Course: Big Data

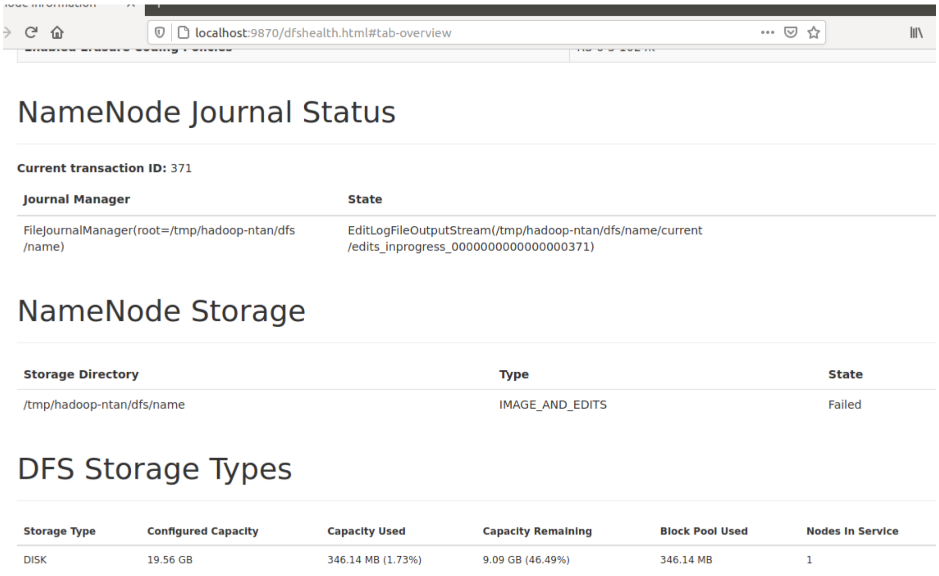
Lab 01

Set up Hadoop

Fill answers of the questions below in the given tables.
Your screenshots must contain commands for required operations.

Question 1:

Set up Hadoop following this tutorial and then take a screenshot of the hdfs-site (localhost:9870) which shows information for NameNode Journal Status, NameNode Storage, DFS Storage Types (Overview page).
For example,



localhost:9870/dfshealth.html#tab-overview

NameNode Journal Status

Current transaction ID: 371

Journal Manager	State
FileJournalManager(root=/tmp/hadoop-ntan/dfs /name)	EditLogFileOutputStream(/tmp/hadoop-ntan/dfs/name/current /edits_inprogress_0000000000000000371)

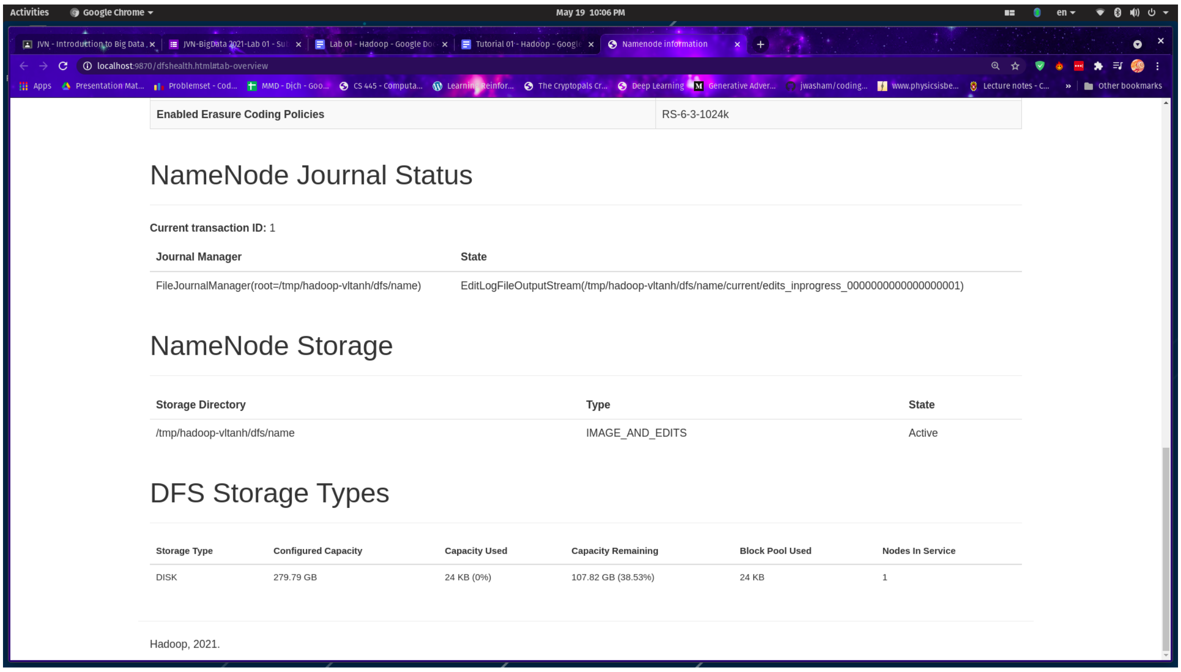
NameNode Storage

Storage Directory	Type	State
/tmp/hadoop-ntan/dfs/name	IMAGE_AND_EDITS	Failed

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	19.56 GB	346.14 MB (1.73%)	9.09 GB (46.49%)	346.14 MB	1

