

1. Pre-requisite

Enable CPU virtualization in the BIOS. For more instructions for this step, you can search it on google using your motherboard information + enable BIOS CPU virtualization. (For some of you, the virtualization is defaulted as enabled but it is good to go to your BIOS setting and check.)

2. Install WSL

Follow the manual installation guide here: <https://docs.microsoft.com/en-us/windows/wsl/install-manual>.

You should follow the instructions till **Step 6**.

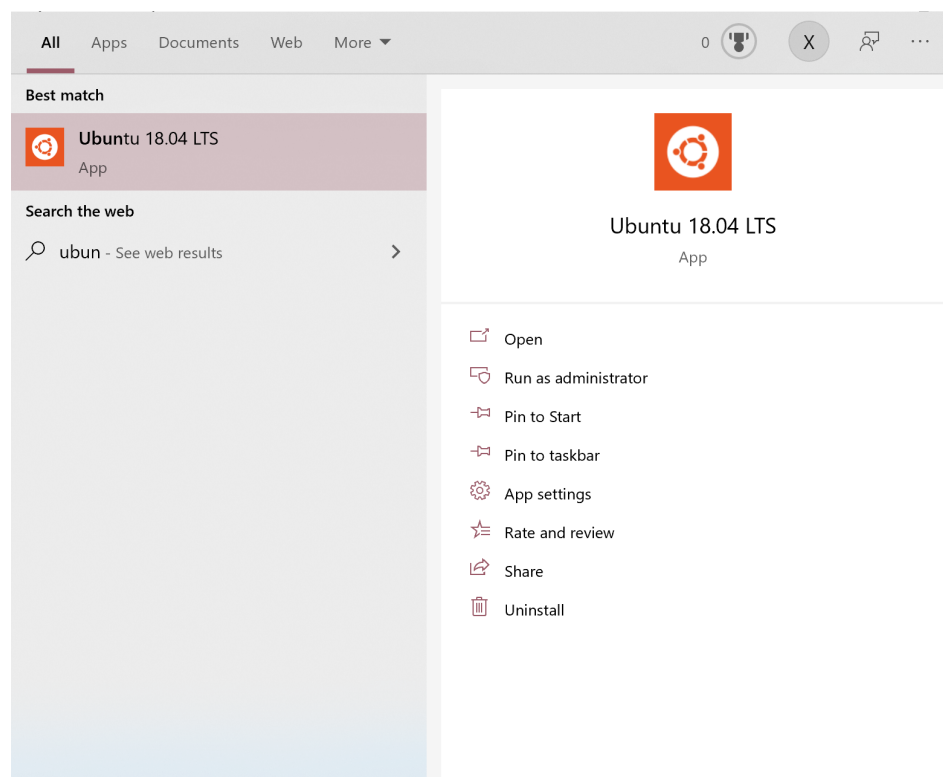
Make sure you follow the instructions closely. If it asked for Administrator access, **right click the Powershell and run it as Administrator**.

3. Download Ubuntu

In "Step 6 - Install your Linux distribution of choice" of the above installation guide. You should choose to install Ubuntu18.04 LTS, which is also located here:

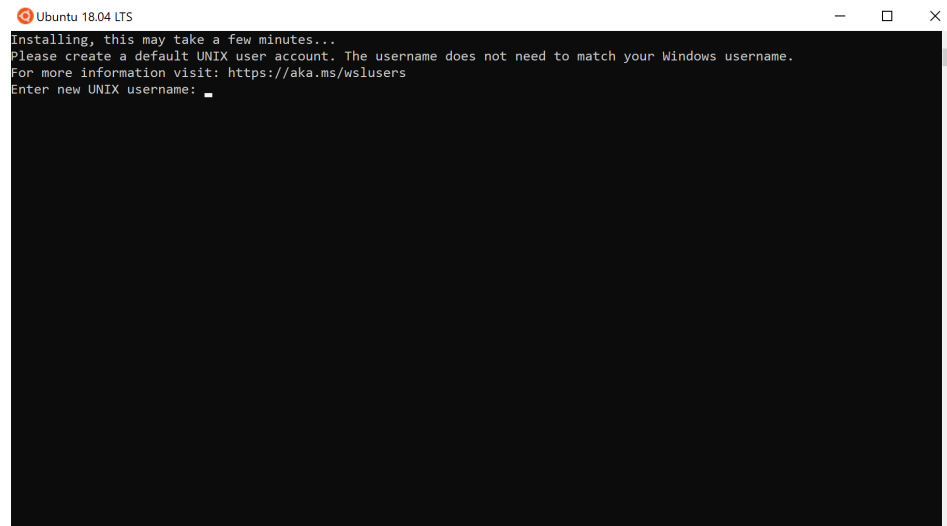
<https://www.microsoft.com/store/apps/9N9TNGVNDL3Q>

4. Open a Ubuntu 18.04 LTS terminal



Notice that you'll be prompted to enter a UNIX username and setup a password. These does not need to

be the same as your windows username or password.



