# 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 <u>lab02.zip</u> 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,

```
Your screenshot goes here

drwxr-xr-x - vtethoang supergroup 0 2024-01-29 23:00 output
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/viethoang/lab02
Found 2 items
-rw-r--- 2 viethoang supergroup 150569 2024-02-04 14:01 /user/viethoang/lab02/LICENSE.txt
drwxr-xr-x - viethoang supergroup 0 2024-02-04 15:04 /user/viethoang/lab02/lab02
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$
```

```
iethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/viethoang/lab02/lab02
Found 365 items
rw-r--r--
                 2 viethoang supergroup
                                                            1 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_01.txt
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_02.txt
rw-r--r--
                                                                                     /user/viethoang/lab02/lab02/2021_01_03.txt
/user/viethoang/lab02/lab02/2021_01_04.txt
                                                            2 2024-02-04 15:03
ΓW-Γ--Γ--
                 2 viethoang supergroup
                 2 viethoang
                                 supergroup
                                                            1 2024-02-04 15:03
                                                            2 2024-02-04 15:04
                                                                                     /user/viethoang/lab02/lab02/2021_01_05.txt
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_06.txt
/user/viethoang/lab02/lab02/2021_01_07.txt
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                            1 2024-02-04 15:04
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_08.txt
/user/viethoang/lab02/lab02/2021_01_09.txt
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                            1 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_10.txt
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_11.txt
/user/viethoang/lab02/lab02/2021_01_12.txt
                                                            2 2024-02-04 15:04
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                 2 viethoang
                                 supergroup
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                                                     /user/viethoang/lab02/lab02/2021_01_13.txt
                                                            2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_14.txt
2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_15.txt
                 2 viethoang supergroup
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_16.txt
/user/viethoang/lab02/lab02/2021_01_17.txt
                                                            2 2024-02-04 15:04
                 2 viethoang supergroup
                 2 viethoang
                                                            2 2024-02-04 15:04
                                 supergroup
                                                            2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_18.txt
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_19.txt
/user/viethoang/lab02/lab02/2021_01_20.txt
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                            2 2024-02-04 15:04
                 2 viethoang supergroup
                                                           2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_21.txt
2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_22.txt
2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_23.txt
                 2 viethoang supergroup
                 2 viethoang supergroup
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                                                     /user/viethoang/lab02/lab02/2021_01_24.txt
/user/viethoang/lab02/lab02/2021_01_25.txt
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                 2 viethoang
                                 supergroup
                 2 viethoang supergroup
                                                            2 2024-02-04 15:04
                                                                                     /user/viethoang/lab02/lab02/2021_01_26.txt
                                                           2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_27.txt
2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_28.txt
                 2 viethoang supergroup
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_01_29.txt
/user/viethoang/lab02/lab02/2021_01_30.txt
                                                            2 2024-02-04 15:04
                 2 viethoang supergroup
                 2 viethoang
                                                            2 2024-02-04 15:03
                                 supergroup
                                                            2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_01_31.txt
                 2 viethoang supergroup
                                                            1 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_01.txt 2 2024-02-04 15:03 /user/viethoang/lab02/lab02/2021_02_02.txt
                 2 viethoang supergroup
                 2 viethoang supergroup
                                                            2 2024-02-04 15:03 /user/viethoang/lab02/lab02/2021_02_03.txt
1 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_04.txt
2 2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_05.txt
                 2 viethoang supergroup
                 2 viethoang supergroup
                 2 viethoang supergroup
                                                                                     /user/viethoang/lab02/lab02/2021_02_06.txt
/user/viethoang/lab02/lab02/2021_02_07.txt
                                                            2 2024-02-04 15:03
                 2 viethoang supergroup
                 2 viethoang
                                 supergroup
                                                            1 2024-02-04 15:04
                   viethoang supergroup
                                                            2 2024-02-04 15:04
                                                                                     /user/viethoang/lab02/lab02/2021_02_08.txt
                                                              2024-02-04 15:03 /user/viethoang/lab02/lab02/2021_02_09.txt
2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_10.txt
                   viethoang supergroup
                 2 viethoang supergroup
                                                              2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_11.txt
2024-02-04 15:04 /user/viethoang/lab02/lab02/2021_02_12.txt
                 2 viethoang supergroup
```

#### 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
                                      0 2021-05-18 15:30 lab02/_05
drwxr-xr-x
            - ntan supergroup
drwxr-xr-x
                                     0 2021-05-18 15:30 lab02/ 06
            - 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
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
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/vi
ethoang/lab02
Found 13 items
drwxr-xr-x

    viethoang supergroup

                                 0 2024-02-04 15:19 /user/viethoang/
lab02/_01
drwxr-xr-x

    viethoang supergroup

                                0 2024-02-04 15:19 /user/viethoang/
lab02/_02
drwxr-xr-x
           lab02/ 03
           - viethoang supergroup
                                  0 2024-02-04 15:19 /user/viethoang/
drwxr-xr-x
lab02/ 04

    viethoang supergroup

                                     0 2024-02-04 15:19 /user/viethoang/
drwxr-xr-x
lab02/_05

    viethoang supergroup

                                     0 2024-02-04 15:20 /user/viethoang/
drwxr-xr-x
lab02/_06

    viethoang supergroup

                                     0 2024-02-04 15:20 /user/viethoang/
drwxr-xr-x
lab02/ 07
           - viethoang supergroup
                                     0 2024-02-04 15:20 /user/viethoang/
drwxr-xr-x
lab02/_08
           drwxr-xr-x
lab02/_09
           - viethoang supergroup
                                      0 2024-02-04 15:20 /user/viethoang/
drwxr-xr-x
lab02/ 10
                                      0 2024-02-04 15:23 /user/viethoang/
drwxr-xr-x - viethoang supergroup
```

#### 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,

