Python for Geospatial Big Data and Data Science Using the FASRC

Exercise 2

# Run a Python program on a compute-node

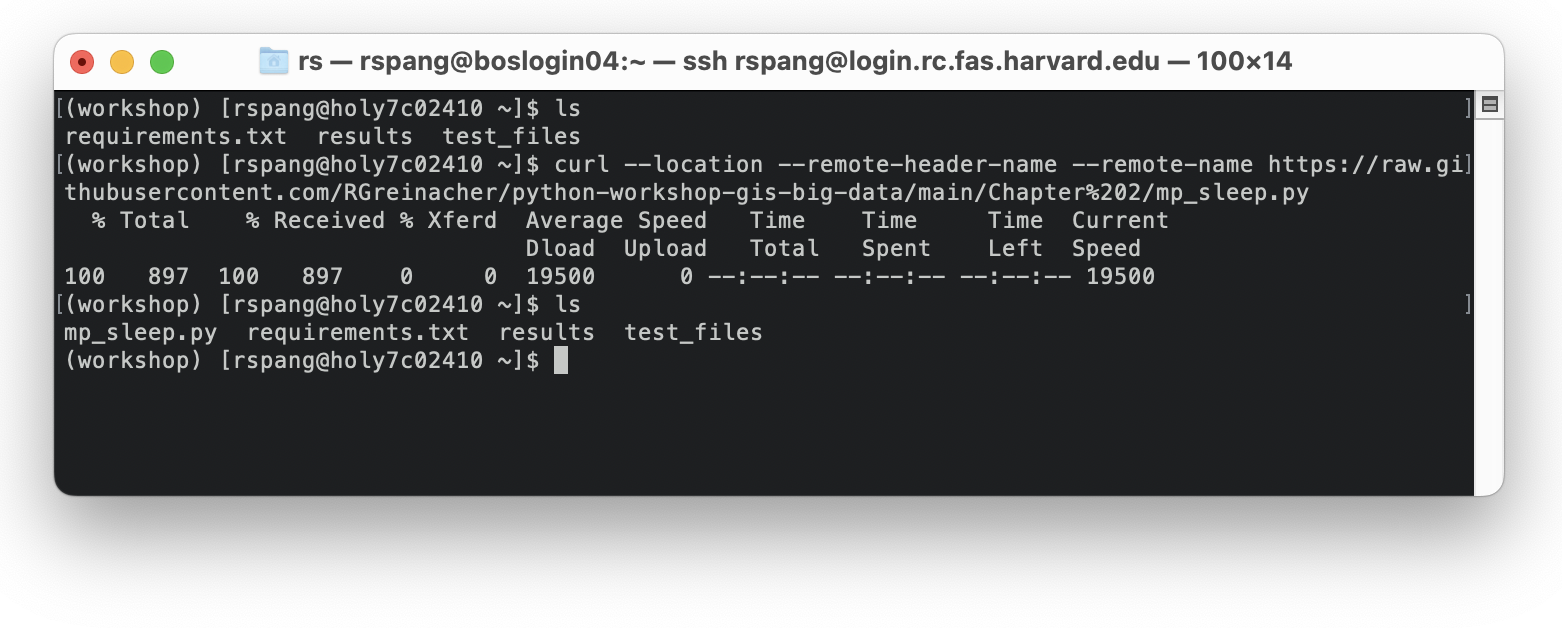
Ensure you are e still connected to a compute node, have the Python module loaded, and our workshop environment activated. If that’s not the case, see Exercise 1.

