

# Course: MMDS

## Lab 02

## HDFS

Fill answers of the questions below in the given tables.

Your screenshots must **contain commands** for required operations.

### Question 1:

Download and unzip [lab02.zip](#) to obtain 365 text files.

Create the folder `/user/<username>/lab02` in your HDFS

Copy the entire **lab02/** from your local filesystem to `/user/<username>/lab02` in HDFS

Take two screenshots to show your results.

For example,

```
Found 1 items
drwxr-xr-x  - ntan supergroup          0 2021-05-18 15:27 /user/ntan/lab02/lab02

Found 365 items
-rw-r--r--  1 ntan supergroup          1 2021-05-18 15:27 /user/ntan/lab02/lab02/2021_01_01.txt
-rw-r--r--  1 ntan supergroup          2 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_02.txt
-rw-r--r--  1 ntan supergroup          2 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_03.txt
-rw-r--r--  1 ntan supergroup          1 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_04.txt
-rw-r--r--  1 ntan supergroup          2 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_05.txt
-rw-r--r--  1 ntan supergroup          2 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_06.txt
-rw-r--r--  1 ntan supergroup          1 2021-05-18 15:26 /user/ntan/lab02/lab02/2021_01_07.txt
```

Your screenshot goes here

```
phuc@phuc-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/phuc/lab02
Found 1 items
drwxr-xr-x  - phuc supergroup          0 2024-02-02 12:23 /user/phuc/lab02/lab02
phuc@phuc-VirtualBox:~/Desktop/hadoop-3.2.1$
```

```

phuc@phuc-VirtualBox: ~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls -R /user/phuc/lab02
drwxr-xr-x  - phuc supergroup 0 2024-02-02 12:23 /user/phuc/lab02/lab02
-rw-r--r--  1 phuc supergroup 1 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_01.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_02.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_03.txt
-rw-r--r--  1 phuc supergroup 1 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_04.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_05.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_06.txt
-rw-r--r--  1 phuc supergroup 1 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_07.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_08.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_09.txt
-rw-r--r--  1 phuc supergroup 1 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_10.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_11.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_12.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_13.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_14.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_15.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_16.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_17.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_18.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_19.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_20.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_21.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_22.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_23.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_24.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_25.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_26.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_27.txt
-rw-r--r--  1 phuc supergroup 2 2024-02-02 12:22 /user/phuc/lab02/lab02/2021_01_28.txt

```

## Question 2:

Create 12 folders, corresponding to 12 months in a year, in `/user/<username>/lab02` in HDFS. Notice: folder names must in the format like `_01`, `_02`, ..., `_12`

Take a screenshot to show your result.

Hint:

- loop in bash scripts
- `$(printf ...)` for formatting strings in bash scripts.

For example,

```

Found 13 items
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_01
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_02
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_03
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_04
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_05
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_06
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_07
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_08
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_09
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_10
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_11
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:30 lab02/_12
drwxr-xr-x - ntan supergroup 0 2021-05-18 15:27 lab02/lab02

```

Your screenshot goes here

```

^C^Cphuc@phuc-VirtualBox:~/Desktop/hadoop-3.2.1$ for i in {1..12}
> do
> bin/hdfs dfs -mkdir /user/phuc/lab02/"$(printf "%02d" $i)"
> done
phuc@phuc-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/phuc/lab02
Found 13 items
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_01
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_02
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_03
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_04
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_05
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_06
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_07
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_08
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_09
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_10
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_11
drwxr-xr-x - phuc supergroup 0 2024-02-02 17:00 /user/phuc/lab02/_12
drwxr-xr-x - phuc supergroup 0 2024-02-02 16:57 /user/phuc/lab02/_bash2d

```

--

### Question 3:

Text files in **lab02/** have a filename format like **YYYY\_MM\_DD.txt**

Move each text file in **lab02/** to the folder which is corresponding to the month in its filename. E.g. **20201\_01\_01.txt** to **\_01/**

Take a screenshot to show the content in **lab02/\_01/**

Take a screenshot to show the sizes (in bytes) of the 12 folders.

