

BÁO CÁO LAB1 LƯU TRỮ XỬ LÝ DỮ LIỆU LỚN : CÀI ĐẶT HDFS

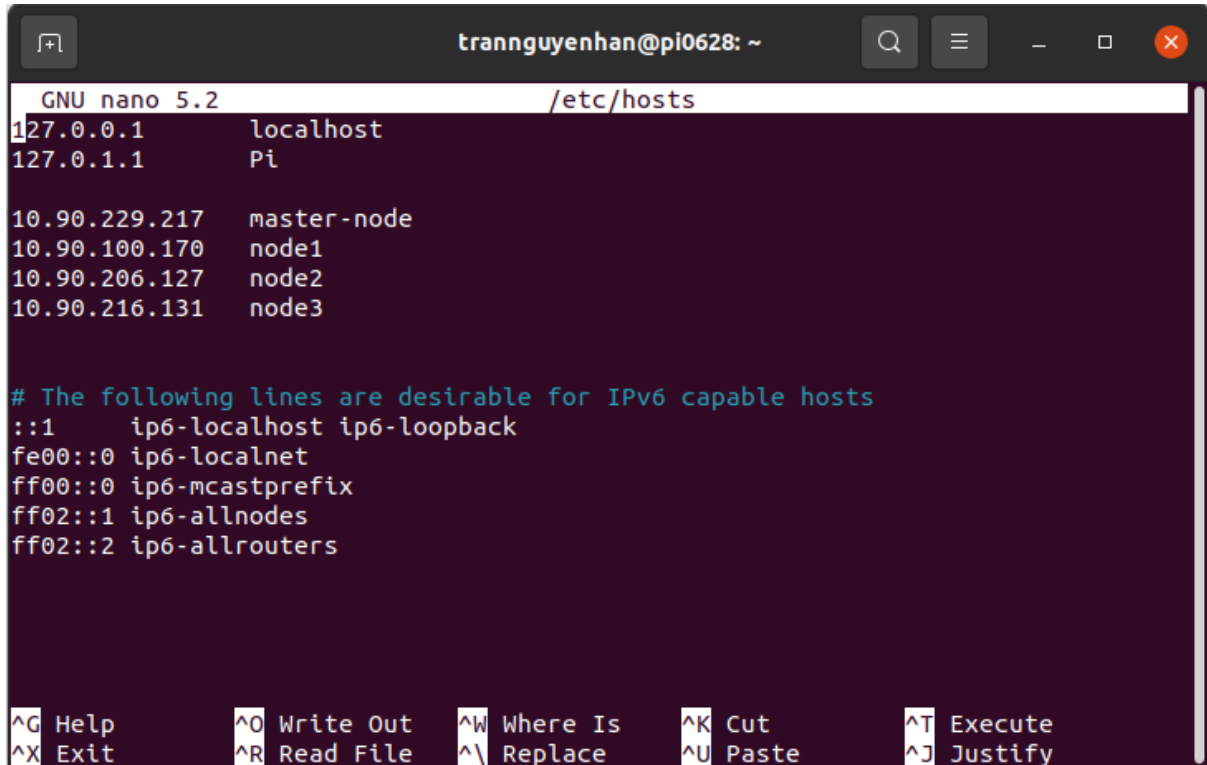


Nhóm	MTHH	
Thành viên	Nguyễn Quang Huy	20183554
	Trần Quang Minh	20183594
	Ngô Song Việt Hoàng	20183542
	Nguyễn Văn Thanh	20183632
GVHD	TS. Đào Thành Chung	