Your screenshot goes here



Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 1

Journal Manager	State
FileJournalManager(root=/tmp/hadoop-vltanh/dfs/name)	EditLogFileOutputStream(/tmp/hadoop-vltanh/dfs/name/current/edits_inprogress_0000000000000000001)

NameNode Storage

Storage Directory	Type	State
/tmp/hadoop-vltanh/dfs/name	IMAGE_AND_EDITS	Active

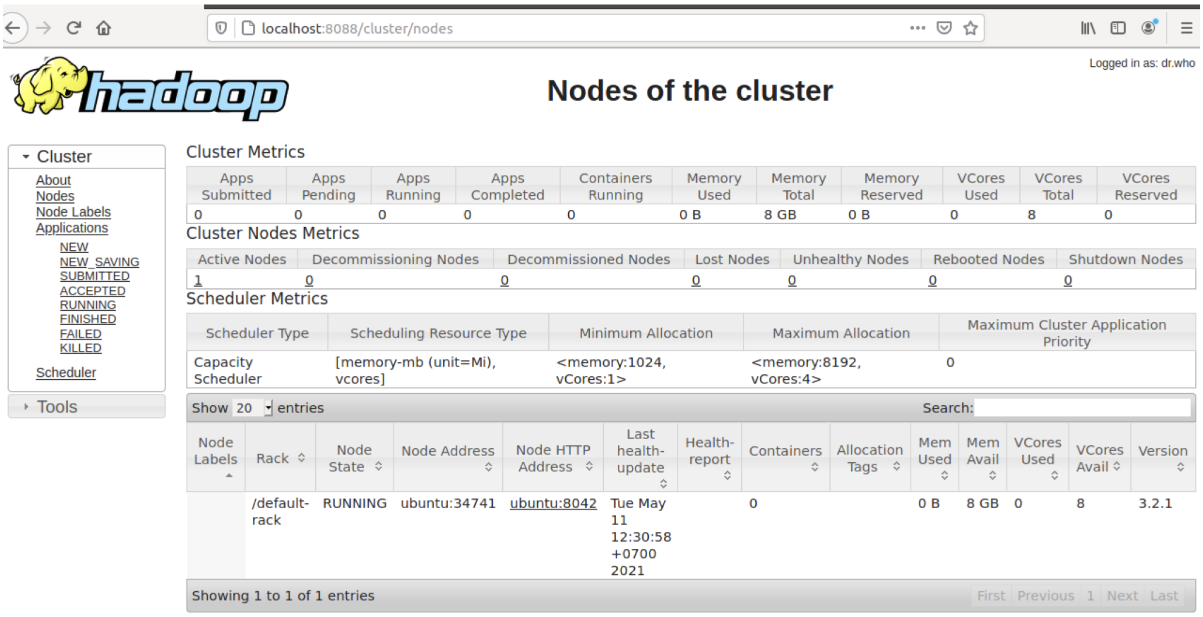
DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	279.79 GB	24 KB (0%)	107.82 GB (38.53%)	24 KB	1

Hadoop, 2021.

Question 2:

Set up Hadoop following this tutorial and then take a screenshot of the yarn-site (localhost:8088) which shows information for nodes (http://localhost:8088/cluster/nodes). For example,



Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

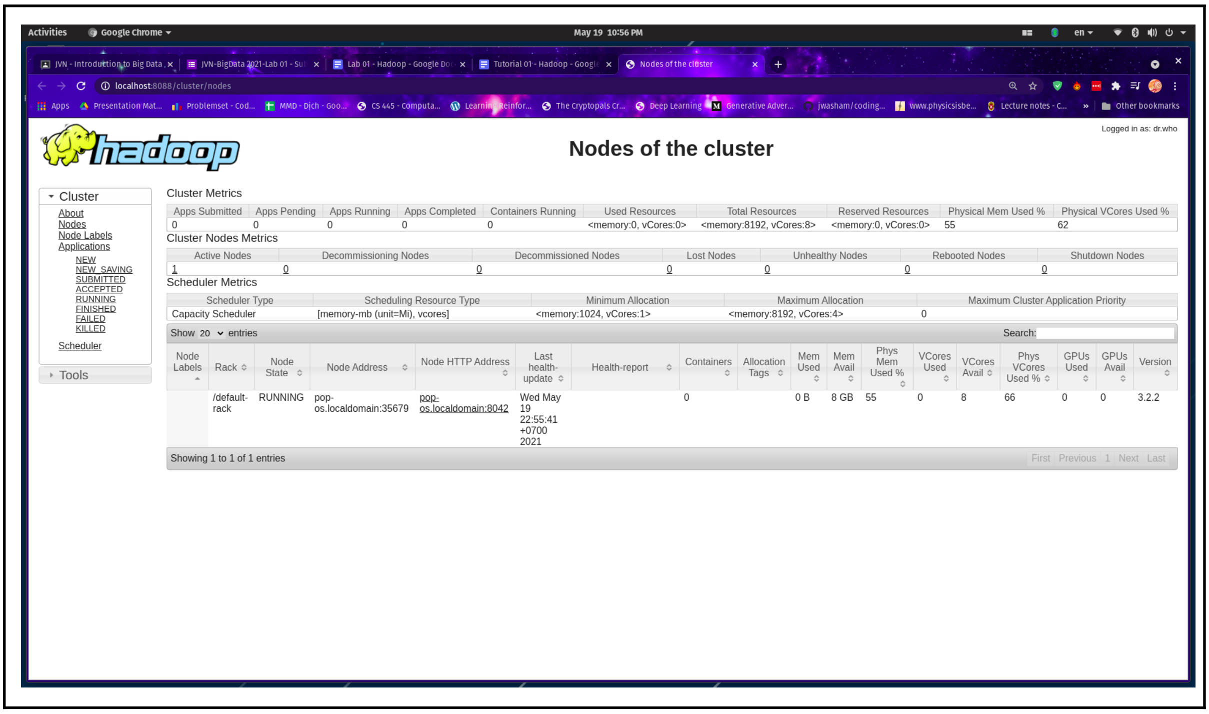
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries Search:

