Module 2: Jupyter Notebooks: Using Atlas

Applied AI is powered by code--Python code.

In most cases, especially in the early testing and development stages, AI code is most easily written, and run, using Jupyter Notebooks, an open standard allowing documents to mix formatted text and images with interactive code and the results from running that code.

> Info: Jupyter is a free, open standard maintained by The Jupyter Project. While it supports over 40 programming languages, its name is derived from the initial three: **Ju**lia, **Pyt**hon and **R**. Jupyter grew out of the iPython (interactive Python) notebook project, which explains the `.ipynb` file extension used for Jupyter notebooks.
>
> The "literate programming" method facilitated by code notebooks facilitates documentation, iterative code development, and visually rich presentation of code and results.
>
> After initial development, it is still common to export code as a text file that can be run as a script. This is easier to automate and for long-running scripts. So, while notebooks are a popular development and educational tool, they do not completely replace plain-text script files, especially for production runs.

Part of this module will get you up and running in one of many available Jupyter systems. We will walk you through some of the options and provide step-by-step instructions to get started.

Module 2 Objectives:

Students will be familiar with Jupyter notebooks as the predominant tool for AI data analysis and model development. Students will be able to:

a. Select the appropriate compute environment and start a Jupyter Notebook.
b. Explain what Jupyter Notebooks are, and why they are used in data science applications.
c. Understand what Markdown is and what makes it useful.
d. Run code blocks and embed images and graphs in Notebooks.

Step 1: Create your repository from the template

Most of the *Practicum AI* exercises will start with you using one of our template repositories to create your own repository. This allows you to get the latest version of our exercises and put a copy in your own space where you can work on the exercises using git and GitHub.

> **Note**: You will need a GitHub account for this. You can create your account at https://github.com/ Click the **Sign up** button.
>
> **Please note that for official USDA work, you may *not* use a free GitHub account. Instead, you must use USDA's GitHub Enterprise Cloud platform. Your unit can purchase GitHub Enterprise Cloud**

licenses through the SLIM system, which is also used for purchasing other centrally managed software. If you are only using GitHub for the Practicum AI course, you may use a free account.

- The template for this module is at: https://github.com/PracticumAI/computing-for-ai
- From that site, click the green **Use this template** button and select **Create a new repository**.
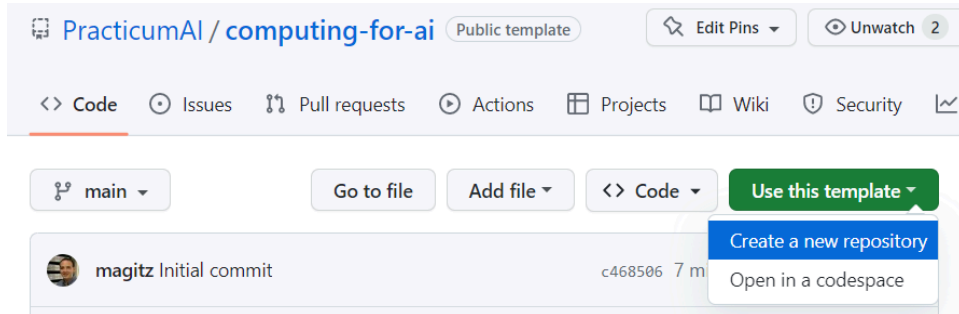
- On the new window, you may need to select the owner. **Select your GitHub account**. Then type a name for the new repository. We suggest using the same name as the template: **computing-for-ai** in this case.
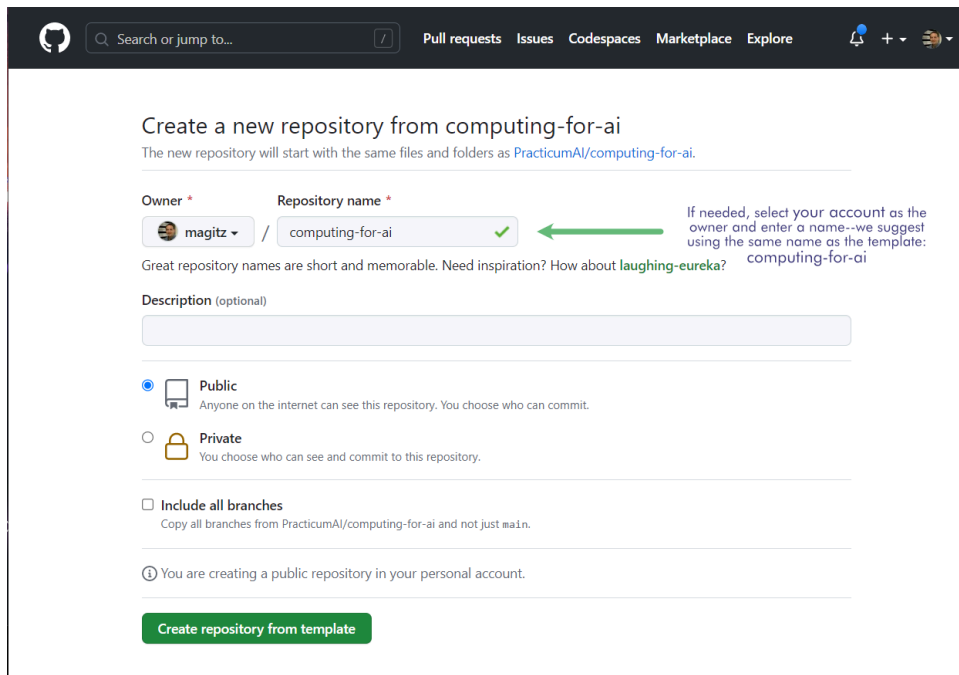
- Click the **Create repository from template** button.
- This will now create your own copy of the repository. This is your own playground for you to work in. GitHub will show you the template from which you generated the repository so you can always get back to the source, but this is yours to use.
- Now you are ready to set up a Jupyter session in step 2.