1. Mô tả kết quả cài đặt

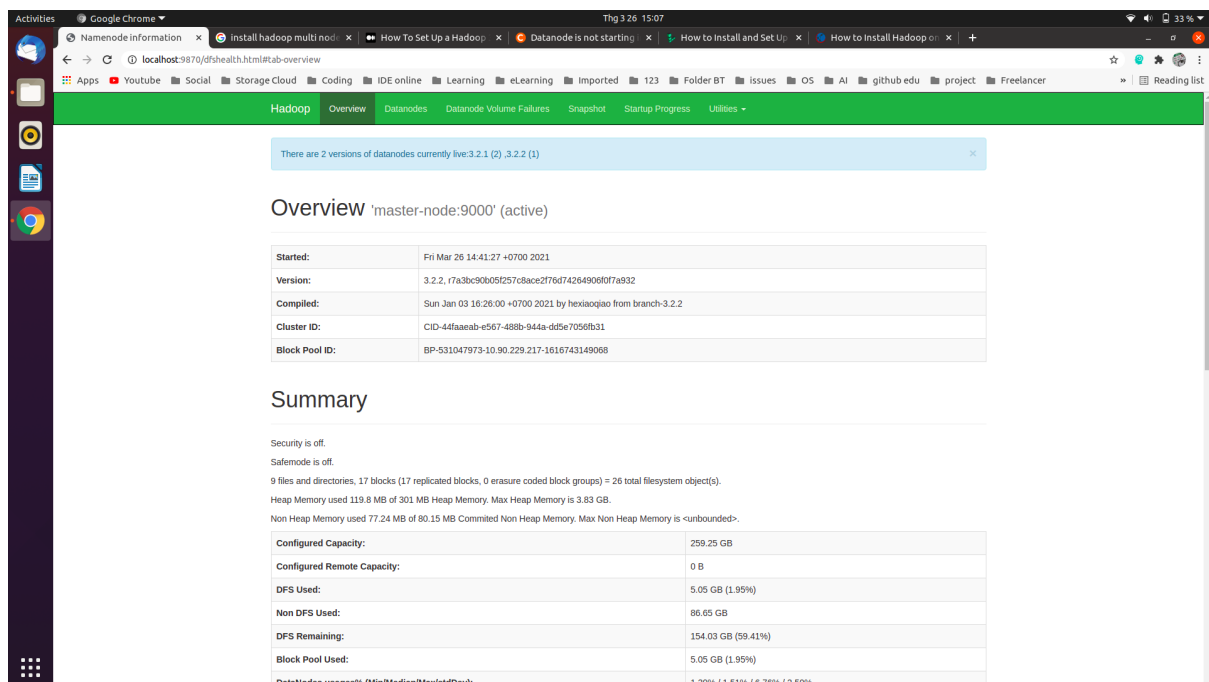
Nhóm cấu hình thành công HDFS trên 4 máy với 3 máy là datanode.

Ảnh file cấu hình các node :



```
tranguyenhan@pi0628: ~  
GNU nano 5.2 /etc/hosts  
127.0.0.1 localhost  
127.0.1.1 pi  
  
10.90.229.217 master-node  
10.90.100.170 node1  
10.90.206.127 node2  
10.90.216.131 node3  
  
# The following lines are desirable for IPv6 capable hosts  
::1 ip6-localhost ip6-loopback  
fe00::0 ip6-localnet  
ff00::0 ip6-mcastprefix  
ff02::1 ip6-allnodes  
ff02::2 ip6-allrouters  
  
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute  
^X Exit      ^R Read File  ^_ Replace    ^U Paste      ^J Justify
```

Tổng quan về cấu hình cài đặt sau khi chạy thành công (**live node** = 3) :



There are 2 versions of datanodes currently live: 3.2.1 (2), 3.2.2 (1)

Overview 'master-node:9000' (active)

Started:	Fri Mar 26 14:41:27 +0700 2021
Version:	3.2.2, (7a3bc90b05f257c8ace2f76d74264906f0f7a932)
Compiled:	Sun Jan 03 16:26:00 +0700 2021 by hexiaojiao from branch-3.2.2
Cluster ID:	CID-44faaeab-e567-488b-944a-dd5e7056fb31
Block Pool ID:	BP-531047973-10.90.229.217-1616743149068

Summary

Security is off.
Safemode is off.
9 files and directories, 17 blocks (17 replicated blocks, 0 erasure coded block groups) = 26 total filesystem object(s).
Heap Memory used 119.8 MB of 301 MB Heap Memory. Max Heap Memory is 3.83 GB.
Non Heap Memory used 77.24 MB of 80.15 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	259.25 GB
Configured Remote Capacity:	0 B
DFS Used:	5.05 GB (1.95%)
Non DFS Used:	86.65 GB
DFS Remaining:	154.03 GB (59.41%)
Block Pool Used:	5.05 GB (1.95%)
DataNodes usages% (Min/Median/Max/stdDev):	1.39% / 1.51% / 6.76% / 2.50%

Activities

Google Chrome

Thg 3 26 15:08

localhost:9870/dfshealth.html#tab-overview

install hadoop multi nod... How To Set Up a Hadoop... Datanode is not starting... How to Install and Set Up... How to Install Hadoop... project Freelancer

Summary

Security is off.

