**Instructions for Project Part A**

Every project group has been assigned a server PC with a good configuration that you can access via SSH. **PyLucene is pre-installed on those servers** so that will save you a lot of hassle regarding installing PyLucene.

* You can also run your web servers on port 8080 that you will be able to access via browsers on your own PC.

Please note that PyLucene is recommended over PyElasticsearch as the TAs will be able to provide support in PyLucene, and you would not need to start an elasticsearch instance every time while running PyElasticsearch.

**Instructions on how to access server PC:**

1. Make sure ssh is installed on your PC (ssh is available for Windows, Linux and Mac)  
2. If you have already set up your bolt account, then skip this step. To set up bolt account go to [https://www.cs.ucr.edu/department-intranet](https://www1.cs.ucr.edu/department-intranet) (click login/password reset) to set your bolt account password.  
3. Open a terminal, type "ssh netid@bolt.cs.ucr.edu", and replace "netid" with your net id. This will ask for a password. Input your password here.  
4. Once logged in to bolt, type "cs242\_login". This will log in to your server. Every member of the same team will have access to the same server so collaboration can be done easily.  
5. You can install required python libraries via "*pip3*" command.   
6. Run any python code via "*python3*" command.  
7. You can install Jupyter notebooks on the server, create notebooks from your local PC and transfer files easily.

If you need any assistance regarding accessing the server PC, please feel free to come to our office hours.

You can install Scrapy on the server PC assigned to you (and on your local machine too).

**Scrapy Installation Guidelines:**

To install Scrapy on the server PC, run the command:

*pip3 install scrapy*

[export PATH=$HOME/.local/bin:$PATH]

For your local machine, it is recommended that you use a **virtual environment** such as conda, venv etc. to install Scrapy, rather than pip as Scrapy installation may conflict with already installed python packages in your local machine. You can follow this [link](https://docs.scrapy.org/en/latest/intro/install.html#intro-install) for detailed instructions about installing Scrapy in a virtual environment.   
Once you have Scrapy installed, you need to create a project.

**Create and Run Your Project:**

1) First enter a directory where you’d like to store your code and run the command:

*scrapy startproject Projectname*

Replace "Projectname" with your project name.

2) This will create a Projectname directory with a bunch of files and a directory named ‘spiders’ inside it.

3) You will create a .py file that will contain your Spider/Crawler code inside the spiders directory.

4) A sample code that crawls all quotes from a webpage is provided here. Copy this file into your spiders directory to run.

5) In order to run the code, go to the project’s top-level directory and run:

*scrapy crawl quotes -O ‘filename.json’*

Note that "quotes" is the name of the spider I created in my code. You need to give a unique name to each of your crawlers ( if you have multiple). Scrapy will run the crawler whose name you provide in the command.

You can find more details about Scrapy [here](https://docs.scrapy.org/en/latest/intro/tutorial.html).

**Note:** The example codes shown in the class are [here](https://drive.google.com/drive/folders/1e1ve87PjTI-Wm6VUW2q-t5wqa2acIA-q?usp=drive_link).

**Copying files from server PC to local machine & vice versa**

Use the “**scp**” command: “**scp** *source destination*” (Replace source & destination appropriately)

* For copying files from your local machine to server PC:
  + First copy your file from personal computer to bolt:  
     “scp <filepath>/filename [netid@bolt.cs.ucr.edu](mailto:netid@bolt.cs.ucr.edu):~/
  + After logging in using “cs242\_login”, scp again:   
    “scp netid@bolt.cs.ucr.edu:~/filename ~/”
* For copying files from your server PC to local machine:
  + First login to server PC using “cs242\_login” from bolt, then copy file from server PC to bolt:  
    “scp <filepath>/filename [netid@bolt.cs.ucr.edu](mailto:netid@bolt.cs.ucr.edu):~/”
  + Then copy file from bolt to local machine:  
    “scp netid@bolt.cs.ucr.edu:~/filename <filepath>/filename”

**How to install Jupyter Lab on CS242 server**:

* pip3 install jupyterlab
* pip3 install markupsafe==2.0.1
* Create a default config file using command: jupyter notebook --generate-config
  + If you are getting an error with “jupyter command not found” Try this:
    - “pip install jupyter”
    - On terminal type “vim ~/.bashrc”
    - At the end of the file add “export PATH=$PATH:~/.local/bin”
    - Save the file, then run “source ~/.bashrc”
    - Try the generate config command again
* Create a password using command: jupyter notebook password
  + Your hashed password is stored in ~/.jupyter/jupyter\_notebook\_config.json
* Edit ~/.jupyter/jupyter\_notebook\_config.py, uncomment and edit the following settings:

c.ServerApp.ip = '\*'

c.ServerApp.open\_browser = False

c.ServerApp.password = 'your\_hashed\_password'

c.ServerApp.port = 8888

[**Note: some of the commands may change with time. So check the Internet for the most up-to-date commands**]

**How to run Jupyter Lab in bolt**

* Run Jupyter Lab:  
  jupyter lab --ip 0.0.0.0 --port 8888
* Test in your browser: Visit (UCR VPN must be on if outside campus):  
  http://class-0**XX**.cs.ucr.edu:8888   
  (Replace “**XX**” with the “**project number + 40**” associated with your group, For example, if your group number is 1 then “XX” will be 41, if your group number is 11, “XX” will be 51)
* Login with your password in Jupyter Lab