5. Make a download directory

After you've setup your username and password, use the following code to make a download directory in your home directory.

You should copy the code to your ubuntu terminal and execute them line by line.

```
mkdir Download
cd Download
```

6. Install Anaconda in wsl

Now that you are working in the Download directory, use the following code to install Anaconda:

```
sudo apt-get update
sudo apt-get install curl
curl https://repo.anaconda.com/archive/Anaconda3-2021.11-Linux-x86_64.sh \
--output Anaconda3-2021.11-Linux-x86_64.sh
```

Now, you should find the `Anaconda3-2021.11-Linux-x86_64.sh` file in the directory and you can type `ls` in your terminal to make sure.

```
ls
```

```
xilin@DESKTOP-ETS1GF6: ~/Download
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  libcurl4
The following packages will be upgraded:
  curl libcurl4
2 upgraded, 0 newly installed, 0 to remove and 165 not upgraded.
Need to get 378 kB of archives.
After this operation, 14.3 kB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 curl amd64 7.58.0-2ubuntu3.16 [159 kB]
Get:2 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcurl4 amd64 7.58.0-2ubuntu3.16 [220 kB]
Fetched 378 kB in 1s (455 kB/s)
(Reading database ... 29293 files and directories currently installed.)
Preparing to unpack .../curl_7.58.0-2ubuntu3.16_amd64.deb ...
Unpacking curl (7.58.0-2ubuntu3.16) over (7.58.0-2ubuntu3.10) ...
Preparing to unpack .../libcurl4_7.58.0-2ubuntu3.16_amd64.deb ...
Unpacking libcurl4:amd64 (7.58.0-2ubuntu3.16) over (7.58.0-2ubuntu3.10) ...
Setting up libcurl4:amd64 (7.58.0-2ubuntu3.16) ...
Setting up curl (7.58.0-2ubuntu3.16) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
Processing triggers for libc-bin (2.27-3ubuntu1.2) ...
xilin@DESKTOP-ETS1GF6:~/Download$ curl https://repo.anaconda.com/archive/Anaconda3-2021.11-Linux-x86_64.sh \
> --output Anaconda3-2021.11-Linux-x86_64.sh
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
100 580M 100 580M    0     0 52.9M    0  0:00:10  0:00:10  --:--:-- 53.9M
xilin@DESKTOP-ETS1GF6:~/Download$ ls
Anaconda3-2021.11-Linux-x86_64.sh
xilin@DESKTOP-ETS1GF6:~/Download$
```

Check the SHA-256 Hash and make sure the output matches the following (this step verifies the data integrity of the Anaconda installer files).

```
fedf9e340039557f7b5e8a8a86affa9d299f5e9820144bd7b92ae9f7ee08ac60
```

```
sha256sum Anaconda3-2021.11-Linux-x86_64.sh
```

Now that you have the installer, you can use the following code and follow the instructions to install anaconda.

```
bash Anaconda3-2021.11-Linux-x86_64.sh
```

When the installation is done, close the current window and open a new Ubuntu window. Now you should see a *(base)* before your username.

7. Install Java 8

Use the following code to install java 8 to your machine:

```
sudo apt install openjdk-8-jdk
```

8. Download Apache Spark

Now, we'll get the Spark package using the following code:

```
wget https://d1cdn.apache.org/spark/spark-3.2.0/spark-3.2.0-bin-hadoop3.2.tgz
```