Download the “mp\_sleep.py” program from the course GitHub page. Copy the following code and execute it on a compute-node:

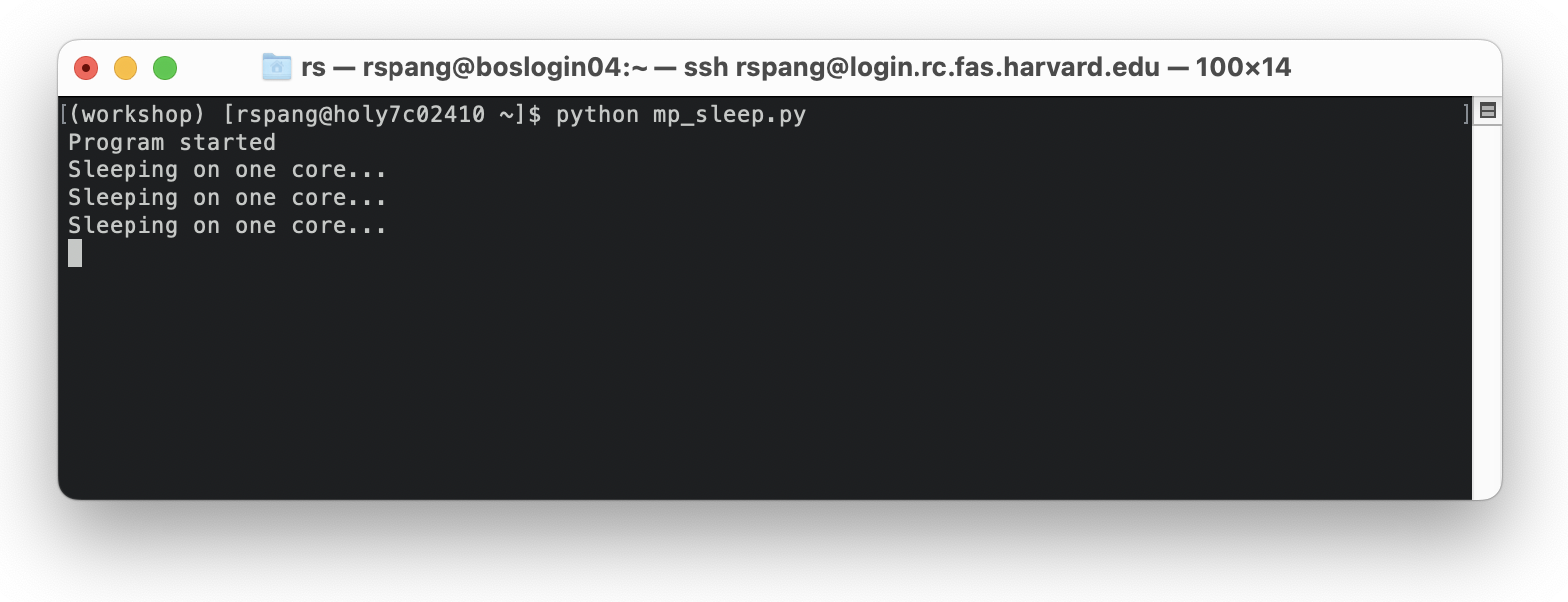
curl --location --remote-header-name --remote-name https://raw.githubusercontent.com/RGreinacher/python-workshop-gis-big-data/main/Chapter%202/mp\_sleep.py

The screenshot below shows the contents of my home folder, then the curl command, and lastly the new contents of my home folder, having a new file: mp\_sleep.py

Caution: generally, be careful what to download from the internet and make sure you trust the resource you download. It makes sense to inspect the file in your browser first.



Execute the script you just downloaded:



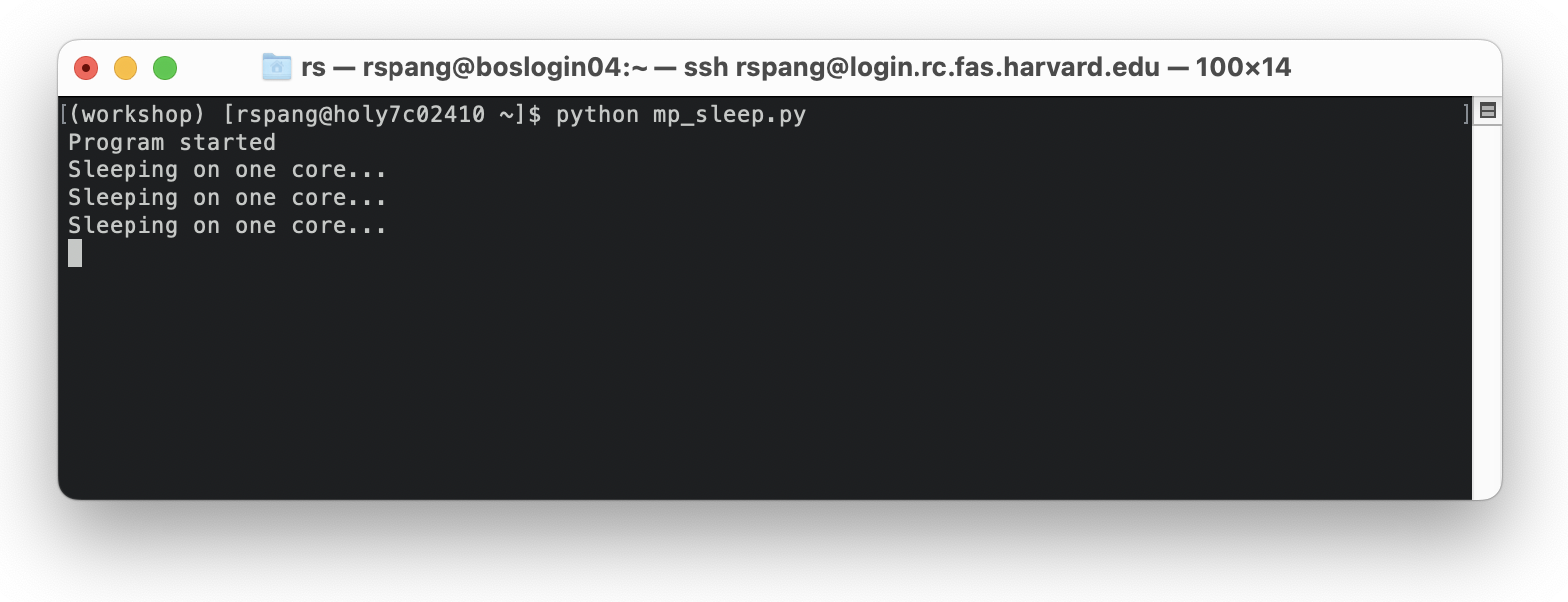
# Monitor the execution in a second window

An interactive shell connection runs in your terminal; while a program is running, you cannot use the same terminal for different purposes. However, you can start a second terminal and start a second SSH connection. While this will allow you to run SLURM monitoring tools, it is not guaranteed that you will be connected to the same node if you request a second (interactive) session with SLURM.