Safemode is off.

9 files and directories, 17 blocks (17 replicated blocks, 0 erasure coded block groups) = 26 total filesystem object(s).

Heap Memory used 119.8 MB of 301 MB Heap Memory. Max Heap Memory is 3.83 GB.

Non Heap Memory used 77.24 MB of 80.15 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	259.25 GB
Configured Remote Capacity:	0 B
DFS Used:	5.05 GB (1.95%)
Non DFS Used:	86.65 GB
DFS Remaining:	154.03 GB (59.41%)
Block Pool Used:	5.05 GB (1.95%)
DataNodes usages% (Min/Median/Max/stdDev):	1.39% / 1.51% / 6.76% / 2.50%
Live Nodes	3 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Fri Mar 26 14:41:27 +0700 2021
Last Checkpoint Time	Fri Mar 26 14:19:09 +0700 2021
Enabled Erasure Coding Policies	RS-6-3-1024k

Activities

Google Chrome

Thg 3 26 15:08

localhost:9870/dfshealth.html#tab-datanode

install hadoop multi nod... How To Set Up a Hadoop... Datanode is not starting... How to Install and Set Up... How to Install Hadoop... project Freelancer

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Datanode Information

✓ In service

● Down

✓ Decommissioning

● Decommissioned

● Decommissioned & dead

✓ Entering Maintenance

● In Maintenance

● In Maintenance & dead

Datanode usage histogram

In operation

Show 25 entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓ node1-9866 (10.90.100.170:9866)	http://node1-9864	1s	26m	122.53 GB	17	1.7 GB (1.39%)	3.2.1
✓ node2-9866 (10.90.216.131:9866)	http://node2-9864	1s	26m	24.47 GB	16	1.65 GB (6.76%)	3.2.2
✓ viethoang-9866 (10.90.206.127:9866)	http://viethoang-9864	0s	26m	112.25 GB	17	1.69 GB (1.51%)	3.2.1