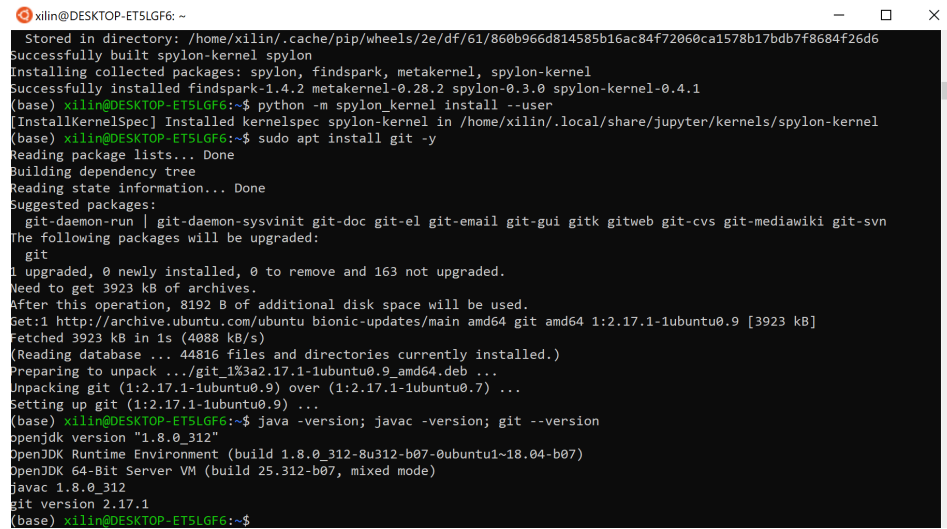
9. Install Scala and Spark

Use the following code to install spylon-kernel and git

```
pip install spylon-kernel
python -m spylon_kernel install --user
sudo apt install git -y
```

Then, you should check the versions and see if they match mine:

```
java -version; javac -version; git --version
```



```
xilin@DESKTOP-ETSJGF6: ~
└─$ python -m pip install spylon
Stored in directory: /home/xilin/.cache/pip/wheels/2e/df/61/860b96d814585b16ac84f72060ca1578b17bdb7f8684f26d6
Successfully built spylon-kernel spylon
Installing collected packages: spylon, findspark, metakernel, spylon-kernel
Successfully installed findspark-1.4.2 metakernel-0.28.2 spylon-0.3.0 spylon-kernel-0.4.1
(base) xilin@DESKTOP-ETSJGF6:~$ python -m spylon_kernel install --user
[InstallKernelSpec] Installed kernelspec spylon-kernel in /home/xilin/.local/share/jupyter/kernels/spylon-kernel
(base) xilin@DESKTOP-ETSJGF6:~$ sudo apt install git -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
Suggested packages:
  git-daemon-run | git-daemon-sysvinit git-doc git-el git-email git-gui gitk gitweb git-cvs git-mediawiki git-svn
The following packages will be upgraded:
  git
1 upgraded, 0 newly installed, 0 to remove and 163 not upgraded.
Need to get 3923 kB of archives.
After this operation, 8192 B of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 git amd64 1:2.17.1-1ubuntu0.9 [3923 kB]
Fetched 3923 kB in 1s (4088 kB/s)
(Reading database ... 44816 files and directories currently installed.)
Preparing to unpack .../git_1%3a2.17.1-1ubuntu0.9_amd64.deb ...
Unpacking git (1:2.17.1-1ubuntu0.9) over (1:2.17.1-1ubuntu0.7) ...
Setting up git (1:2.17.1-1ubuntu0.9) ...
(base) xilin@DESKTOP-ETSJGF6:~$ java -version; javac -version; git --version
openjdk version "1.8.0_312"
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~18.04-b07)
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
javac 1.8.0_312
git version 2.17.1
(base) xilin@DESKTOP-ETSJGF6:~$
```

Now, let's unzip the Spark package we just downloaded (Scala 2.13 comes with this pre-built) in Step 8 and move it to the "/opt/spark" directory

```
tar -xvf spark-3.2.0-bin-hadoop3.2.tgz
sudo mv spark-3.2.0-bin-hadoop3.2 /opt/spark
```

10. Configure the environment

This is a very important last step! **Make sure you replace {username} with your UNIX username!** If you are not clear about your username, type `ls /home/`.

```
echo "export SPARK_HOME=/opt/spark" >> ~/.profile
echo "export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin" >> ~/.profile
echo "export PYPARK_PYTHON=/home/{username}/anaconda3/bin/python" >> ~/.profile
source ~/.profile
```