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	VCores Used	VCores Avail	Version
	/default-rack	RUNNING	ubuntu:34741	ubuntu:8042	Tue May 11 12:30:58 +0700 2021		0		0 B	8 GB	0	8	3.2.1

Showing 1 to 1 of 1 entries First Previous 1 Next Last

Your screenshot goes here



Logged in as: dr.who

Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
0	0	0	0	0	<memory:0, vCores:0>	<memory:8192, vCores:8>	<memory:0, vCores:0>	55	62

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries Search:

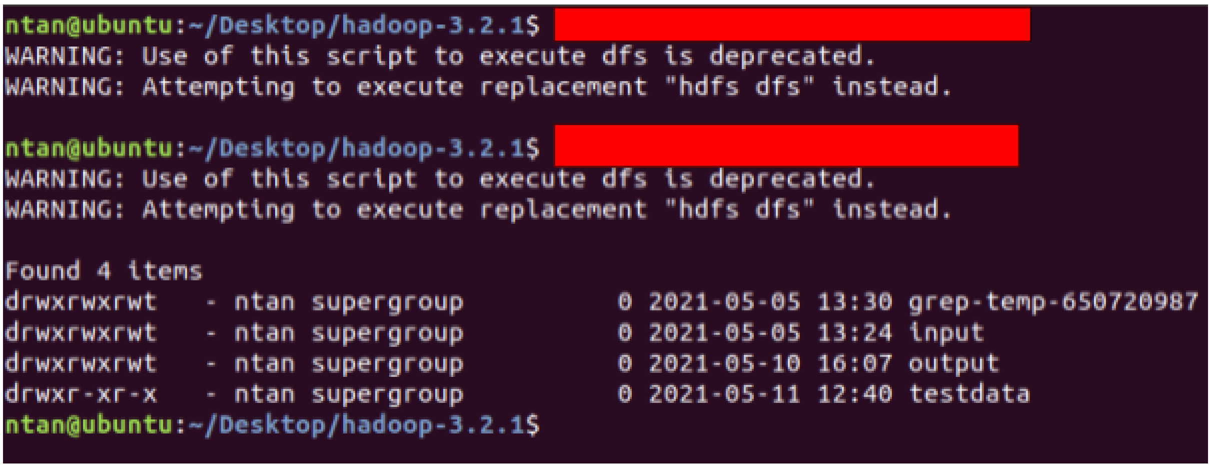
Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys VCores Used %	GPUs Used	GPUs Avail	Version
	/default-rack	RUNNING	pop-os.localdomain:35679	pop-os.localdomain:8042	Wed May 19 22:55:41 +0700 2021		0		0 B	8 GB	55	0	8	66	0	0	3.2.2

Showing 1 to 1 of 1 entries First Previous 1 Next Last

Question 3:

