Recitation01

August 24, 2020

1 Recitation 1: setting up our machines!

- Download and install Python (optional)
- Download and install Python IDE (optional)
- Download and install Jupyter Notebooks
- Github account
- Access Jupyter notebook on the LU HPCC

1.1 Download and install Python (optional)

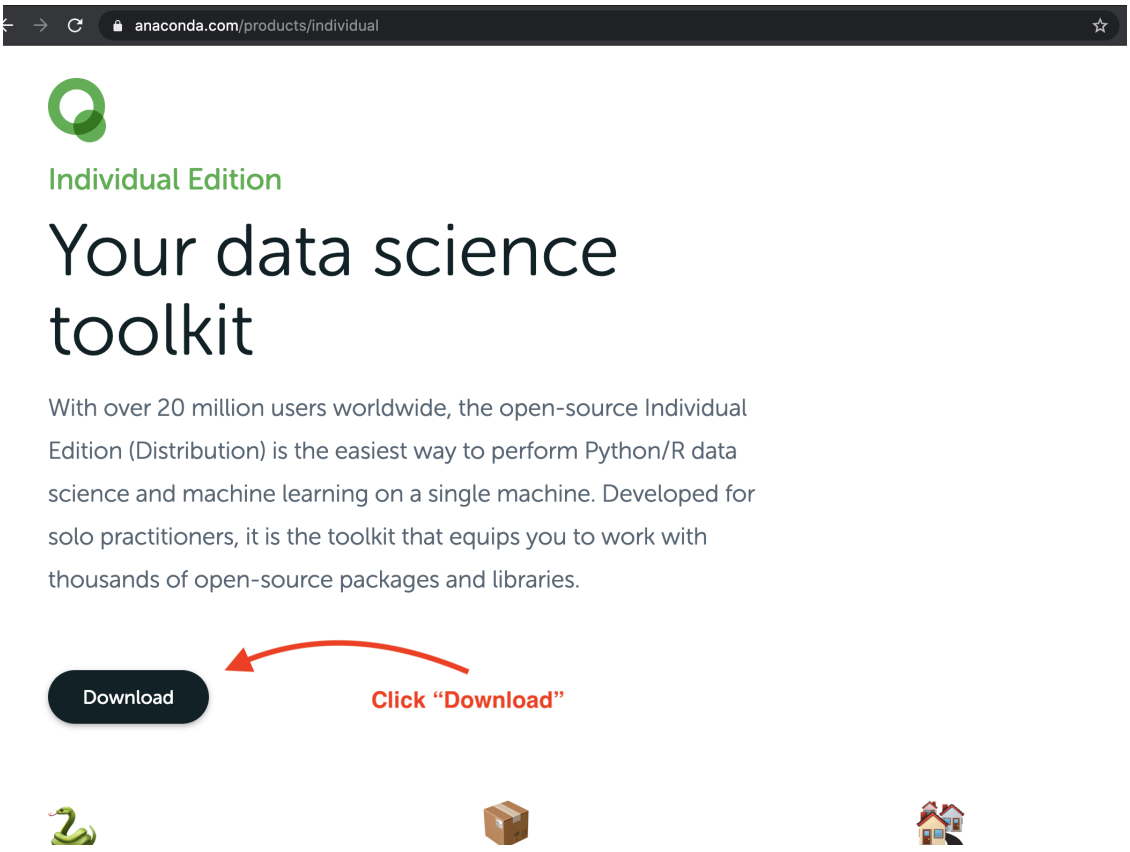
Python 3 is a versatile programming language that is used for just about anything you can think of. We'll use Python for statistical computation in Population Health Data Science (PHDS)-I, and if you're interested, in Population Health Data Science (PHDS)-II and PHDS-III.

If you would like to download Python 3 on your own, local, computer, follow the below steps. This is optional. There is no requirement to install Python 3 on your machine. But often students find it easier to work with programming languages like Python on their own machine.