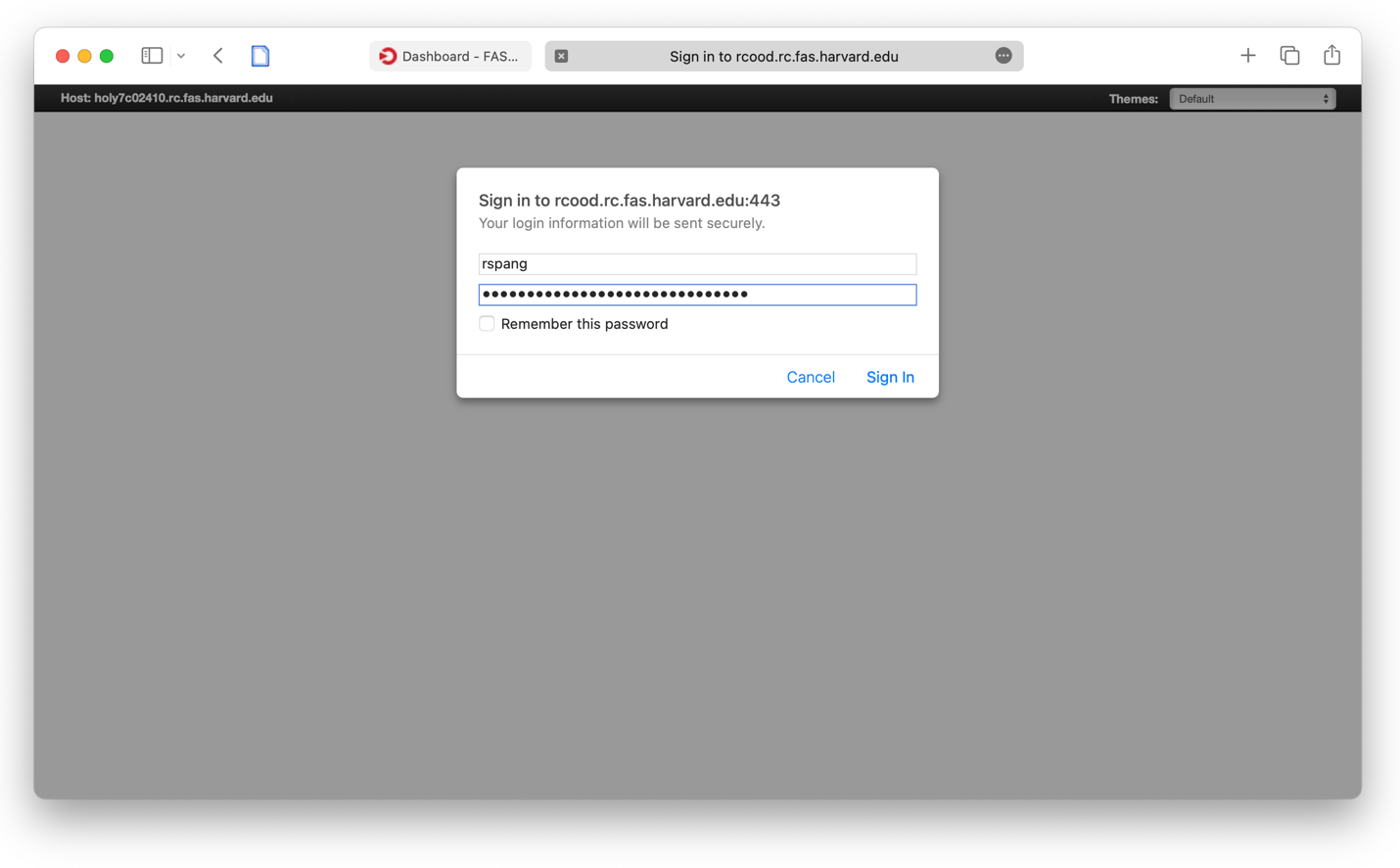
However, there is a web-tool that allows you to connect to the same instance from your browser! Open the following link