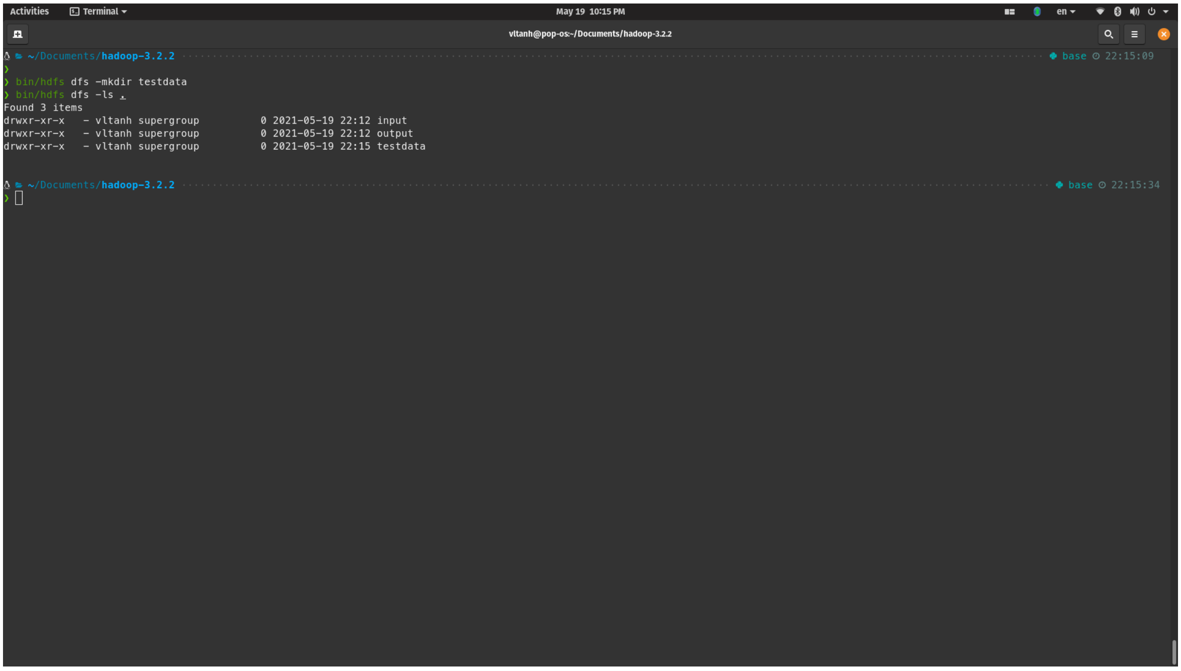
Create a folder, named **testdata**, using HDFS commands and then show list of files/folders in the default folder (assume that you had already created the folder /user/<username> following this tutorial).

For example,



```
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Found 4 items
drwxrwxrwt   - ntan supergroup          0 2021-05-05 13:30 grep-temp-650720987
drwxrwxrwt   - ntan supergroup          0 2021-05-05 13:24 input
drwxrwxrwt   - ntan supergroup          0 2021-05-10 16:07 output
drwxr-xr-x   - ntan supergroup          0 2021-05-11 12:40 testdata
ntan@ubuntu:~/Desktop/hadoop-3.2.1$
```

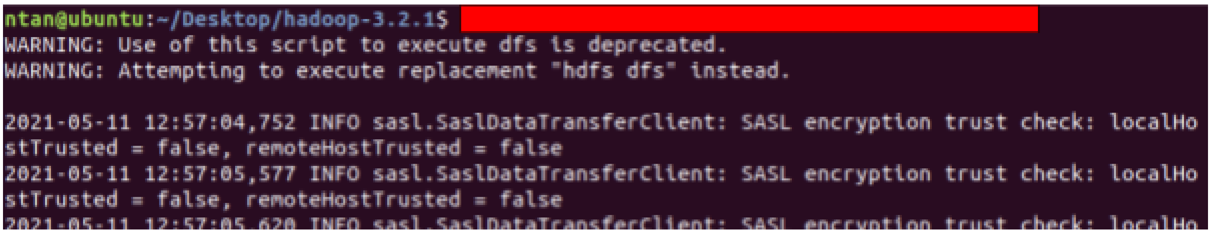
Your screenshot goes here



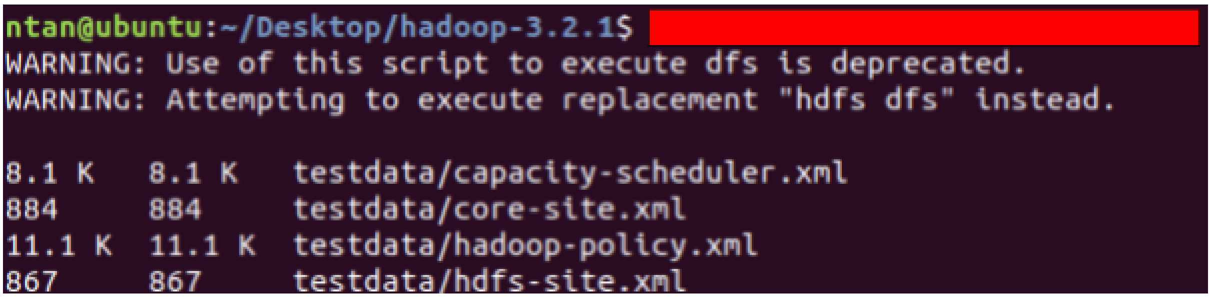
```
~/Documents/hadoop-3.2.2 ·························································· base 22:15:09
❯
❯ bin/hdfs dfs -mkdir testdata
❯ bin/hdfs dfs -ls .
Found 3 items
drwxr-xr-x   - vltanh supergroup          0 2021-05-19 22:12 input
drwxr-xr-x   - vltanh supergroup          0 2021-05-19 22:12 output
drwxr-xr-x   - vltanh supergroup          0 2021-05-19 22:15 testdata


~/Documents/hadoop-3.2.2 ·························································· base 22:15:34
❯
```