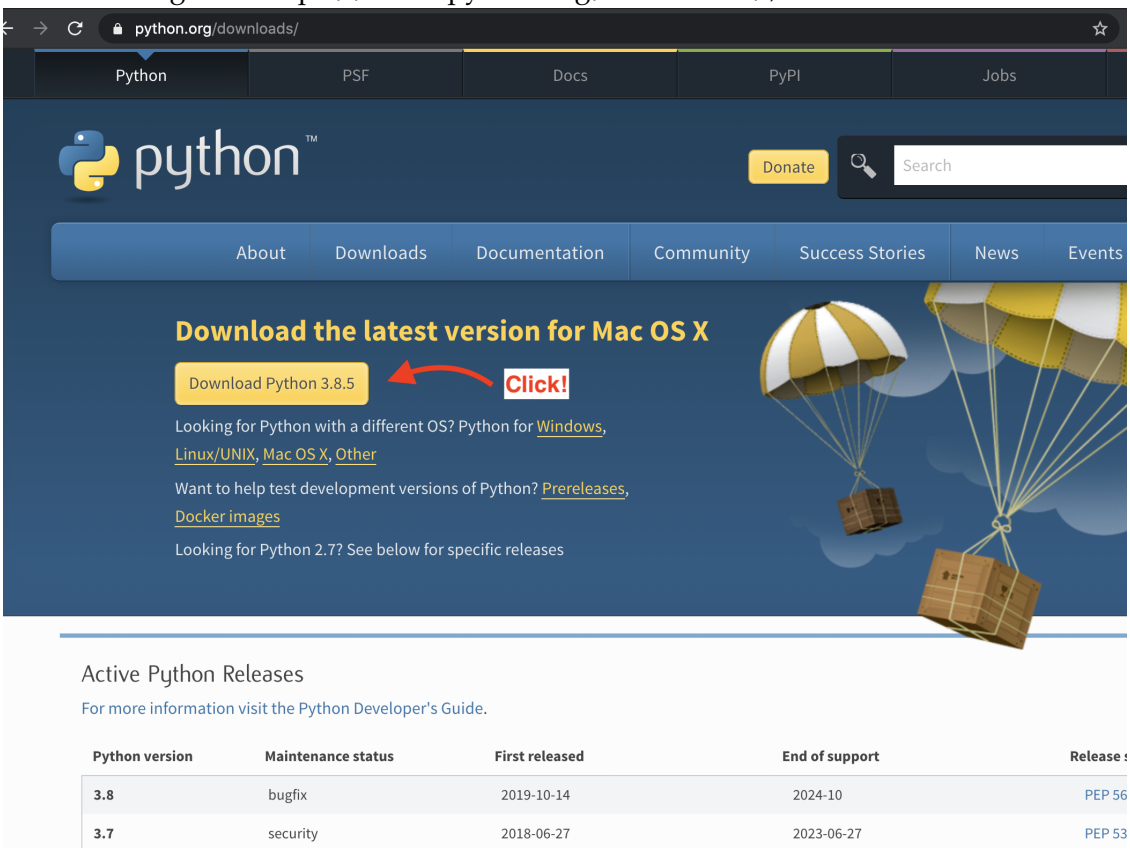
1.1.1 Installation

You have two options for installing Python 3. I recommend installing Anaconda, a suite of the Python language and set of handy packages we'll use throughout the course. The other option is to install just Python 3.

Anaconda To install Anaconda, Python 3 and additional useful packages, use a browser to navigate to https://www.anaconda.com/products/individual, click "Download", and follow the default installation instructions.


anaconda.com/products/individual
Individual Edition
Your data science toolkit
With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.
Download
Click “Download”

Just Python 3 Use a browser to navigate to https://www.python.org/downloads/, and click on Download Python 3.X.X



Click on the downloaded package and follow the default installation prompts from the installer.

1.2 Download and install Python IDE (optional)

An integrated development environment, or IDE for short, is a computer program written to make your work easier. Some IDEs are general purpose software for making a list of programming languages easier to use, others were written for a single language. Python has a few IDEs that have had very good reviews. If you download Python 3 to your local machine, it may be worth checking out the following IDEs:

- Spyder
- PyCharm (the free community edition)
- Pyzo

If you try one of these IDEs and love (or hate) using it, let me know!

1.3 Download and install Jupyter Notebooks (optional)

A Jupyter notebook is software that integrates the Python programming language with a graphical interface and the ability to include text in between lines of code. Jupyter is a useful way to analyze data and share your results with others (for example your professor).

If you installed Anaconda than the software for Jupyter notebooks is pre-installed. You're pretty much done! First open an Anaconda navigator (instructions here

https://docs.anaconda.com/anaconda/user-guide/getting-started/) and then click Jupyter Notebook and install.