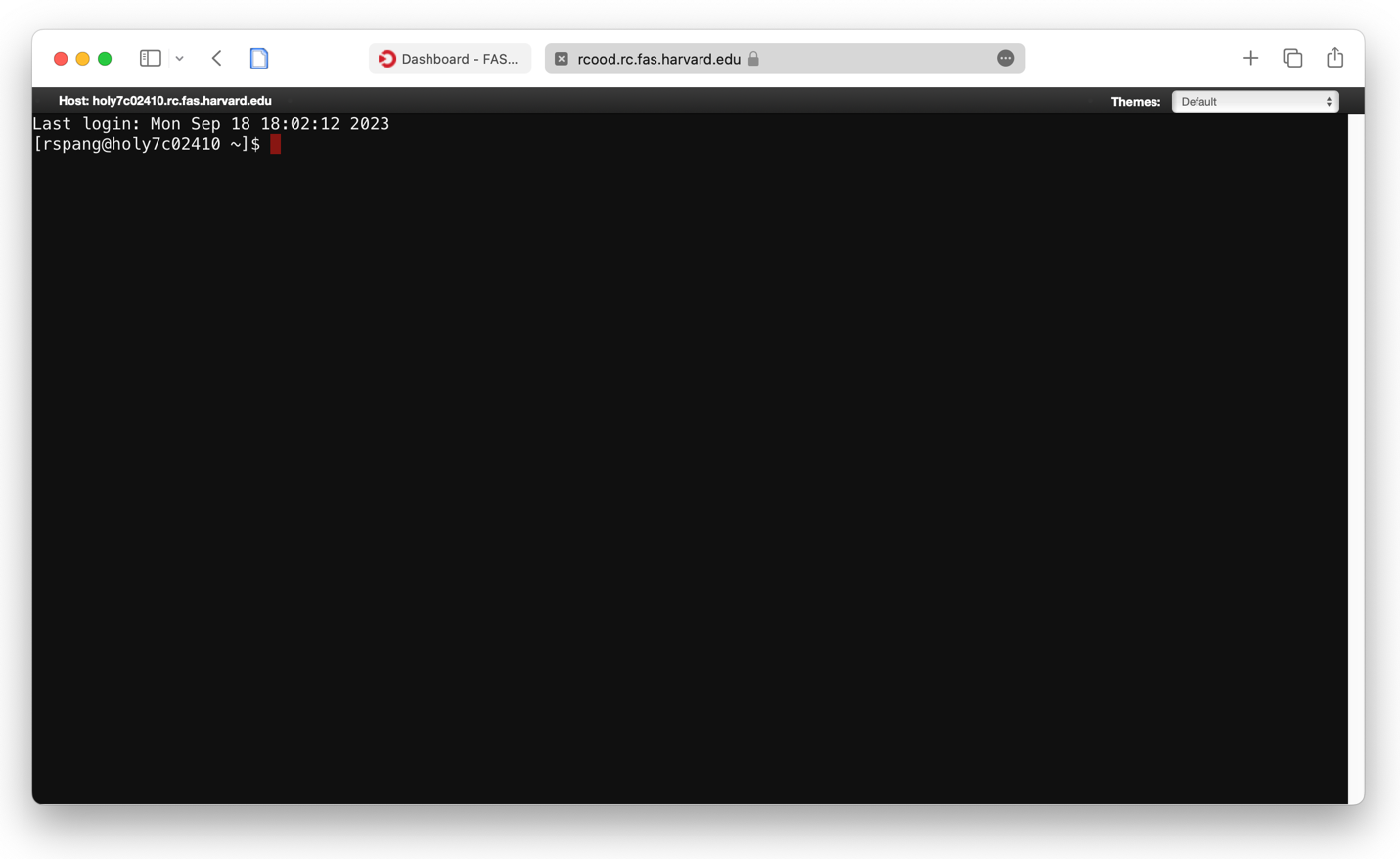
https://rcood.rc.fas.harvard.edu/pun/sys/shell/ssh/COMPUTE\_NODE.rc.fas.harvard.edu

in a browser; replace “COMPUTE\_NODE” with the node-ID you are connected to.



In this example, the node-ID is be “holy7c02410”.

