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
```

```
alfiee@alfiee-VirtualBox: ~/Desktop/hadoop-3.2.1

alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - alfiee supergroup          0 2024-02-02 22:49 /user
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/alfiee/l
ab02/lab02
Found 365 items
-rw-r--r--   1 alfiee supergroup          1 2024-02-02 23:24 /user/alfiee/lab02/
lab02/2021_01_01.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:23 /user/alfiee/lab02/
 ~b02/2021_01_02.txt
ox v-r--r--   1 alfiee supergroup          2 2024-02-02 23:23 /user/alfiee/lab02/
lab02/2021_01_03.txt
-rw-r--r--   1 alfiee supergroup          1 2024-02-02 23:23 /user/alfiee/lab02/
lab02/2021_01_04.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:24 /user/alfiee/lab02/
lab02/2021_01_05.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:23 /user/alfiee/lab02/
lab02/2021_01_06.txt
-rw-r--r--   1 alfiee supergroup          1 2024-02-02 23:24 /user/alfiee/lab02/
lab02/2021_01_07.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:24 /user/alfiee/lab02/
lab02/2021_01_08.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:24 /user/alfiee/lab02/
lab02/2021_01_09.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
```

```
/home/alfiee/Desktop/hadoop-3.2.1
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/alfiee/l
ab02/
Found 14 items
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:53 /user/alfiee/lab02/
 __01
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:53 /user/alfiee/lab02/
 __02
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:53 /user/alfiee/lab02/
 __03
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:53 /user/alfiee/lab02/
 __04
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
 __05
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
tu Software
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
 __07
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
 __08
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
 __09
drwxr-xr-x   - alfiee supergroup          0 2024-02-03 00:54 /user/alfiee/lab02/
 __10
```

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/

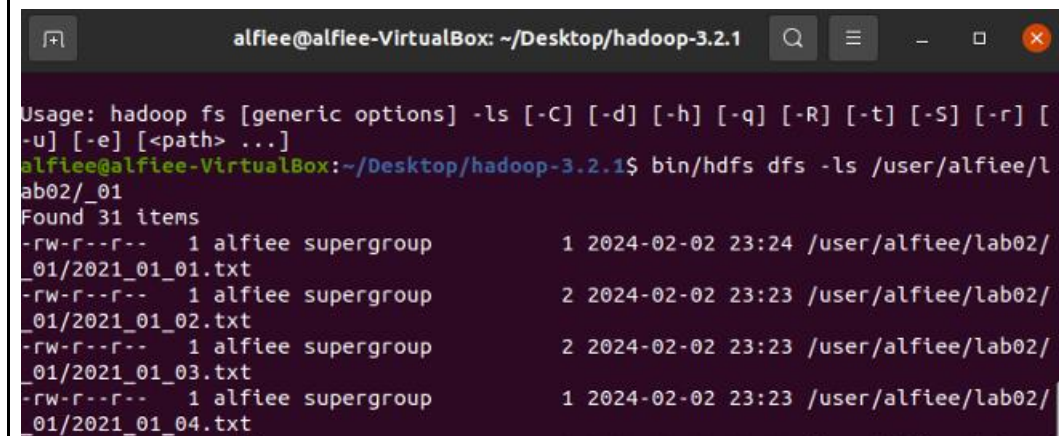Take a screenshot to show the content in **lab02/_01/**

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
```

```
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
```

```
alfiee@alfiee-VirtualBox: ~/Desktop/hadoop-3.2.1

Usage: hadoop fs [generic options] -ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [
-u] [-e] [<path> ...]
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -ls /user/alfiee/l
ab02/_01
Found 31 items
-rw-r--r--   1 alfiee supergroup          1 2024-02-02 23:24 /user/alfiee/lab02/
_01/2021_01_01.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:23 /user/alfiee/lab02/
_01/2021_01_02.txt
-rw-r--r--   1 alfiee supergroup          2 2024-02-02 23:23 /user/alfiee/lab02/
_01/2021_01_03.txt
-rw-r--r--   1 alfiee supergroup          1 2024-02-02 23:23 /user/alfiee/lab02/
_01/2021_01_04.txt
```

```
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -du /user/alfiee/l
ab02
58   58   /user/alfiee/lab02/_01
52   52   /user/alfiee/lab02/_02
58   58   /user/alfiee/lab02/_03
56   56   /user/alfiee/lab02/_04
58   58   /user/alfiee/lab02/_05
56   56   /user/alfiee/lab02/_06
58   58   /user/alfiee/lab02/_07
58   58   /user/alfiee/lab02/_08
56   56   /user/alfiee/lab02/_09
58   58   /user/alfiee/lab02/_10
56   56   /user/alfiee/lab02/_11
58   58   /user/alfiee/lab02/_12
0    0    /user/alfiee/lab02/create_month.sh
0    0    /user/alfiee/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
```

```
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ sudo vi lab02_question4.sh
alfiee@alfiee-VirtualBox:~/Desktop/hadoop-3.2.1$ ./lab02_question4.sh
Month: 01 Day: 10\n

_
Month: 01 Day: 01\n

_
Month: 01 Day: 07\n
a
Month: 01 Day: 04\n

_
Month: 02 Day: 10\n
<
Month: 02 Day: 01\n

_
Month: 02 Day: 07\n

_
Month: 02 Day: 04\n
y
Month: 03 Day: 10\n
3
Month: 03 Day: 01\n
y
Month: 03 Day: 07\n
```

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

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).