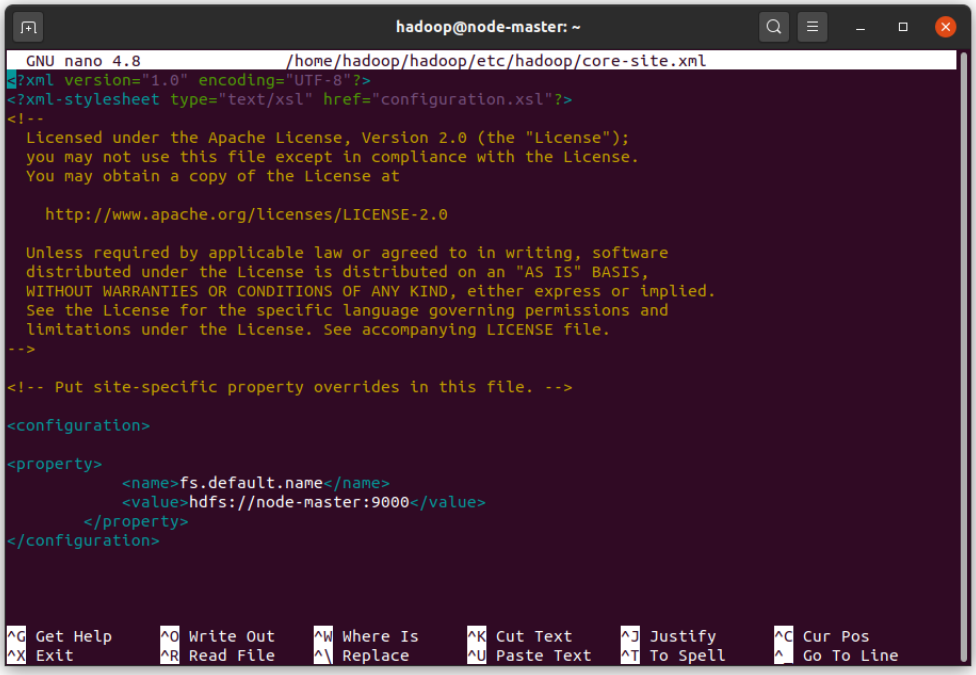

Figure 4: Set NameNode Location

- *Dfs.replication = 2*

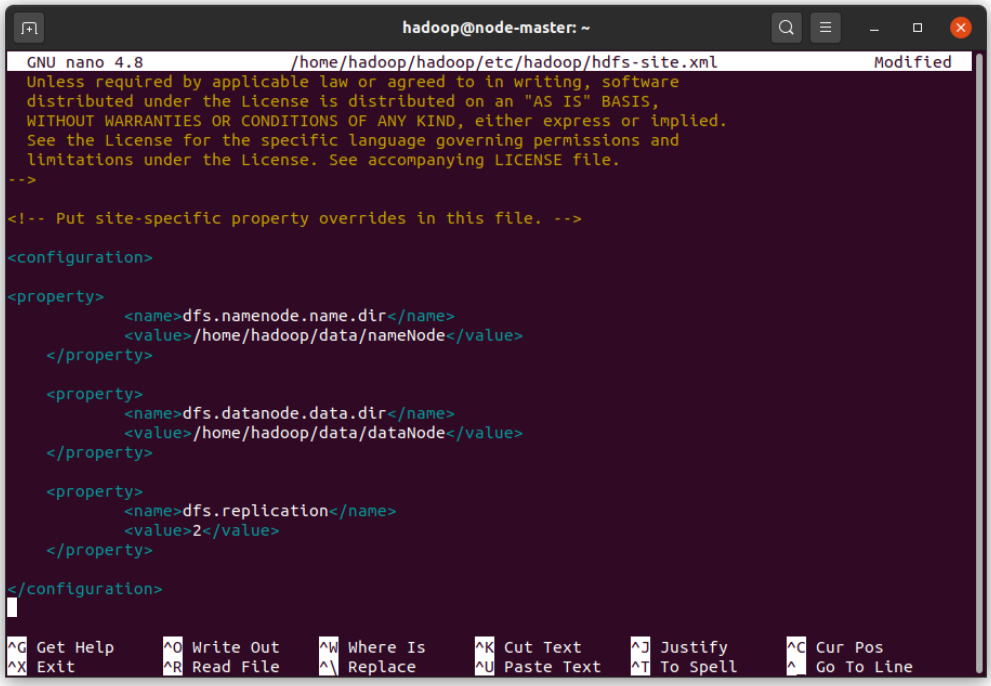

Figure 5: Set path for HDFSPermalink

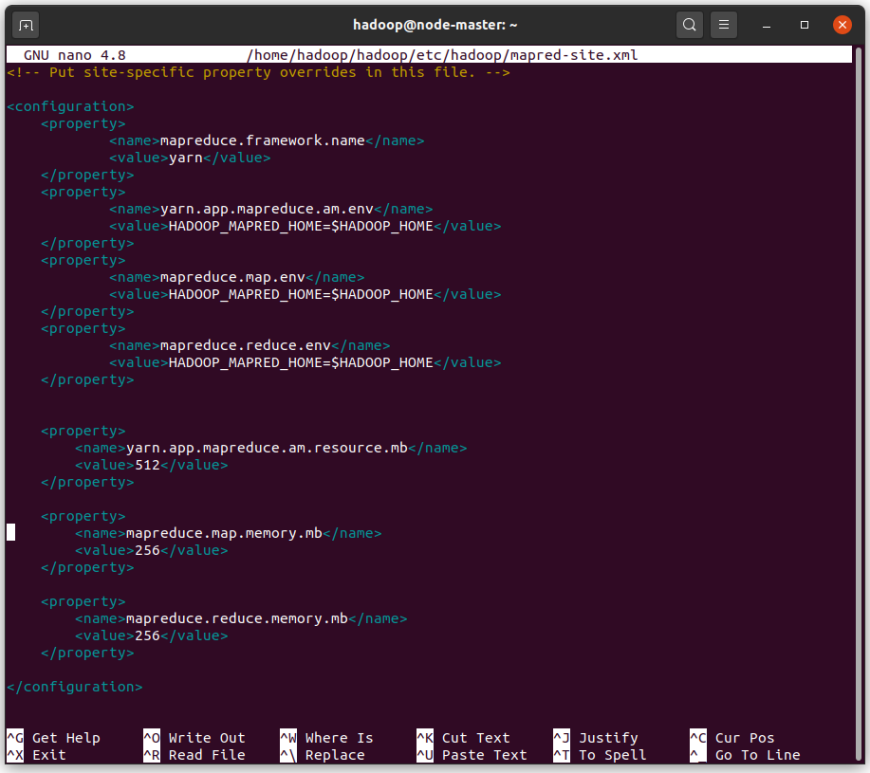

Figure 6: Set YARN as Job Scheduler

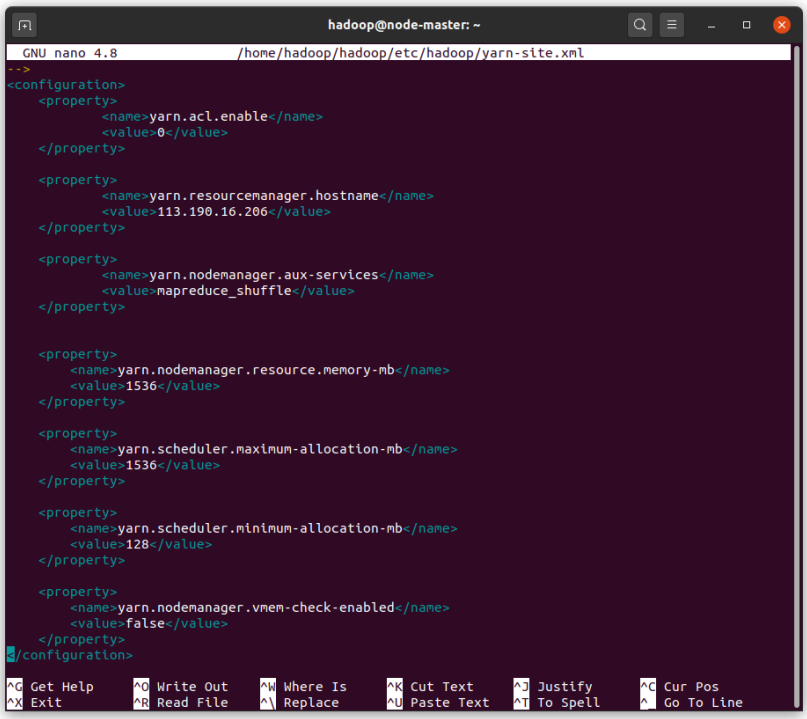