If you decided to manually install Python 3 then you'll need to also install Jupyter Notebooks.

1.3.1 For a Mac

Open a "terminal" window and run the following command: curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py

1.3.2 For Windows

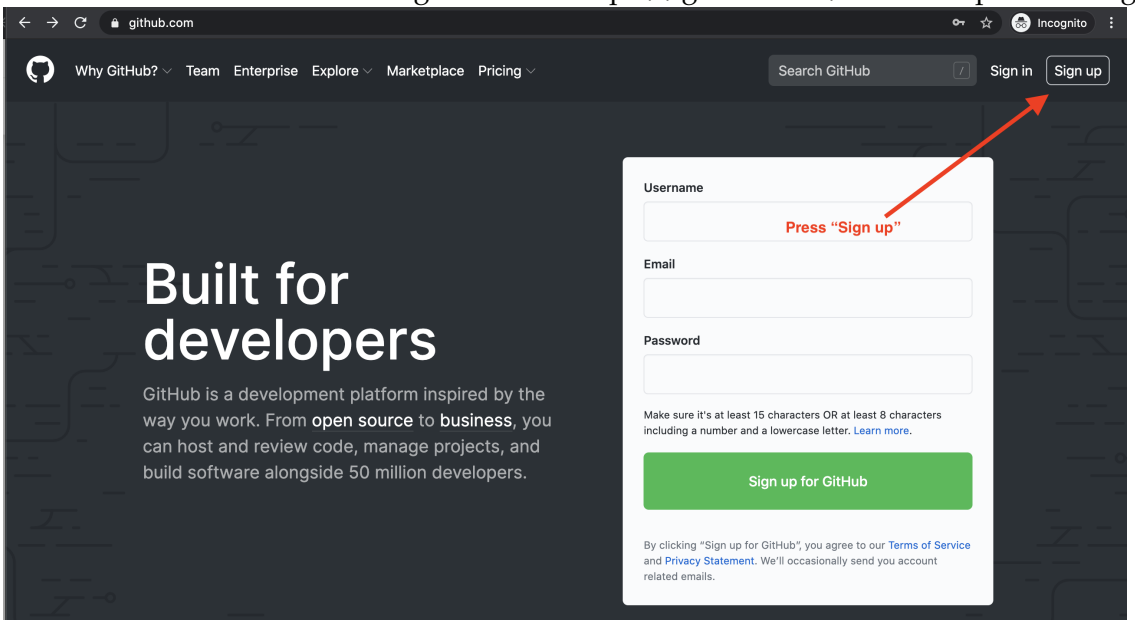
First follow this tutorial to install PIP and then type the following command into a windows terminal: python -m pip install jupyter

