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:

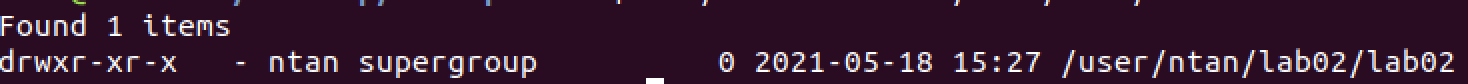
Download and unzip [lab02.zip](https://drive.google.com/file/d/1qfWHmFXZrQmwLfcp_FcgNRX2IDSqOYQ8/view?usp=sharing) to obtain 365 text files.

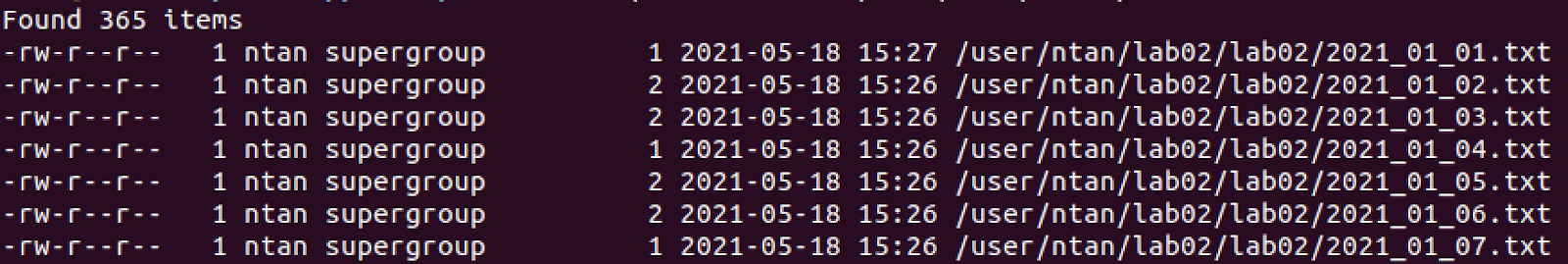
Create the folder **/user/<username>/lab02** in your HDFS

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* |
| --- |

## Question 2:

Create 12 folders, corresponding to 12 months in a year, in **/user/<username>/lab02** in HDFS.

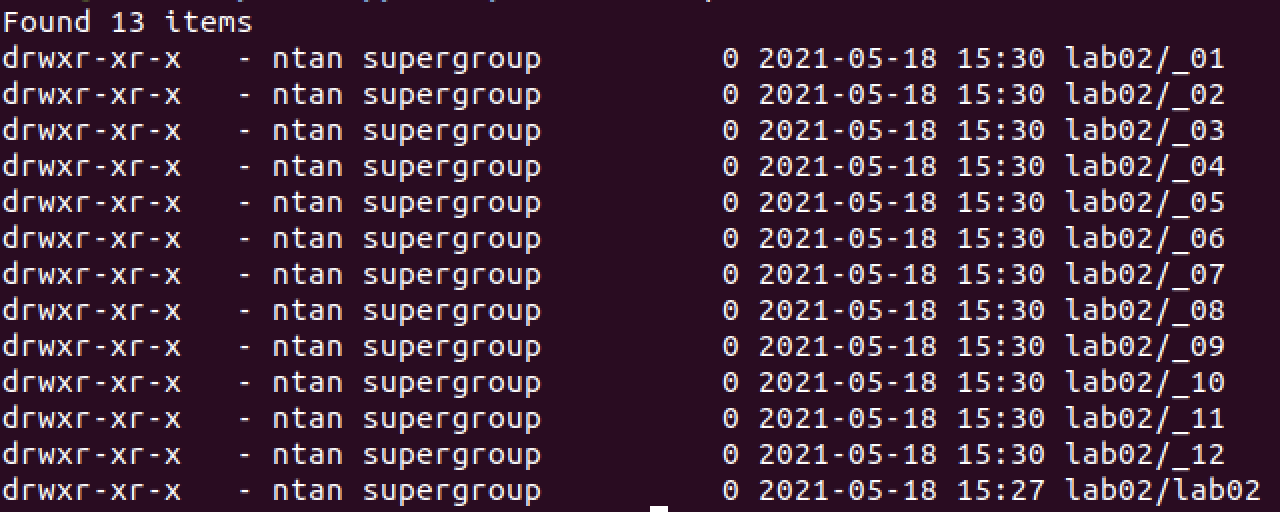
Notice: folder names must in the format like \_01, \_02, …, \_12

Take a screenshot to show your result.