Hint: loop in bash scripts

For example,

Loop `i` in `{1..12}` :

Group file by `i` month ex: `2021_01_01 == _01`

Move those file to folder `_0i`

```
Found 31 items
-rw-r--r--  1 ntan supergroup      1 2021-05-18 15:27 lab02/_01/2021_01_01.txt
-rw-r--r--  1 ntan supergroup      2 2021-05-18 15:26 lab02/_01/2021_01_02.txt
-rw-r--r--  1 ntan supergroup      2 2021-05-18 15:26 lab02/_01/2021_01_03.txt
-rw-r--r--  1 ntan supergroup      1 2021-05-18 15:26 lab02/_01/2021_01_04.txt
-rw-r--r--  1 ntan supergroup      2 2021-05-18 15:26 lab02/_01/2021_01_05.txt
```

```
58 58 lab02/_01
52 52 lab02/_02
58 58 lab02/_03
56 56 lab02/_04
58 58 lab02/_05
56 56 lab02/_06
58 58 lab02/_07
58 58 lab02/_08
56 56 lab02/_09
58 58 lab02/_10
56 56 lab02/_11
58 58 lab02/_12
```

*Your screenshot goes here*



```
phuc@phuc-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -du -s -h /user/phuc/lab02/*
58 58 /user/phuc/lab02/_01
52 52 /user/phuc/lab02/_02
58 58 /user/phuc/lab02/_03
56 56 /user/phuc/lab02/_04
58 58 /user/phuc/lab02/_05
56 56 /user/phuc/lab02/_06
58 58 /user/phuc/lab02/_07
58 58 /user/phuc/lab02/_08
56 56 /user/phuc/lab02/_09
1 1 /user/phuc/lab02/_1
58 58 /user/phuc/lab02/_10
56 56 /user/phuc/lab02/_11
58 58 /user/phuc/lab02/_12
0 0 /user/phuc/lab02/_bash2d
0 0 /user/phuc/lab02/lab02
```

#### Question 4:

Using **cat** to display the content of all files (in month order)

- day 10th
- day 01st
- day 07th
- day 04th

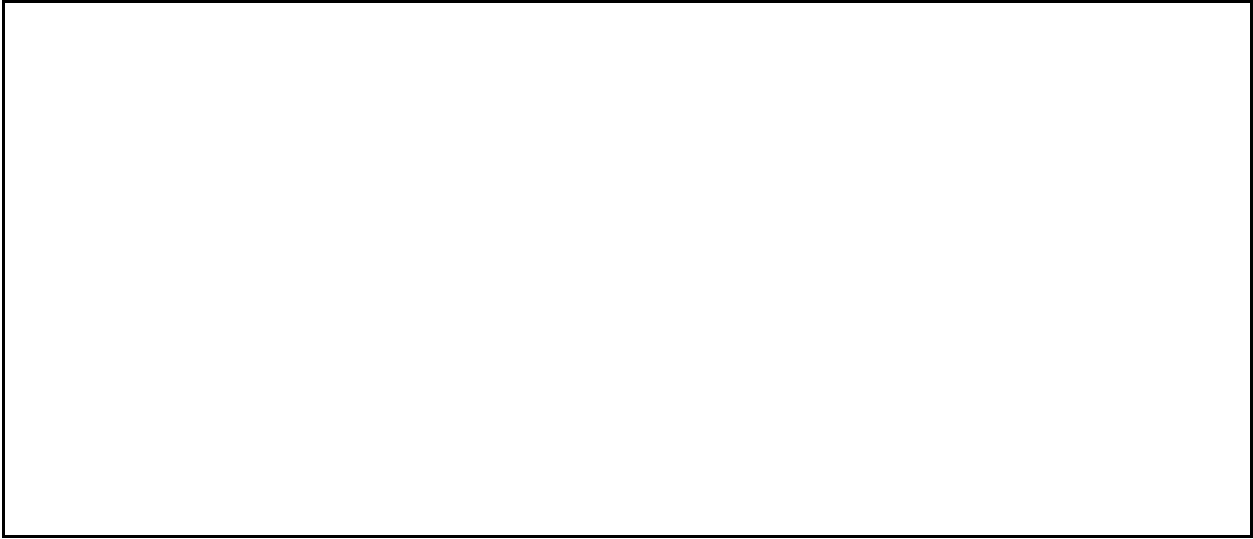
Take 4 screenshots of the corresponding results above.

For example,

```
2021-05-18 15:57:20,510 INFO s
se, remoteHostTrusted = false
e
c
2
1
2
1
a
9
3
7
f
e
```

*Your screenshot goes here*

[illegible]





## Question 5:

Concatenate contents in the 4 screenshots of Question 4 to form the completed message. Finally, write it in the given table.

*Your answer goes here*

\_<3\_i\_love\_you\_for\_a\_thousand\_years\_<3\_.

## Submission Notice

- Export your answer file as pdf
- Rename the pdf following the format:  
lab02\_<student number>\_HoTen.pdf

E.g: **lab02\_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).