1.4 Git and GitHub

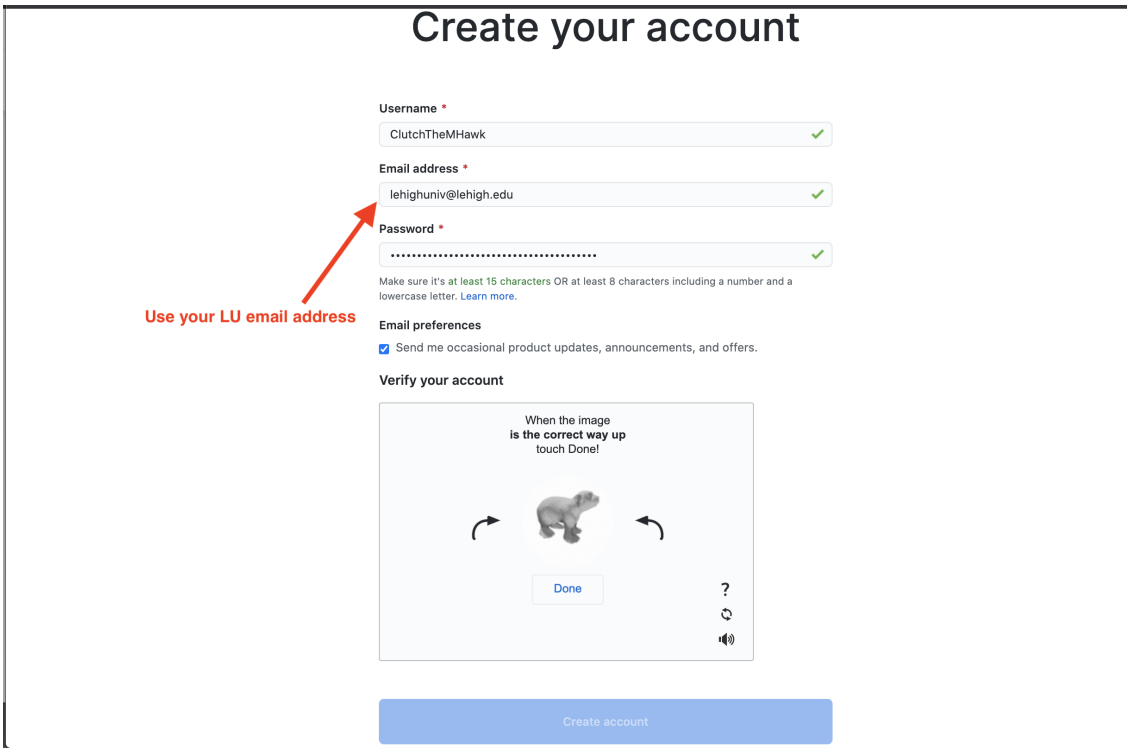
Git is a version control language—a programming language that keeps successive versions of your work to help you rewind to previous code and collaborate with a team. Code, data, visualizations, and any other documents needed for a project are all stored in a **git repository**. One repository should correspond to one project. GitHub is a set of servers allowing users to create accounts, their own Git repositories, and most importantly share their projects with others.

1.4.1 Creating a GitHub Account

Use a browser to navigate to https://github.com/ and press "Sign up".



Clicking "Sign up" takes you to a form where you'll need to enter * A username * An email address (use your LU-issued email address) * A password