Step 2: Launch Jupyter on Atlas

Using Atlas requires that you have a SCINet account.

Launching a Jupyter Lab session on Atlas involves specifying the resources that your session will use. The job scheduler on Atlas needs information about the resources your job will need and account to use. See the details below for the settings to use.

We will use the web interface Open OnDemand (OOD) to launch and connect to Jupyter sessions. OOD provides an easy interface to provide the resource request, submit the job to the scheduler, and connect to the Jupyter server once it starts.

Steps to launch a JupyterLab session

- Login at: https://atlas-ood.hpc.msstate.edu/pun/sys/dashboard
- Use the USDA LincPass method. If you do not have a LincPass, see **alternative log-in instructions on the SCINet website**.
- Click on the **Interactive Apps** menu and select **Jupyter**
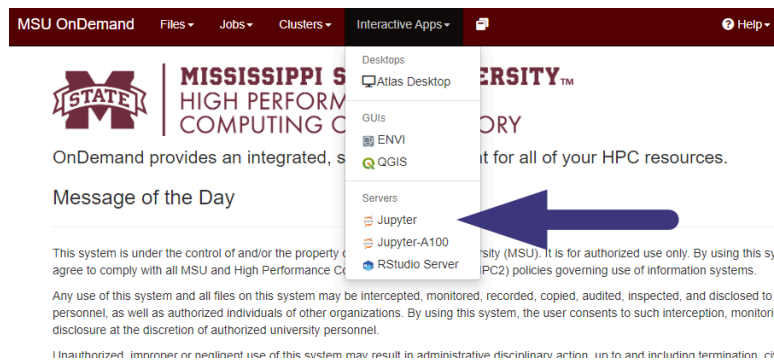
- The next screen allows you to request resources to run a job on Atlas.
- See the image for reference, but the suggested resource request for most of the *Practicum AI* courses is:

- o Lab or Notebook: **JupyterLab**
- o Account Name: **scinet_workshop2**
- o Partition Name: **atlas**
- o QOS: **normal**
- o Number of hours: **2**
- o Number of nodes: **1**
- o Number of tasks: **1**
- o Additional Slurm Parameters:

    --reservation=workshop --mem=16G
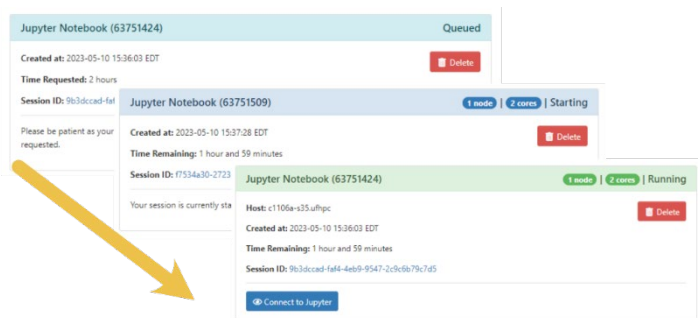
- After entering the information, click the Launch button

- On the next page, your job will have a card.
  The card will start with a light blue border with "Queued" in the upper right. That means your job has been submitted to the scheduler and is waiting for the scheduler to find resources (or is waiting for resources to become available in your group's allocation). Once the resources are available, the card will change to a dark blue border with "Starting". Finally, when your job is ready, the card will have a green border saying "Running" and a button to click to "**Connect to Jupyter**".