11. A few final remarks

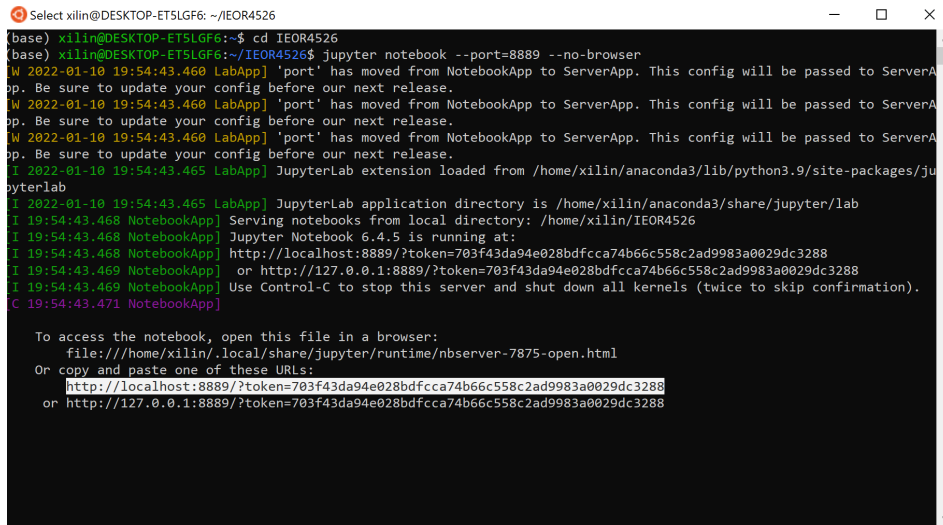
Now we are done installing Spark! You can close the terminal and restart a new one. Before we try out Spark in a Jupyter notebook, I recommend you to make a new directory and name it IEO4526 using the following code:

```
mkdir IEO4526
```

go to the `\\wsl$\Ubuntu-18.04\home\{username}` directory (press the windows key and R key at the same time and paste this directory there, also make sure to replace {username} with your UNIX username) and create a shortcut to the IEO4526 folder on your desktop. You should keep the files from this class (class notebooks and homework assignments) there so that you can easily access them.

Now, in the newly opened terminal, type the following command to open the Jupyter notebook. Copy the address shown in the terminal to a web browser to open the notebook.

```
cd IEOR4526
jupyter notebook --port=8889 --no-browser
```



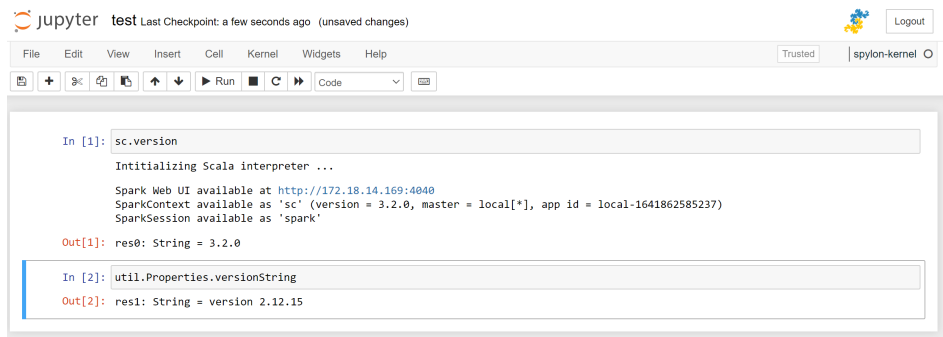
```
Select xilin@DESKTOP-ETS1GF6: ~/IEOR4526
(base) xilin@DESKTOP-ETS1GF6:~$ cd IEOR4526
(base) xilin@DESKTOP-ETS1GF6:~/IEOR4526$ jupyter notebook --port=8889 --no-browser
[W 2022-01-10 19:54:43.460 LabApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2022-01-10 19:54:43.460 LabApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2022-01-10 19:54:43.460 LabApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[I 2022-01-10 19:54:43.465 LabApp] JupyterLab extension loaded from /home/xilin/anaconda3/lib/python3.9/site-packages/jupyterlab
[I 2022-01-10 19:54:43.465 LabApp] JupyterLab application directory is /home/xilin/anaconda3/share/jupyter/lab
[I 19:54:43.468 NotebookApp] Serving notebooks from local directory: /home/xilin/IEOR4526
[I 19:54:43.468 NotebookApp] Jupyter Notebook 6.4.5 is running at:
[I 19:54:43.468 NotebookApp] http://localhost:8889/?token=703f43da94e028bdfcca74b66c558c2ad9983a0029dc3288
[I 19:54:43.469 NotebookApp] or http://127.0.0.1:8889/?token=703f43da94e028bdfcca74b66c558c2ad9983a0029dc3288
[I 19:54:43.469 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
^C 19:54:43.471 NotebookApp]

To access the notebook, open this file in a browser:
file:///home/xilin/.local/share/jupyter/runtime/nbserver-7875-open.html
Or copy and paste one of these URLs:
http://localhost:8889/?token=703f43da94e028bdfcca74b66c558c2ad9983a0029dc3288
or http://127.0.0.1:8889/?token=703f43da94e028bdfcca74b66c558c2ad9983a0029dc3288
```

After you've opened jupyter, create a new spylon-kernel notebook.



In the notebook, you can type `sc.version` to check if the version you have is 3.2.0 and type `util.Properties.versionString` to check if your scala version is 2.12.15



Now, your device is ready! Congrats!