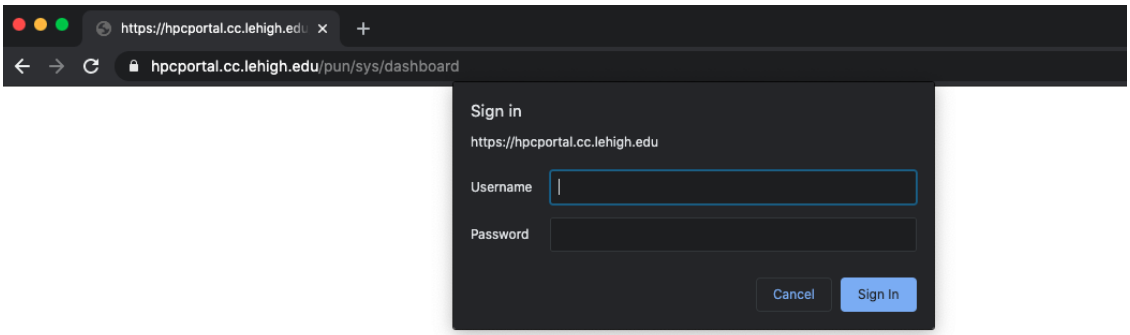


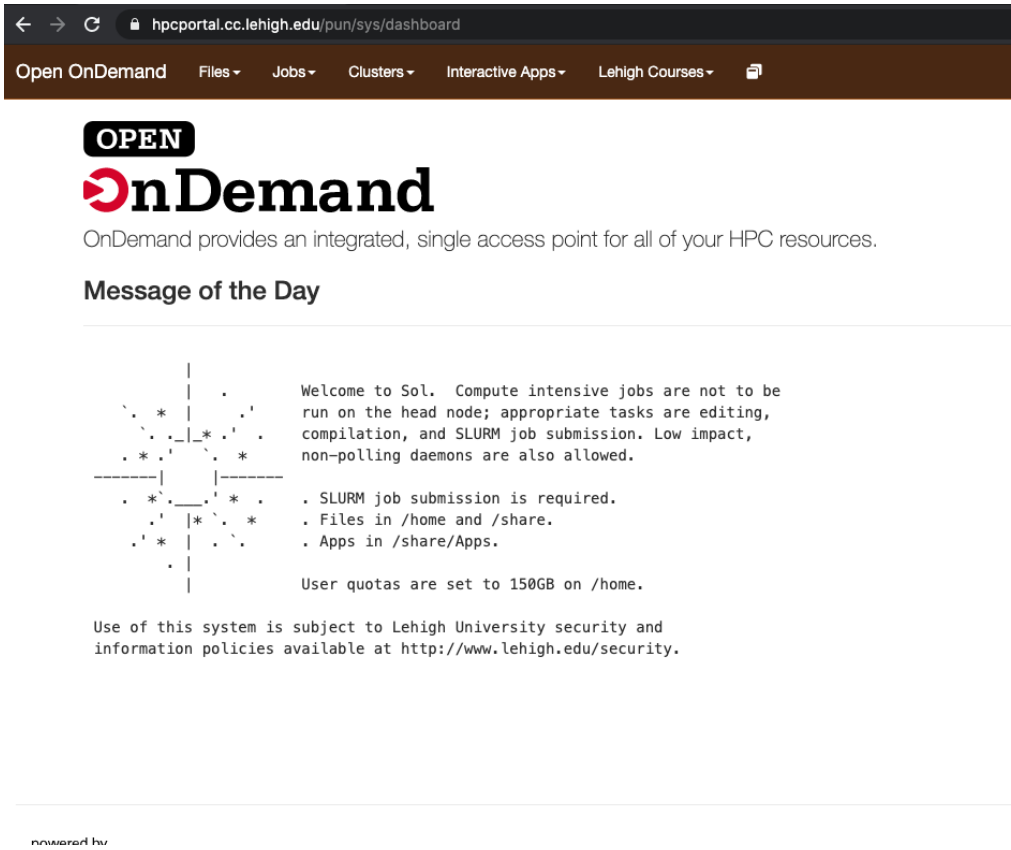
1.4.2 Installing Git

We will work with Git using GitHubDesktop. To download GitHubDesktop, navigate to https://desktop.github.com/, click on "Download for XXX" where XXX is Mac, Windows, or Linux, and follow the defaults given by the GitHubDesktop installer.

1.5 Access Jupyter notebook on the LU HPCC

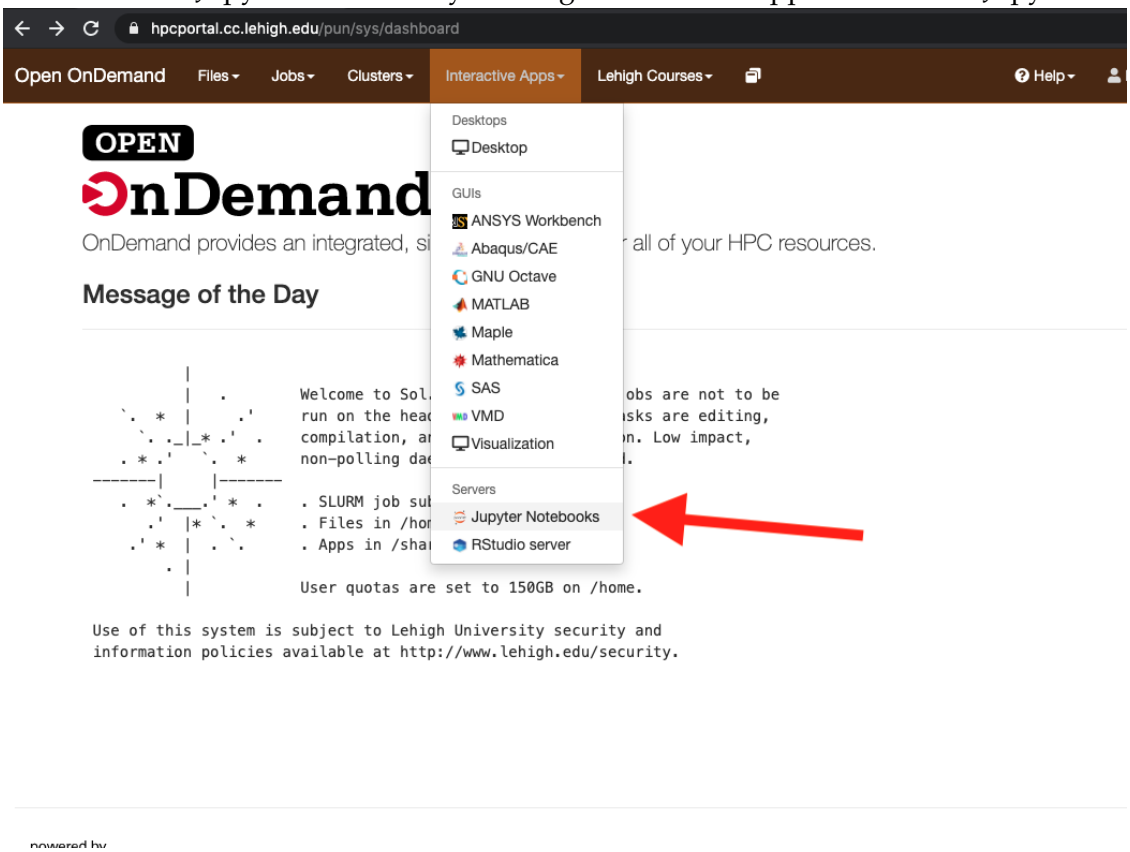
If you do not want to download Python 3 and Jupyter to your local computer, we can use resources provided by Lehigh to use Jupyter and Python. To access Jupyter Notebooks and python via Lehigh University's cluster computer, use a browser to navigate to https://hpcportal.cc.lehigh.edu/. You'll be asked for a username and password. Type in your Lehigh username, password, and press enter.