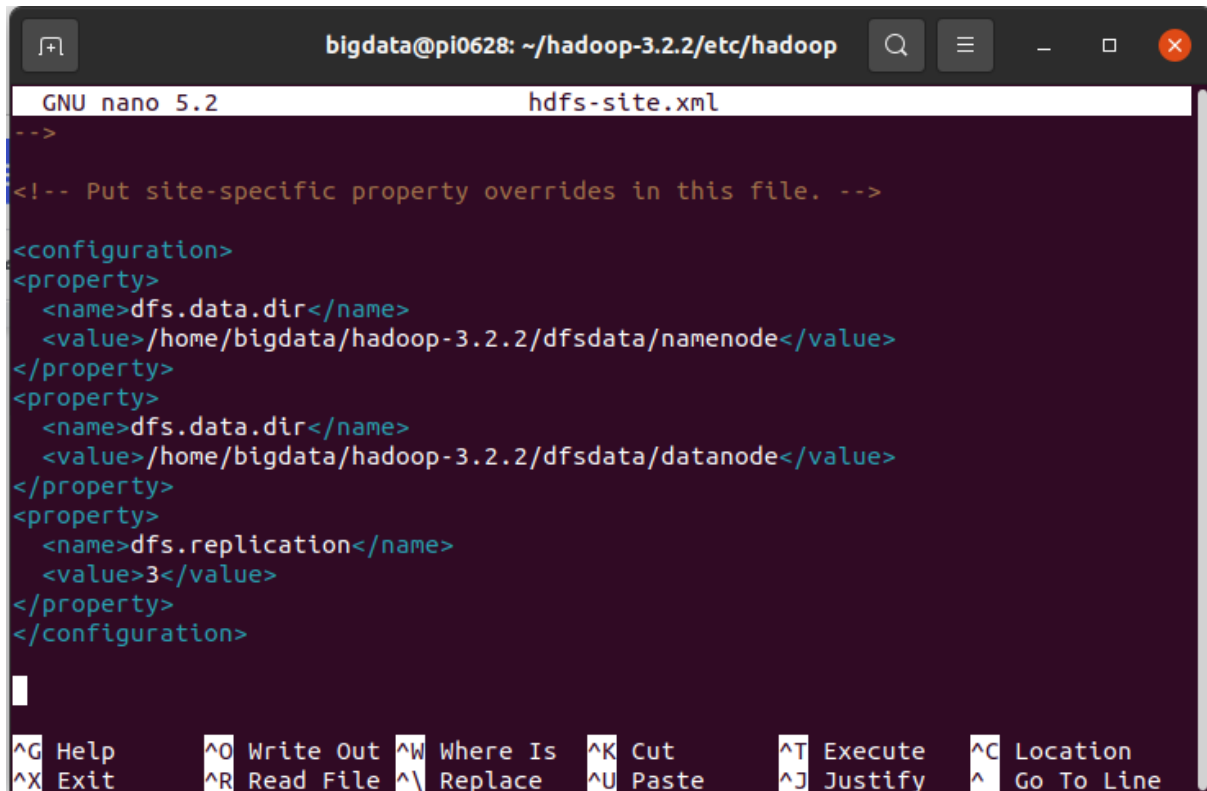
Showing 1 to 3 of 3 entries

Previous 1 Next

Entering Maintenance

2. Bằng chứng tạo bản sao

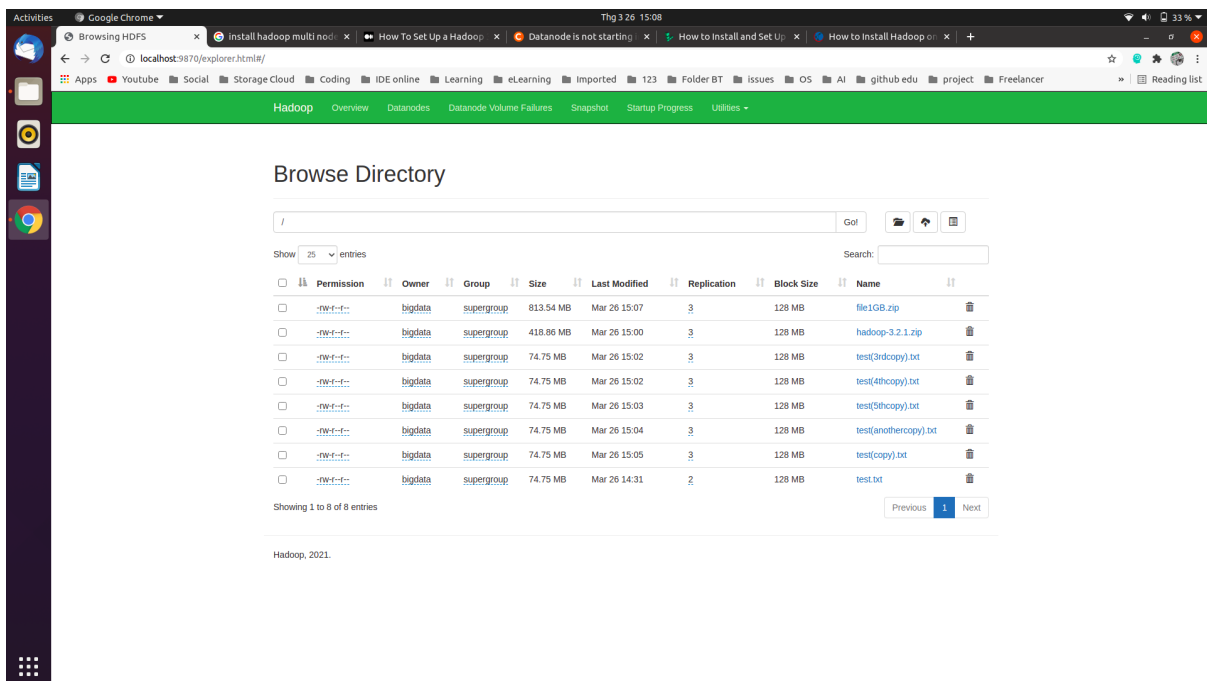
Nhóm đã cấu hình tạo ra 3 bản sao :



```
bigdata@pi0628: ~/hadoop-3.2.2/etc/hadoop
GNU nano 5.2 hdfs-site.xml
-->
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
  <name>dfs.data.dir</name>
  <value>/home/bigdata/hadoop-3.2.2/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/bigdata/hadoop-3.2.2/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
</configuration>
```