Question 4:

Use HDFS commands to copy all xml files from **<Hadoop folder>/etc/hadoop** to folder **testdata/** and then show the size of each file in **testdata/**.
For example,

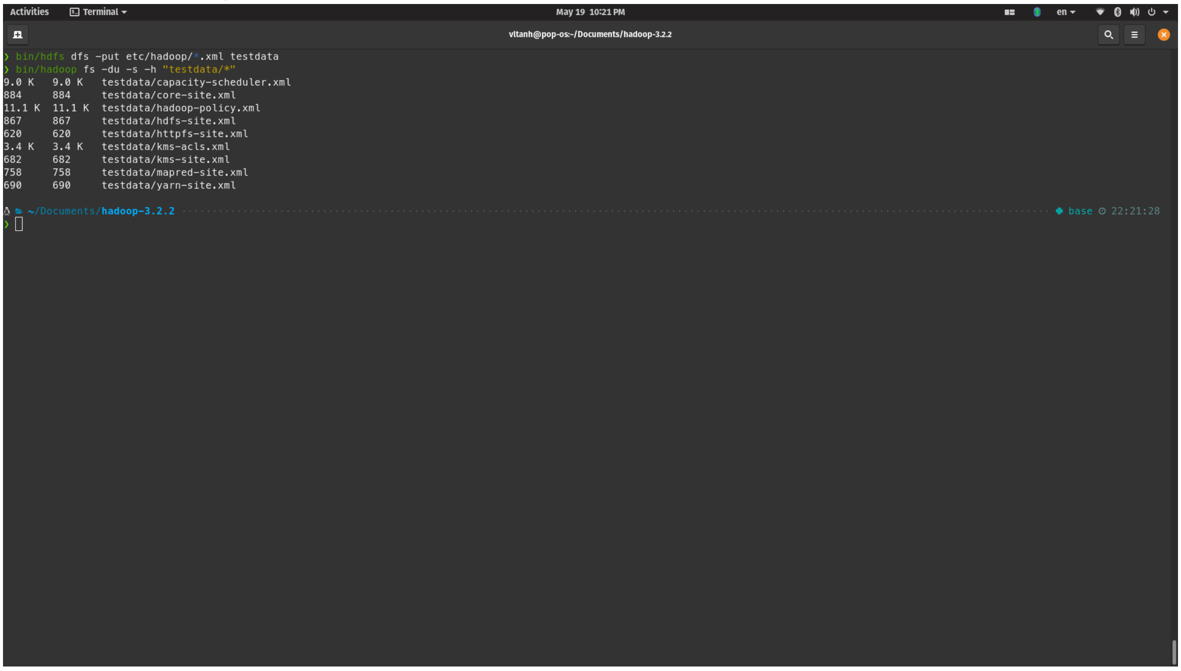


```
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

2021-05-11 12:57:04,752 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
stTrusted = false, remoteHostTrusted = false
2021-05-11 12:57:05,577 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
stTrusted = false, remoteHostTrusted = false
2021-05-11 12:57:05,620 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
```



```
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

8.1 K   8.1 K   testdata/capacity-scheduler.xml
884     884     testdata/core-site.xml
11.1 K  11.1 K  testdata/hadoop-policy.xml
867     867     testdata/hdfs-site.xml
```