Figure 7: Configure YARN

1.3. Configure Memory Allocation

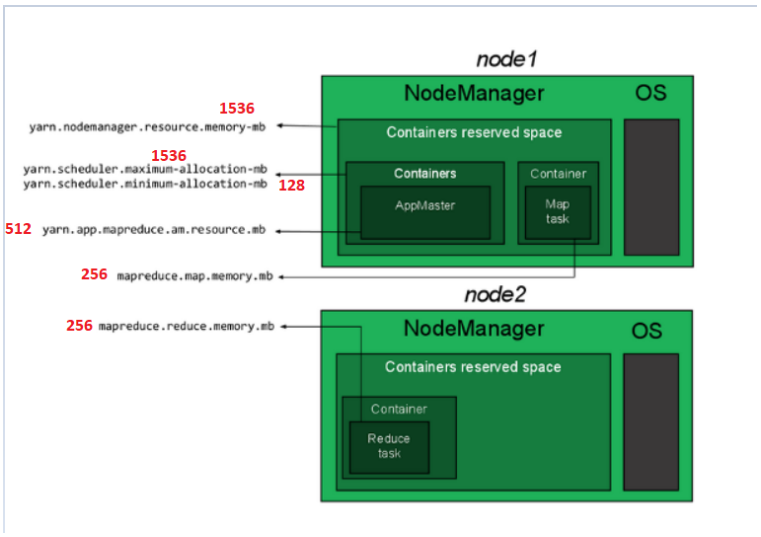

Figure 8: Sample config for 2GB Nodes (will change in next lab)

1.4. Duplicate Config Files On Each Node

- First, install ssh
- Generate an SSH key
- Make a new file master.pub in the /home/hadoop/.ssh. Paste your public key into
- Copy key:

```
cat ~/.ssh/master.pub >> ~/.ssh/authorized_keys
```

- Note:

- We use this command for ssh node1, node2 without rq passwd

```
Ssh-copy-id –I ~/ssh/id_rsa.pub hadoop@note-master

Ssh-copy-id –I ~/ssh/id_rsa.pub hadoop@note1

Ssh-copy-id –I ~/ssh/id_rsa.pub hadoop@note2
```