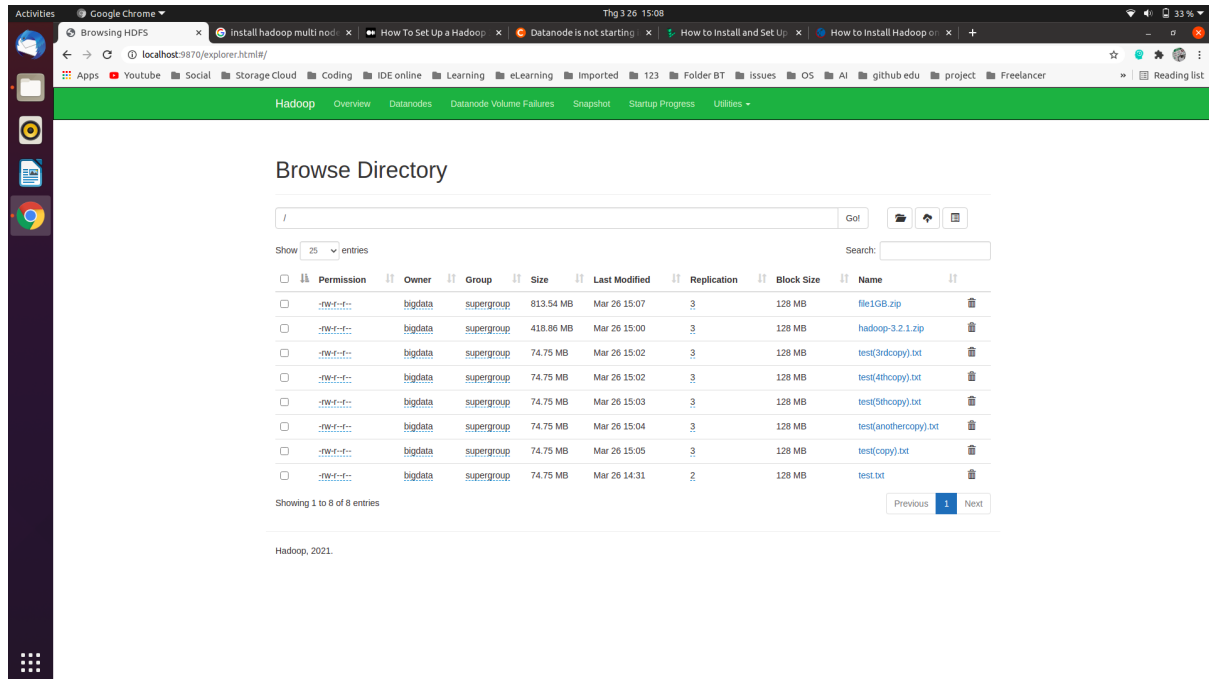
Replication = 3



Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	bigdata	supergroup	813.54 MB	Mar 26 15:07	3	128 MB	file1GB.zip
-rw-r--r--	bigdata	supergroup	418.86 MB	Mar 26 15:00	3	128 MB	hadoop-3.2.1.zip
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 15:02	3	128 MB	test(3rdcopy).txt
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 15:02	3	128 MB	test(4thcopy).txt
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 15:03	3	128 MB	test(5thcopy).txt
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 15:04	3	128 MB	test(anothercopy).txt
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 15:05	3	128 MB	test(copy).txt
-rw-r--r--	bigdata	supergroup	74.75 MB	Mar 26 14:31	2	128 MB	test.txt