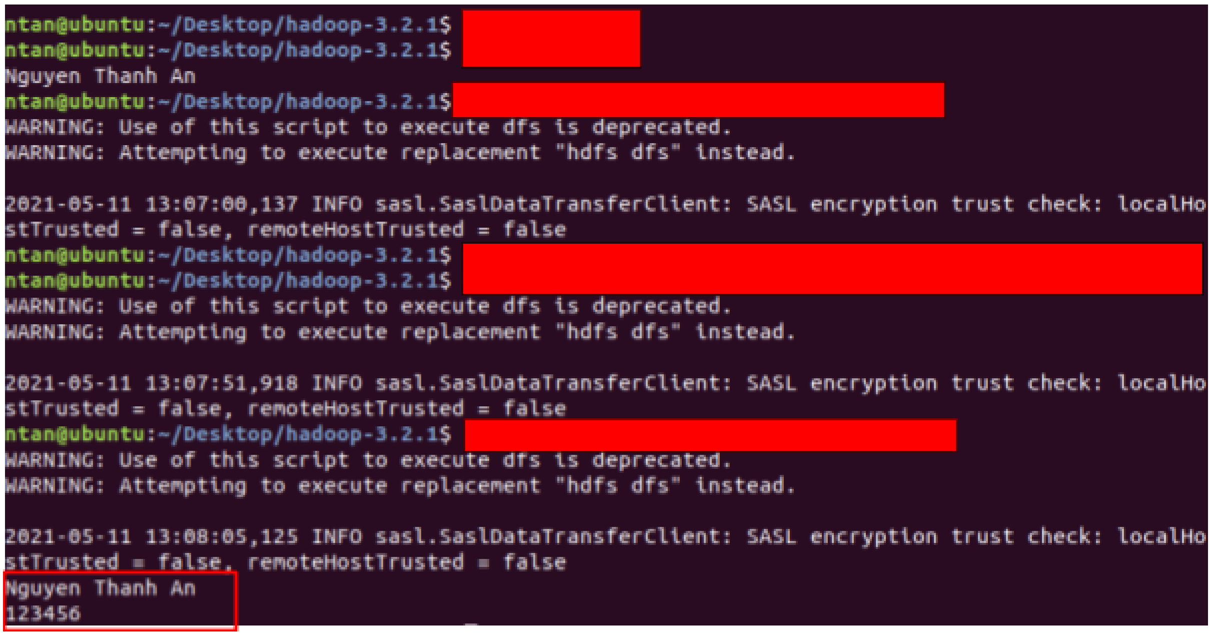
Your screenshot goes here



```
❯ bin/hdfs dfs -put etc/hadoop/*.xml testdata
❯ bin/hadoop fs -du -s -h "testdata/*"
9.0 K   9.0 K   testdata/capacity-scheduler.xml
884     884     testdata/core-site.xml
11.1 K  11.1 K  testdata/hadoop-policy.xml
867     867     testdata/hdfs-site.xml
620     620     testdata/httpfs-site.xml
3.4 K   3.4 K   testdata/kms-acls.xml
682     682     testdata/kms-site.xml
758     758     testdata/mapred-site.xml
690     690     testdata/yarn-site.xml

~/Documents/hadoop-3.2.2 ·························· base 22:21:28
❯
```

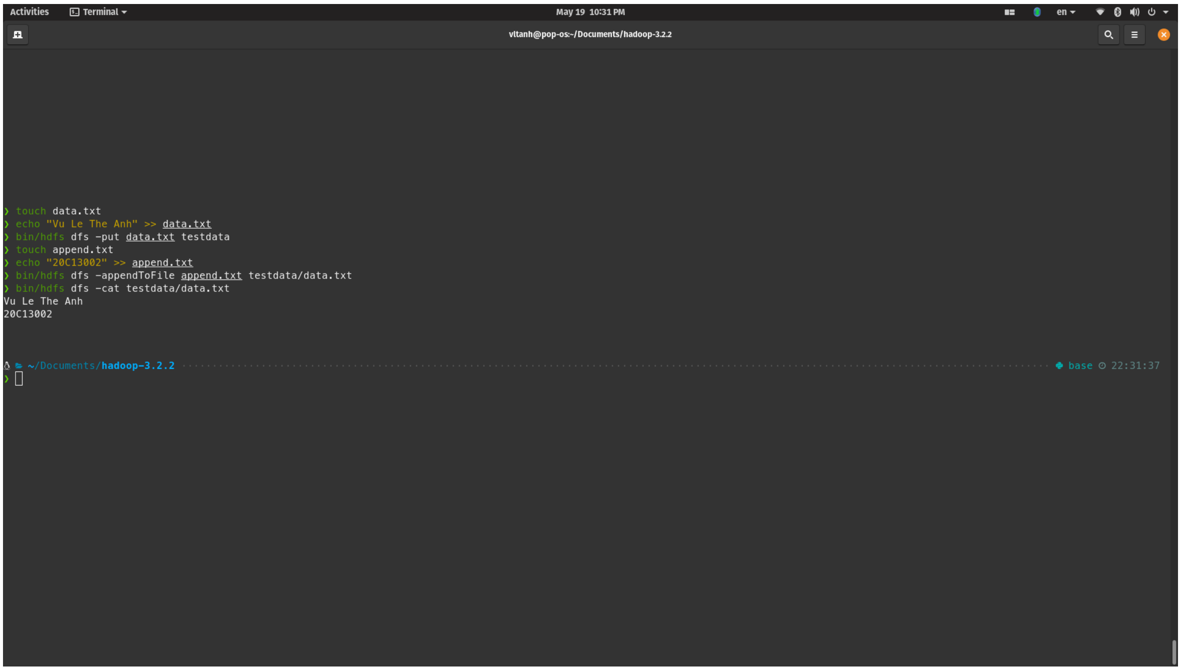
Question 5:

Create a text file, named **data.txt**, in your local machine.
Write your fullname in **data.txt**.
Upload **data.txt** to **testdata/**
Create a text file, named **append.txt**, in your local machine.
Write your student number in **append.txt**.
Append the content of **append.txt** to **data.txt**
Display the content of **data.txt**
For example,



```
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
ntan@ubuntu:~/Desktop/hadoop-3.2.1$
Nguyen Thanh An
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

2021-05-11 13:07:00,137 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
stTrusted = false, remoteHostTrusted = false
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
ntan@ubuntu:~/Desktop/hadoop-3.2.1$
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

2021-05-11 13:07:51,918 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
stTrusted = false, remoteHostTrusted = false
ntan@ubuntu:~/Desktop/hadoop-3.2.1$ [redacted]
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

2021-05-11 13:08:05,125 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
stTrusted = false, remoteHostTrusted = false
Nguyen Thanh An
123456
```