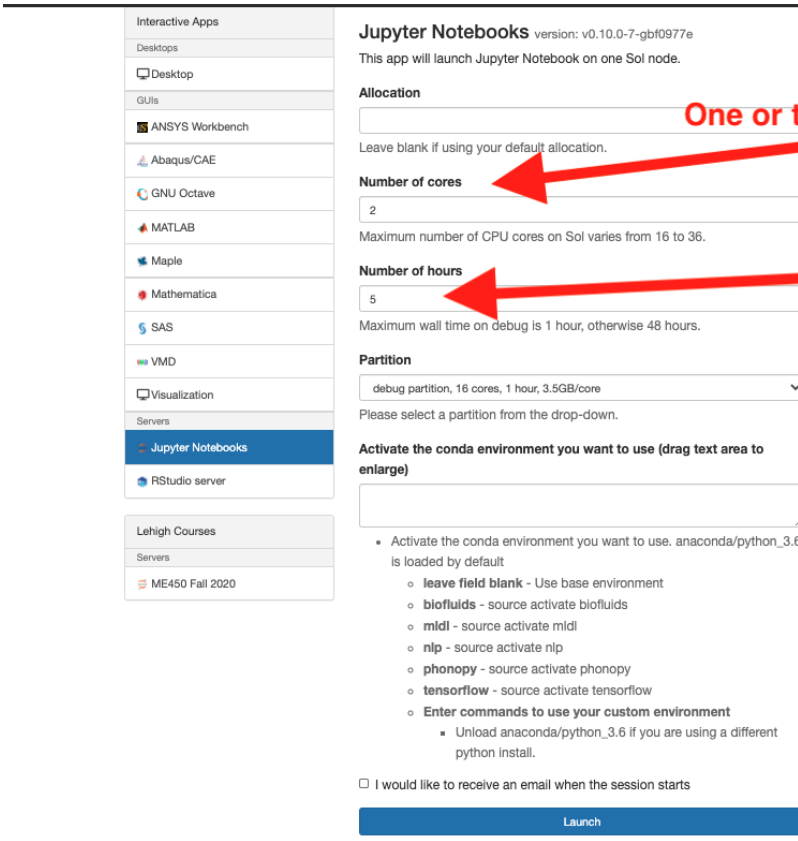




You'll be served the following site

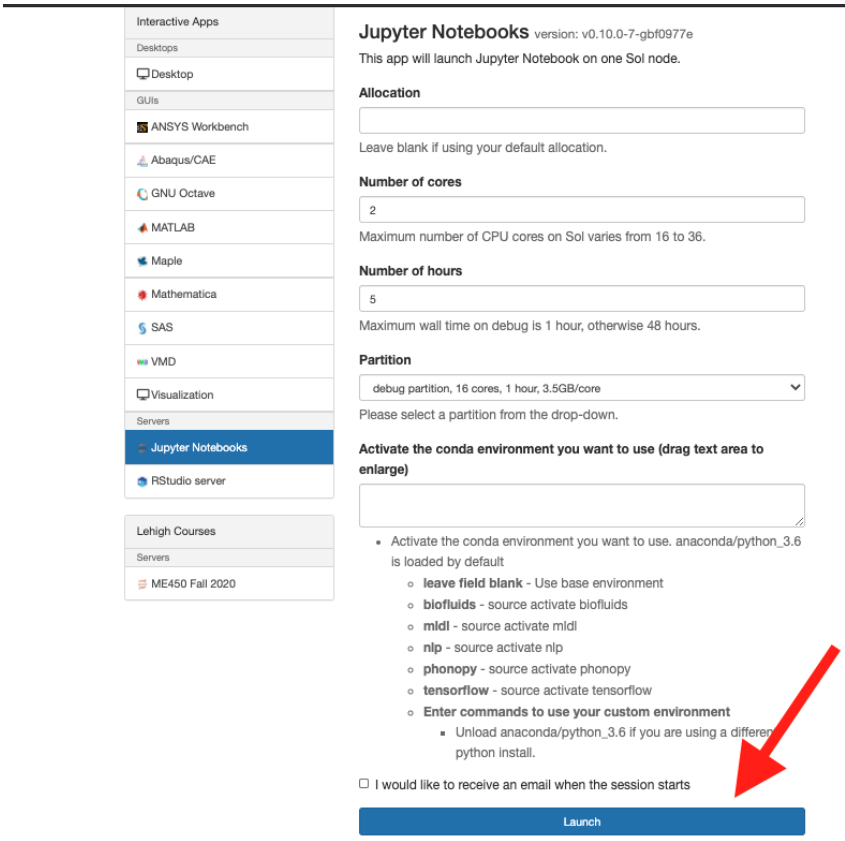
You can use a Jupyter notebook by clicking “Interactive Apps” and then “Jupyter Notebooks”.

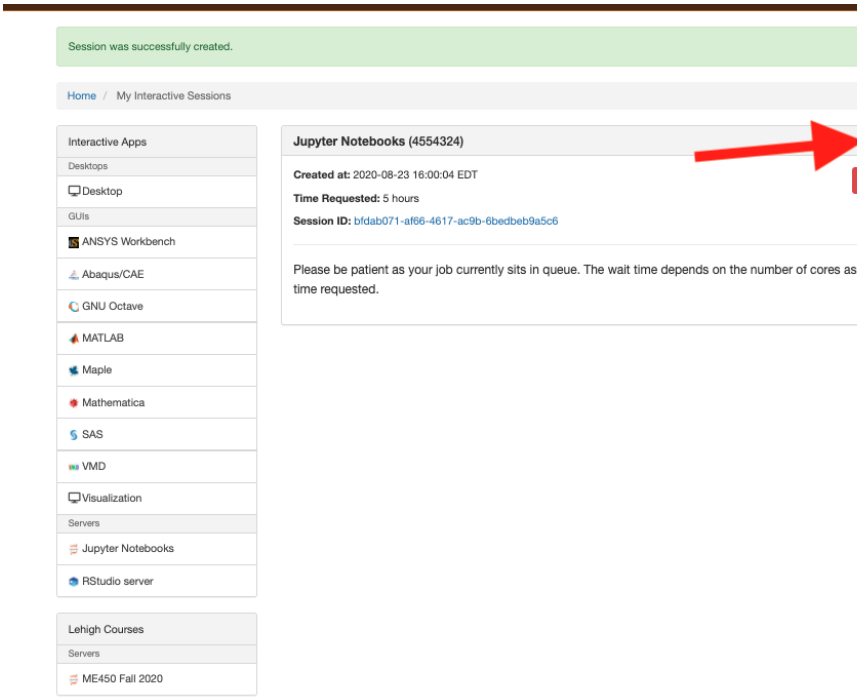




Next you'll be presented with an allocation screen.

When you use a resource on Lehigh's cluster computer you need to request the exact resources you need. After a request is made your computer is put "in line" until the cluster computer allocates your requested resources for the time you requested. If you request more cores or more time it may take a few minutes for the cluster computer to find the right resources for you. When you specify the number of cores and number of hours requested, click "Launch"





You should see your requested "job" Queued.

After a few minutes, when the cluster has made space for your requested time and number of cores, you're screen will change to the following

And finally you will be able to access a Jupyter Notebooks server.