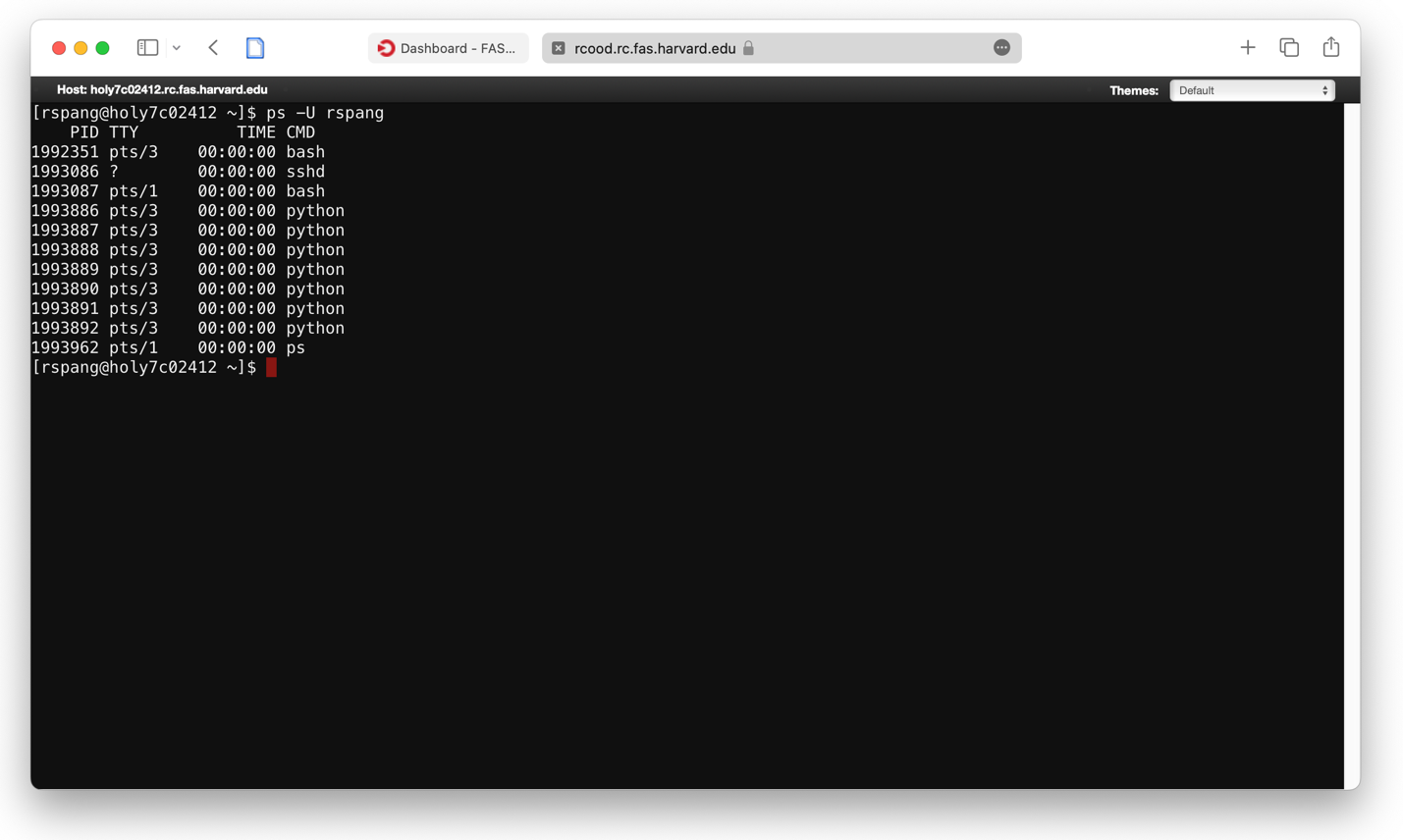
If you are asked to login first, provide your FASRC credentials:

Then, a terminal-like web page shows up.