```
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
```

```
lab02/_01
    58
52
    52
        lab02/_02
58
   58
        lab02/ 03
56
   56
        lab02/ 04
58
    58
        lab02/_05
56
    56
        lab02/ 06
        lab02/_07
58
    58
        lab02/_08
58
    58
56
    56
        lab02/_09
58
    58
        lab02/_10
56
    56
        lab02/_11
        lab02/_12
    58
```

Your screenshot goes here

```
riethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /
Found 2 items
drwx----
                       viethoang supergroup
                                                                       0 2024-01-29 23:05 /tmp
drwxr-xr-x
                                                                       0 2024-02-04 14:00 /user
                    - viethoang supergroup
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/viethoang/lab02/_01
Found 31 items
                                                                       1 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_01.txt
 ------
                    2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_02.txt
2 2024-02-04 15:03 /user/viethoang/lab02/_01/2021_01_03.txt
1 2024-02-04 15:03 /user/viethoang/lab02/_01/2021_01_04.txt
 - rw-r--r--
                    2 viethoang supergroup
 rw-r--r--
                    2 viethoang supergroup
 rw-r--r--
                    2 viethoang supergroup
                                                                     1 2024-02-04 15:03 /user/viethoang/lab02/_01/2021_01_04.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_05.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_06.txt
1 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_07.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_08.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_09.txt
1 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_10.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_11.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_12.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_13.txt
 rw-r--r--
                    2 viethoang supergroup
                    2 viethoang supergroup
                    2 viethoang supergroup
                    2 viethoang supergroup
 rw-r--r--
                    2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_14.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_15.txt
                    2 viethoang supergroup
                    2 viethoang supergroup
                                                                       2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_16.txt
                    2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_17.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_18.txt
                    2 viethoang supergroup
                    2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_19.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_20.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_21.txt
                    2 viethoang supergroup
                    2 viethoang supergroup
                    2 viethoang supergroup
 2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_22.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_23.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_24.txt
                    2 viethoang supergroup
                    2 viethoang supergroup
 2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_25.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_26.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_27.txt
                    2 viethoang supergroup
                    2 viethoang supergroup
                    2 viethoang supergroup
 rw-r--r--
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_28.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_28.txt
2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_30.txt
                    2 viethoang supergroup
 rw-r--r--
 ГW-Г--Г--
                    2 viethoang supergroup
 ------
                    2 viethoang supergroup
 - rw-r--r--
                    2 viethoang supergroup
                                                                      2 2024-02-04 15:04 /user/viethoang/lab02/_01/2021_01_31.txt
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -du /user/viethoang/lab02
58
      116
                /user/viethoang/lab02/_01
               /user/viethoang/lab02/_02
/user/viethoang/lab02/_03
/user/viethoang/lab02/_04
52
       104
58
       116
56
       112
58
       116
                /user/viethoang/lab02/_05
56
       112
               /user/viethoang/lab02/_06
58
                /user/viethoang/lab02/_07
       116
58
       116
                /user/viethoang/lab02/_08
                /user/viethoang/lab02/_09
/user/viethoang/lab02/_10
56
       112
58
       116
                /user/viethoang/lab02/_11
56
       112
                /user/viethoang/lab02/ 12
58
       116
       0
                /user/viethoang/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
2
а
```

### Your screenshot goes here

```
Viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.15 bin/hdfs dfs -cat /user/viethoang/lab02/_01/2021_01_10.txt
2024-02-05 00:00:036,292 INFO sasl.SaslDataTransferCitent: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false of the control of the
```

```
aviethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_05/2021_05_10.txt
2024-02-05 00:04:39,920 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_05/2021_05_01.txt
2024-02-05 00:04:45,426 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_05/2021_05_07.txt
2024-02-05 00:04:52,502 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_05/2021_05_04.txt
2024-02-05 00:04:59,862 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_06/2021_06_10.txt
2024-02-05 00:05:26,623 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_06/2021_06_01.txt
2024-02-05 00:05:23,3196 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
oviethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_06/2021_06_07.txt
2024-02-05 00:05:49,739 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
oviethoang@viethoang-VirtualBox:-/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_06/2021_06_04.txt
2024-02-05 00:05:49,739 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoan
_viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_08/2021_08_10.txt
2024-02-05 00:06:59,489 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false of the control of the 
3viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_11/2021_11_10.txt
2024-02-05 00:08:40,333 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
vviethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_11/2021_11_01.txt
2024-02-05 00:08:45,145 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_11/2021_11_07.txt
2024-02-05:00:08:49,433 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
dviethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_11/2021_11_04.txt
2024-02-05:00:08:54,216 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_12/2021_12_10.txt
2024-02-05:00:09:03,988 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_12/2021_12_01.txt
2024-02-05:00:09:03.3NFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_12/2021_12_07.txt
2024-02-05:00:09:13,896 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -cat /user/viethoang/lab02/_12/2021_12_07.txt
2024-02-05:00:09:13,896 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
viethoang@viethoang-Vir
```

### 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 A day without laughter is a day wasted.

### **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).