3. Bằng chứng lưu trữ 1GB dữ liệu

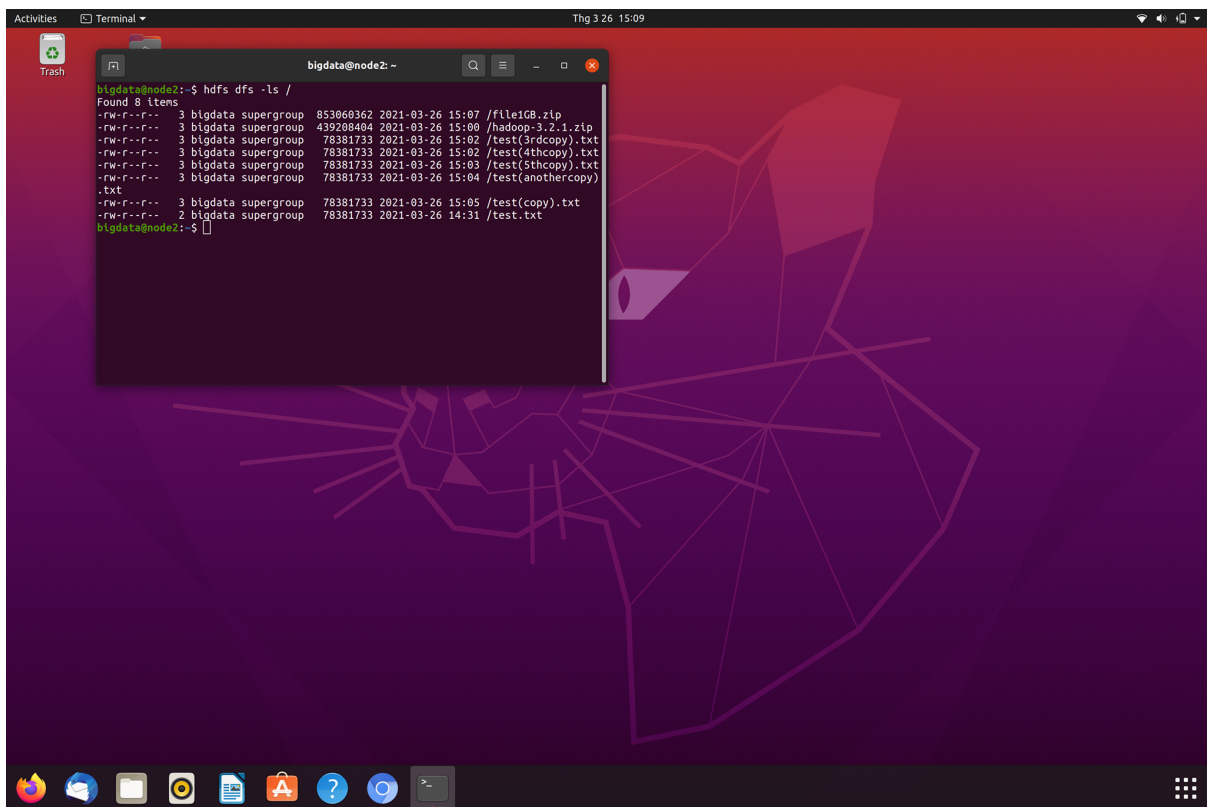
Hình ảnh lưu trữ 1GB dữ liệu trên giao diện UI của HDFS trên website tại đường dẫn localhost:9870 của máy pi0628 (master-node) :



Trên máy viethoang (node3) :

```
bigdata@viethoang: ~/hadoop-3.2.2/dfsdata
bigdata@viethoang:~/hadoop-3.2.2/dfsdata$ jps
15751 Jps
15192 DataNode
bigdata@viethoang:~/hadoop-3.2.2/dfsdata$ jps
16537 DataNode
17551 Jps
bigdata@viethoang:~/hadoop-3.2.2/dfsdata$ hdfs dfs -ls /
Found 8 items
-rw-r--r--  3 bigdata supergroup  853060362 2021-03-26 15:07 /file1GB.zip
-rw-r--r--  3 bigdata supergroup  439208404 2021-03-26 15:00 /hadoop-3.2.1.zip
-rw-r--r--  3 bigdata supergroup   78381733 2021-03-26 15:02 /test(3rdcopy).txt
-rw-r--r--  3 bigdata supergroup   78381733 2021-03-26 15:02 /test(4thcopy).txt
-rw-r--r--  3 bigdata supergroup   78381733 2021-03-26 15:03 /test(5thcopy).txt
-rw-r--r--  3 bigdata supergroup   78381733 2021-03-26 15:04 /test(anothercopy)
.txt
-rw-r--r--  3 bigdata supergroup   78381733 2021-03-26 15:05 /test(copy).txt
-rw-r--r--  2 bigdata supergroup   78381733 2021-03-26 14:31 /test.txt
bigdata@viethoang:~/hadoop-3.2.2/dfsdata$
```