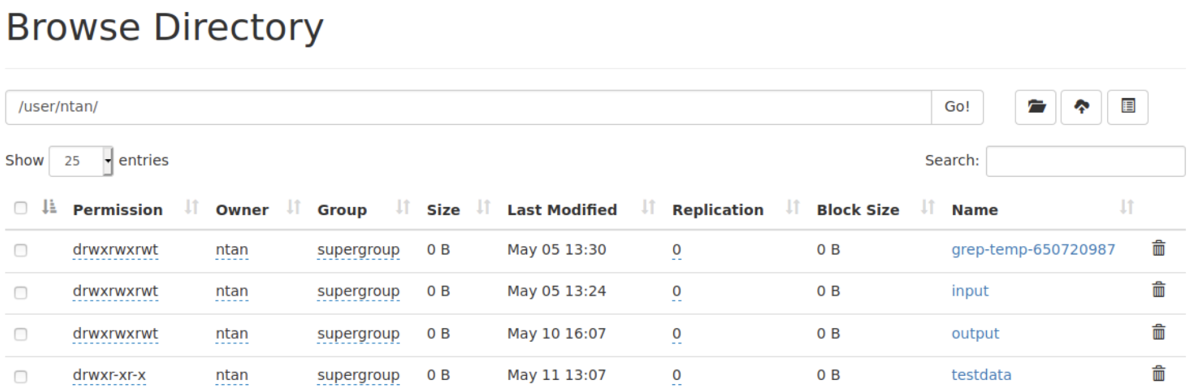
Your screenshot goes here



```
❯ touch data.txt
❯ echo "Vu Le The Anh" >> data.txt
❯ bin/hdfs dfs -put data.txt testdata
❯ touch append.txt
❯ echo "20C13002" >> append.txt
❯ bin/hdfs dfs -appendToFile append.txt testdata/data.txt
❯ bin/hdfs dfs -cat testdata/data.txt
Vu Le The Anh
20C13002


~/Documents/hadoop-3.2.2 ························································ base 22:31:37
❯
```

Question 6:

Go to http://localhost:9870/explorer.html#/ to access the web interface for the file system. For example,



Browse Directory

/user/ntan/ Go!

Show 25 entries Search:

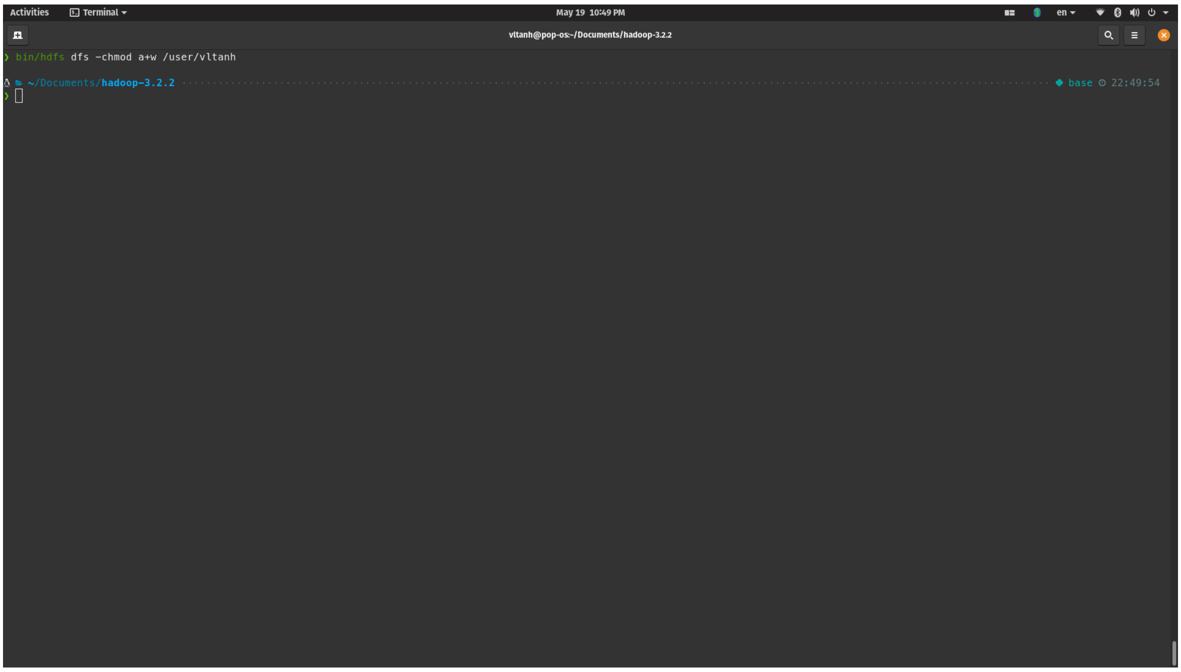
□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
□	drwxrwxrwt	ntan	supergroup	0 B	May 05 13:30	0	0 B	grep-temp-650720987	🗑
□	drwxrwxrwt	ntan	supergroup	0 B	May 05 13:24	0	0 B	input	🗑
□	drwxrwxrwt	ntan	supergroup	0 B	May 10 16:07	0	0 B	output	🗑
□	drwxr-xr-x	ntan	supergroup	0 B	May 11 13:07	0	0 B	testdata	🗑

Create a folder named **testoutput/** in **/user/<username>** and then your operation is denied. Specify the reason, fix the problem, and then conduct the requirement.