- Importance: Config username of node-master, node1, node2 is ***hadoop***

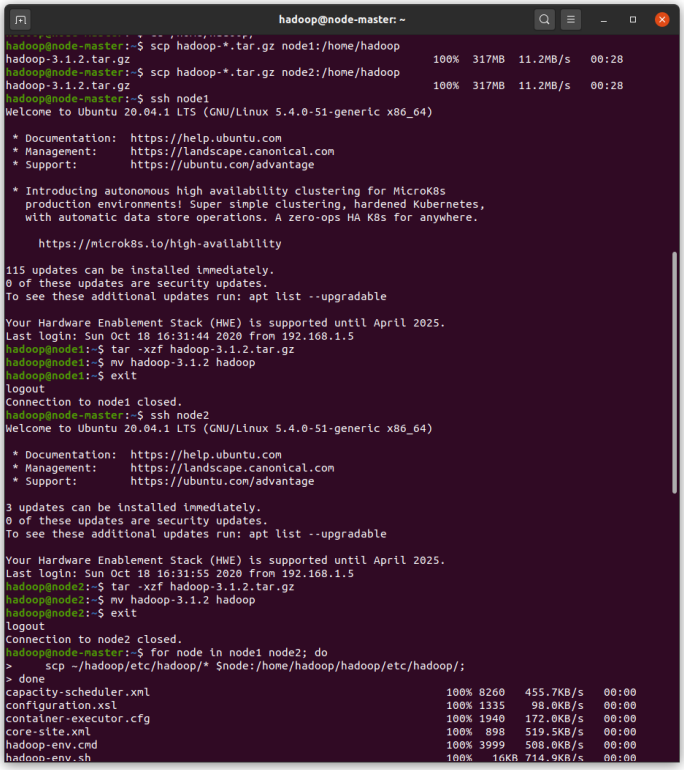

Figure 9: Copy files to node1, node2

1.5. Start HDFS

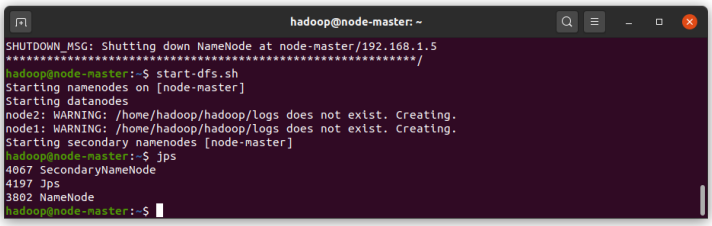

Figure 10: Start the HDFS by running the following script from node-master

- Or can: $ start-all.sh

2. Monitor HDFS Cluter

2.1. Put Data to HDFS

```
-   hdfs dfs -mkdir books
```

- use wget to download file, see here ☺: https://speed.hetzner.de/ for what ever file size you want

```
-   hdfs dfs -put 500MB.bin books
-   hdfs dfs -put 500MB.bin books
```

2.2. View web user interface

- http://192.168.1.5:9870 (192.168.1.5 is my master node id)

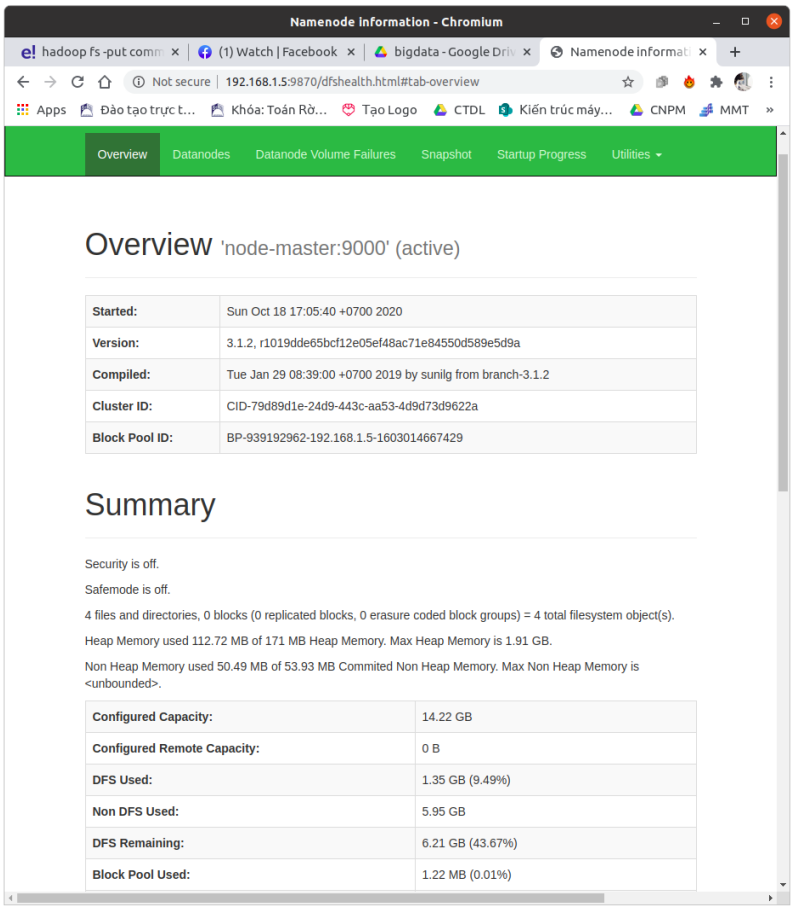

Figure 11: Web user interface

どうもありがとう、先生