Hint:

* loop in bash scripts
* $(printf …) for formatting strings in bash scripts.

For example,



| *Your screenshot goes here* |
| --- |

## 

## Question 3:

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

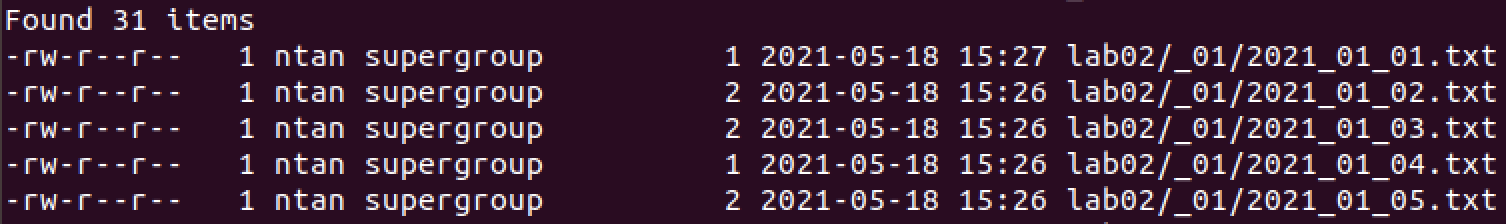
Move each text file in **lab02/** to the folder which is corresponding to the month in its filename. E.g. **20201\_01\_01.txt** to **\_01/**

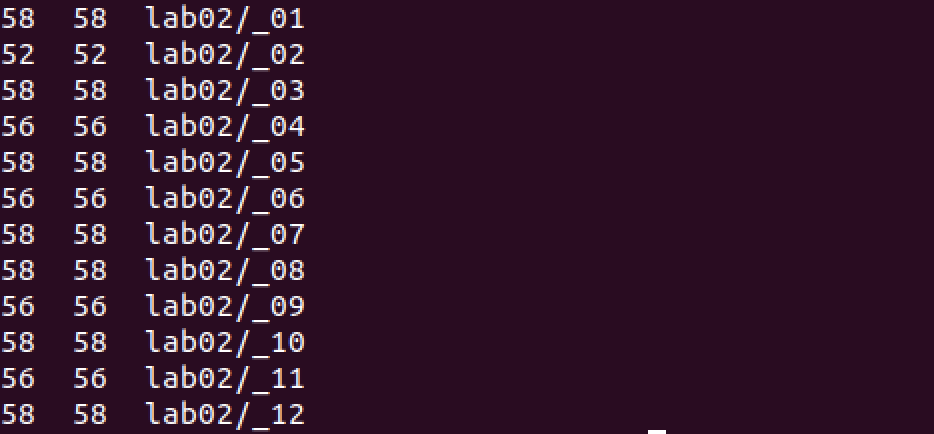
Take a screenshot to show the content in **lab02/\_01/**

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

Hint: loop in bash scripts

For example,





| *Your screenshot goes here* |
| --- |

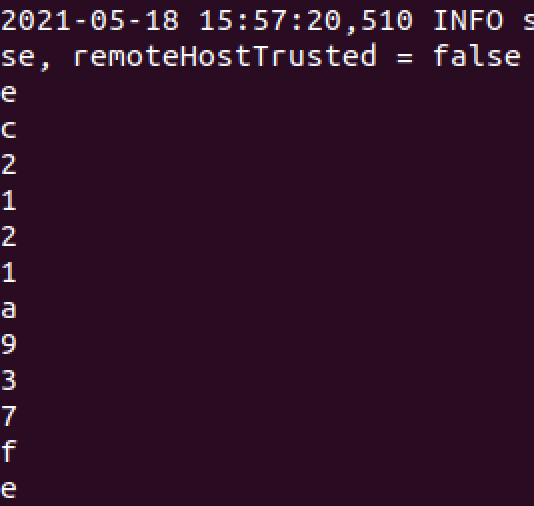
## Question 4:

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,



| *Your screenshot goes here* |
| --- |

## 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).