DE300 Lab 1

1, Create an instance on AWS and SSH your instance

1. Login at https://nu-sso.awsapps.com/start/
2. Navigate by clicking 'mse-tl-dataeng300-EMR'-->''EC2'-->'Launch instance', now you should in a configuration page for the instance that you are going to create. Configure the following:
    - Names and Tag: choose your own name for the instance. You will use this name to identify your instance later, so be specific.
    - Application and OS Images: choose Ubuntu (you are free to try others if you like, I believe AWS linux should be OK for this course)
    - Instance type: stay with the default t2.micro (in real work, you can choose more powerful core depending on your projects)
    - Key pair (IMPORTANT): choose 'create new key pair', and a prompt would pop for you to set up the key file. Enter your preferred file name and leave all other things default. Click 'Create key pair' at the bottom, and then your key file will download to your local machine automatically. Move your downloaded key file to a folder. THe KEY-PATH can look like ~/.ssh/id_rsa/<KEY_FILENAME> (you can choose your own KEY-PATH but make sure it is under ~/,ssh/)
    - Network Settings(IMPORTANT). Click 'EDIT' on the right to change settings. Choose VCP ID: vpc-0f4855643d7cd7d56 (MWAAEnvironment). The series '-0f4855643d7cd7d56' you find in your case can be different, but make sure to choose the one with 'MWAAEnvironment'. Choose a subnet that is public (you see four options when you click the subnet window. Two are labeled as public subsets, which are with IP address 10.192.20…or 10.192.21…. Make sure 'Auto Assign Public IP' is 'enabled'. Choose your own name for 'security group name' (This is mainly for later reference if you create other instances, so that you can choose the same security group set up by selecting this security group name)
    - Configure Storage: Professor Klabjian mentioned that the projects in this course would not be big, so choose 128 or 256 Gib with either gp2 or gp3.
3. Now you are good to launch the instance by clicking the 'Launch instance' on the right of the page. Click the 'EC2' and then click the 'instances' under 'Resources', you can see a list of instances and you should find the one that you just created with your specified name. Wait for several seconds and the instance state will switch to 'running'.
4. SSH your instance.
    - Click your instance id (light blue color), a page with detailed info about your instance will appear. Click the 'connect' will show you four options. Click 'SSH client'.
    - Move to your KEY-PATH folder, and follow the instructions on the 'SSH client' page.At the last step, after running the 'ssh -i ……' command, you will see a message that requires you to create a footprint. Enter 'yes' and you are now at your ubuntu instance on AWS.

2 Further configuration for the security group.

By far, anyone with any IP address can visit your instance, and this is NOT safe. So we would like to configure the security group so that only you with a certain IP address can login to your instance. To do this, click your instance id in the instance list, at the second half of the page, click 'security'. Now you see the clickable security group. Click that and then click the 'Edit inbound rules' and lower right, then enter the IP '165.124.167.0/29' in the blank with the search tool and save rules.

The IP address '165.124.167.0/29' is the one for the Northwestern VPN. In other words, you need to connect to Northwestern VPN before you login to your instance if you do the above setting. Using Northwestern VPN is highly recommended due to safety considerations. If you do not use Northwestern VPN, then you can use 129.105.0.0/16 or 165.124.0.0/16, they are IPs for general Northwestern networks.

3 Download Anaconda distribution

- Download and install the package following the instructions at page https://docs.anaconda.com/free/anaconda/install/linux/
  You can choose installation with the option Linux X86 and run the following command would download the package (as indicated on the webpage):
  curl -O
  https://repo.anaconda.com/archive/Anaconda3-<INSTALLER_VERSION>-Linux-x86_64.sh
  Make sure to replace the <INSTALLER_VERSION> with the version number. The latest is '2024.02-1', and you can see all the versions at https://repo.anaconda.com/archive/. If you want to download the latest version,, your command is
  curl -O https://repo.anaconda.com/archive/Anaconda3-2024.02-1-Linux-x86_64.sh

- Install the package you just downloaded by running (replace the version number below!)
  bash ~/Anaconda3-<INSTALLER_VERSION>-Linux-x86_64.sh

- The default path for the anaconda3 is /home/ubuntu/anaconda3, you need to add this path to the .bashrc file so that your instance knows where to anaconda. For this purpose, you need to ad the following line to your .bashrc file. Since .bashrc is a hidden file, you can see it after you run 'ls -al'. Run 'vim .bashrc' to open the text editor. Move the cursor to the bottom of the file. Hit 'i' to change to insert mode. Copy and paste the following line
  export PATH=$PATH:/home/ubuntu/anaconda3/bin

Then hit ‘esc to switch back to the view mode. Type ‘:wq’ to save your edit and quit the text editor.

- To make .bashrc change in effect,, you may need to logout and login to your instance. (I am not sure) Now in the terminal, if you run ‘which python’ or ‘which python3’, you should see anaconda3 in the output, which indicates you are using the right python distribution.

4 Install the Tmux (if it not installed at your instance)

- If you run ‘tmux’ in your terminal and the output is an error message, then you need to install ‘tmux’. Follow the instruction at https://github.com/tmux/tmux/wiki/Installing: (run ‘apt install tmux’)
- Play the Tmux with the cheat sheet: https://tmuxcheatsheet.com/