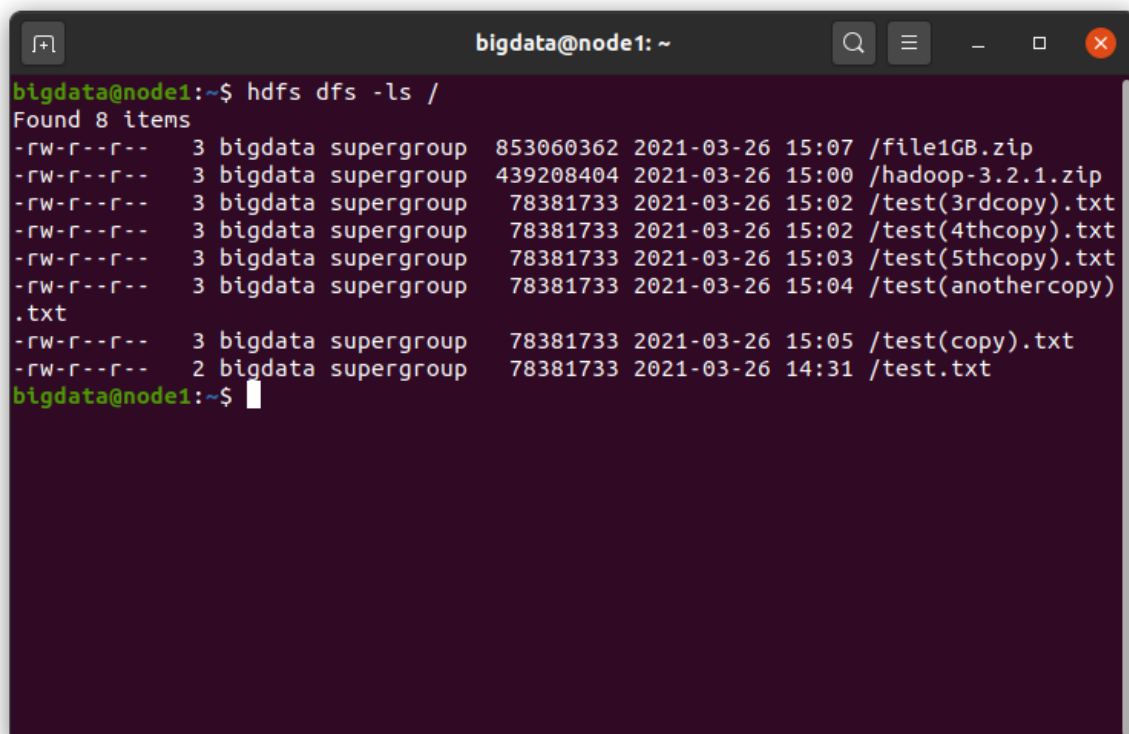
Trên máy node2 (node2) :



The screenshot shows a terminal window titled 'bigdata@node2:~' with the command 'hdfs dfs -ls /' executed. The output lists 8 items in the root directory of the HDFS filesystem. The background of the desktop is a purple and red geometric pattern.

```
bigdata@node2:~$ hdfs dfs -ls /
Found 8 items
-rw-r--r-- 3 bigdata supergroup 853060362 2021-03-26 15:07 /file1GB.zip
-rw-r--r-- 3 bigdata supergroup 439208404 2021-03-26 15:00 /hadoop-3.2.1.zip
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:02 /test(3rdcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:02 /test(4thcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:03 /test(5thcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:04 /test(anothercopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:05 /test(copy).txt
-rw-r--r-- 2 bigdata supergroup 78381733 2021-03-26 14:31 /test.txt
bigdata@node2:~$
```

Trên máy node1 (node1) :



The screenshot shows a terminal window titled 'bigdata@node1:~' with the command 'hdfs dfs -ls /' executed. The output lists 8 items in the root directory of the HDFS filesystem, matching the output from node2.

```
bigdata@node1:~$ hdfs dfs -ls /
Found 8 items
-rw-r--r-- 3 bigdata supergroup 853060362 2021-03-26 15:07 /file1GB.zip
-rw-r--r-- 3 bigdata supergroup 439208404 2021-03-26 15:00 /hadoop-3.2.1.zip
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:02 /test(3rdcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:02 /test(4thcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:03 /test(5thcopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:04 /test(anothercopy).txt
-rw-r--r-- 3 bigdata supergroup 78381733 2021-03-26 15:05 /test(copy).txt
-rw-r--r-- 2 bigdata supergroup 78381733 2021-03-26 14:31 /test.txt
bigdata@node1:~$
```