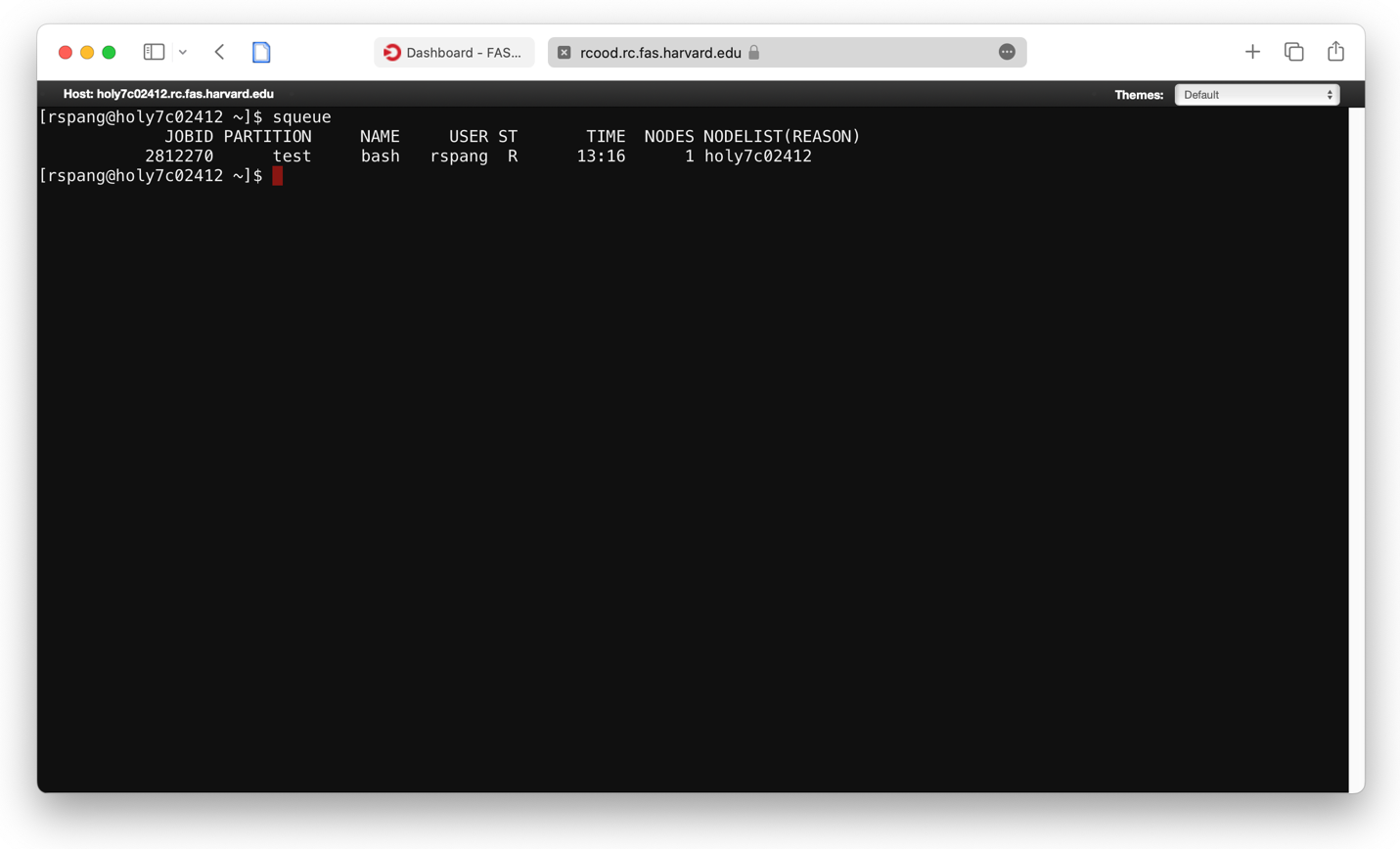
Here, you can run all commands as you would in a normal terminal. For example, you can monitor how many processes you are currently running. Use the following command to obtain a list of all processes in your name:

ps -U username

Replace “username” with your FASRC username.



The SLURM command squeue returns a list of all currently running jobs. This will return (at least) the one interactive session you used to start the python program.



# What you learned in this exercise:

* How to download a python script from the internet on the command line
* How to run a python script that uses multiple processes
* How to start a browser-based shell to connect to an ongoing (interactive) session
* How to monitor processes and jobs while running