- Clicking the "Connect to Jupyter" button will open a new tab and should open with an interface like what is pictured—on future connections, it will normally open with the documents you had opened last time.
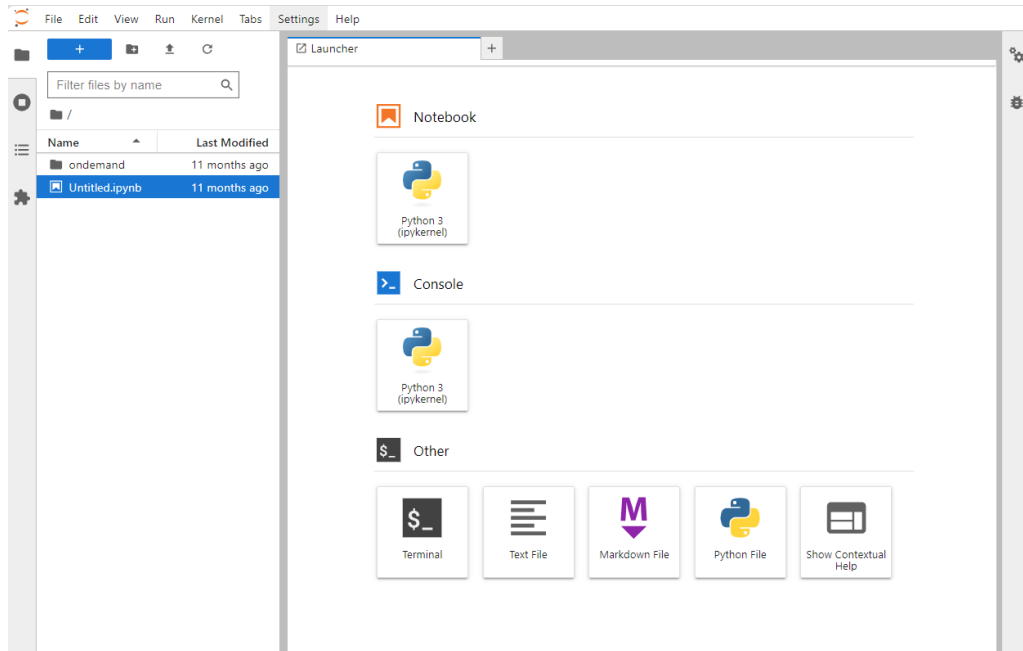

Step 3: Cloning your repository onto the HPC system

- In a different tab, go to your GitHub repository for this course. It should be at https://github.com/ then your GitHub username, then the repository name you provided, e.g. "computing-for-ai".

- Click on the following three buttons:
    1. The Green "<> Code" button
    2. The HTTPS button
    3. The copy button to copy the URI (Universal Resource Identifier—a more general term than the more familiar URL which is just for links).
    o See the image for reference.
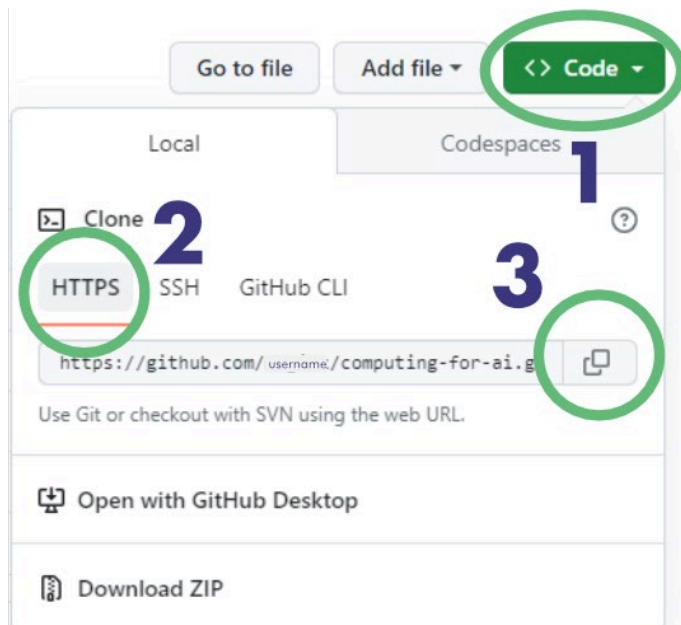
- Return to your Jupyter tab and click on the Terminal Icon to open a terminal on Atlas.
- You won't need the terminal for a lot but will need to use it for git commands.
- We can clone the git repository from GitHub by typing:

```
git clone <paste in your URI copied in step 3 above>
```
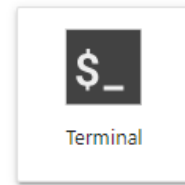
- There should now be a folder called "computing-for-ai" in the left navigation pane. Click the folder to open it.
- **Open the 01_intro_to_jupyter.ipynb** notebook.
- Read through the notebook and follow the exercises.