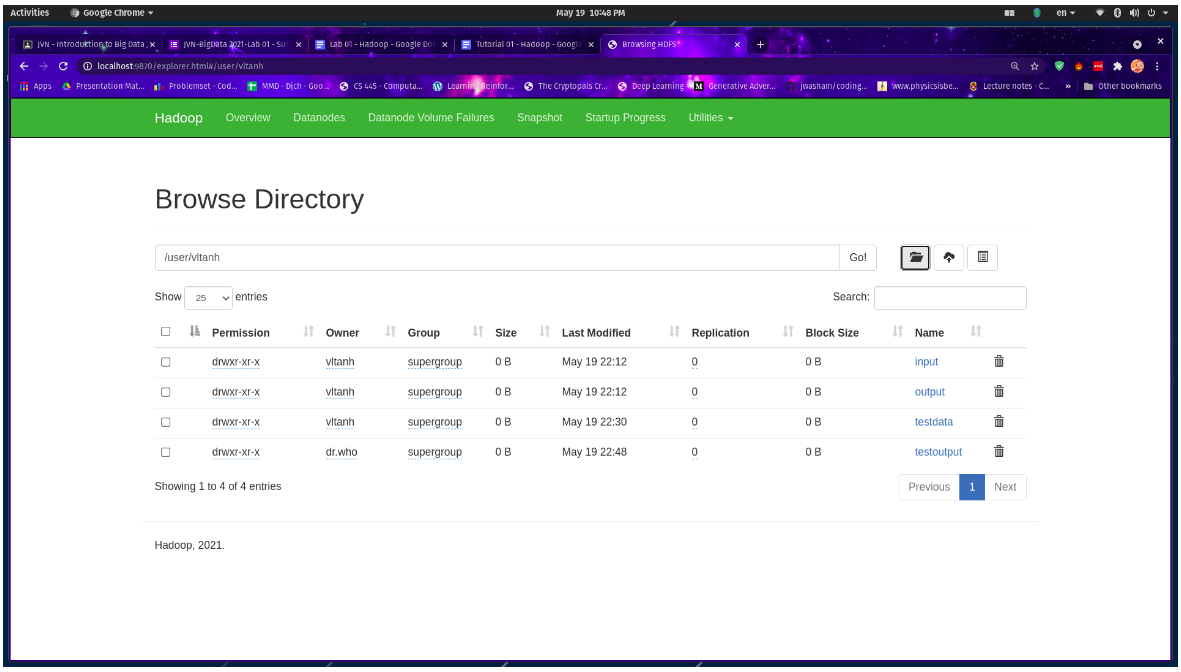
Reason: The error is “Permission denied: user=dr.who, access=WRITE, inode="/user/vltanh":vltanh:supergroup:drwxr-xr-x”. This means that the “/user/vltanh” directory is owned by “vltanh” with 755 (drwxr-xr-x) permissions and therefore only “vltanh” can write to “/user/vltanh” (“dr.who" cannot).

Solution: Although this might not be a recommended operation, give everyone (“dr.who” included) the write permission (to “/user/vltanh”) using “bin/hdfs dfs -chmod a+w /user/vltanh” then proceed.

Your screenshot goes here



```
❯ bin/hdfs dfs -chmod a+w /user/vltanh

 ~/Documents/hadoop-3.2.2 ··································· base  22:49:54
❯ 
```



Browse Directory

/user/vltanh Go!

Show 25 entries

Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	vltanh	supergroup	0 B	May 19 22:12	0	0 B	input
□	drwxr-xr-x	vltanh	supergroup	0 B	May 19 22:12	0	0 B	output
□	drwxr-xr-x	vltanh	supergroup	0 B	May 19 22:30	0	0 B	testdata
□	drwxr-xr-x	dr.who	supergroup	0 B	May 19 22:48	0	0 B	testoutput

Showing 1 to 4 of 4 entries

Previous 1 Next

Hadoop, 2021.

Submission Notice

- Export your answer file as pdf
- Rename the pdf following the format:

  <student number>_HoTen.pdf

  E.g: **123456_NguyenThanhAn.pdf**

  If you have not been assigned a student number yet, then use 123456 instead.
- Careless mistakes in filename, format, question order, etc. are not accepted (0 pts).
- Submission URL: https://forms.gle/zSitTjPTvjUogMBfA
- **Deadline: 17:00, 20/05/2021 (Thu).**