# MSDM5001: homework1

**NOTE:** Submit your report in the PDF format to the course canvas website. Share your codes/scripts along with some necessary instructions (readme) through GitHub, and provide the URL [1] in your report. We will use git clone $<$ *your repo URL* $>$ to download your codes/scripts. After grading, do not forget to delete the GitHub repository.

At the beginning of your report, write the version of your operating system, the version of bash/python, and any additional libraries or packages. A step-by-step instruction for how to execute your codes/scripts is recommended. If TA has difficulties to run your codes/scripts, you may lose your mark.

You may search the Internet or seek help from others, but you are expected to work independently. Cheating or plagiarism in any form is extremely prohibited and may result in disciplinary action. See: `http://acadreg.ust.hk/generalreg.html`

## 1. Linux operating system and memory hierarchy

Install a Linux operating system distribution on your computer. A virtual machine or a Windows subsystem is recommended. You may also use Amazon or Google cloud computing.

1. Open a terminal, run the command "top", and save a screenshot in your report.
2. Use a few Linux commands to collect the hardware information of your computer to draw the memory hierarchy diagram (see, e.g., Slide 54 in Lecture 1). List the used commands and briefly explain what they are used for.
3. Install the Linux "tree" command if your Linux system does not have it, e.g., sudo apt install tree. Run the commands

[1] e.g., https://github.com/dingpan-HK/MSDM5001.git

```
cd /; tree | head -n 15
```

Paste the output into your report and briefly explain what these commands did.

## 2. Bash script

Write a bash script to create 100 directories/folders, whose names are "DDM1, DDM2, DDM3, ..., DDM100". In each directory, generate a text file, "time_till_now.txt", in which the content is
nanoseconds since 1970-01-01 00:00:00 UTC:
$< XXXXXXXXXXXXXXXXXXX >$
The digits in <> should be calculated when you execute the script. (Hint: you may use the Linux command "date". The same command in macOS/UNIX may not work.)

## 3. Regular expression

Write bash or python scripts to get the desired data from the "blocklist.xml" file. You should simply print the whole lines.

1. Print all the text lines with the "blockID" values that start with the letter "i" or "g", and end with digits, e.g., '<emItem blockID="i334" id="{0F827075-B026-42F3-885D-98981EE7B1AE}">'. (Tip: In the xml file, "blockID" is the attribute name and "i334" is the attribute value.)
2. Print all the text lines where the "ID" values are email addresses. Skip the email addresses that are written by regular expressions containing special characters, such as "\, /, ^ ".

00:53 Monday 28th September, 2020