

Prelab 01 | Linux and Bash

INTRODUCTION

This prelab will provide an introduction to the software and computing environment that we will use this semester and instructions to install it on your own computer or use it through a web interface. For most labs we will use Jupyter Lab/Notebook to complete the assignments. This can be done on your own computer, offline if needed. Or, you can access Jupyter Lab online through the Center for Integrated Research Computing (CIRC) bluehive cluster. In this lab we will use the command line environment (bash), which is also accessible via Jupyter Lab/notebook.

Bluehive is a linux server available to the UR research/teaching community. Being able to login and access other computers through the command line is necessary when running computationally intensive programs. Most computational biologists at UR utilize the BlueHive computing cluster https://info.circ.rochester.edu/#BlueHive/Getting_Started/, which has 8,644 processor nodes and a total of 41,272 GB of memory.

Prelab objectives

- Access Jupyter Notebook online at bluehive
- Install Jupyter Lab/Notebook on your computer (optional)
- Download and install file transfer software (optional)

Python and Jupyter Notebook/Lab/Hub

All the labs in this class will be done in a terminal or Python notebook accessed through Jupyter Hub (online) or on your own install of Jupyter. Python, Jupyter Notebook, and Jupyter Lab are described below, followed by instructions for installing them using Anaconda or using the web interface, JupyterHub.

- Python is an interpreted, high-level, general-purpose programming language. As opposed to C/C++ or Java which are compiled, interpreted languages like Python translates each statement into a sequence of one or more subroutines, and then into another language (often machine code).
- Jupyter Notebook (formerly IPython Notebook) is a web-based interactive computational environment for creating Jupyter notebooks documents. A Jupyter Notebook is a document containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension. The main advantage of a Jupyter Notebook is that code (executable) can be interlaced with comments, graphs and other information that is not supposed to be executed. Jupyter Notebooks can run code in R, Python and open bash terminals.
- JupyterLab is the next-generation user interface for Jupyter Notebooks. Its main advance is in a better user experience by arranging multiple documents and activities side by side in the work area using tabs and splitters.
- JupyterHub is a multi-user server for Jupyter Notebooks and is installed on bluehive. This is a good option if your computer is having problems or isn't available for some reason. However, using Jupyter Notebooks on bluehive can be slower, especially when everyone tries to login at the same time.

Dual-authentication and VPN

In order to login to the bluehive server from off campus you will need to setup dual-authentication in order to login to the university's virtual private network (VPN).

A VPN enables a private network over a public one for enhanced security and functionality. When working on campus and already connect to the campus network a VPN isn't necessary. To enroll in dual-authentication follow the instructions here:

<https://tech.rochester.edu/services/two-factor-authentication/>

Once you have enrolled, install a VPN client here:

<http://tech.rochester.edu/services/remote-access-vpn/>

Start the VPN client software and login using your university NETID and PASSWORD credentials. Note that when logging in you will receive a push or SMS passcode on your phone that must be entered into the VPN client.

Once you've connect to the VPN, or you are already on the university network, you can access Jupyter Notebooks through bluehive. No connection is needed to download and install the software.

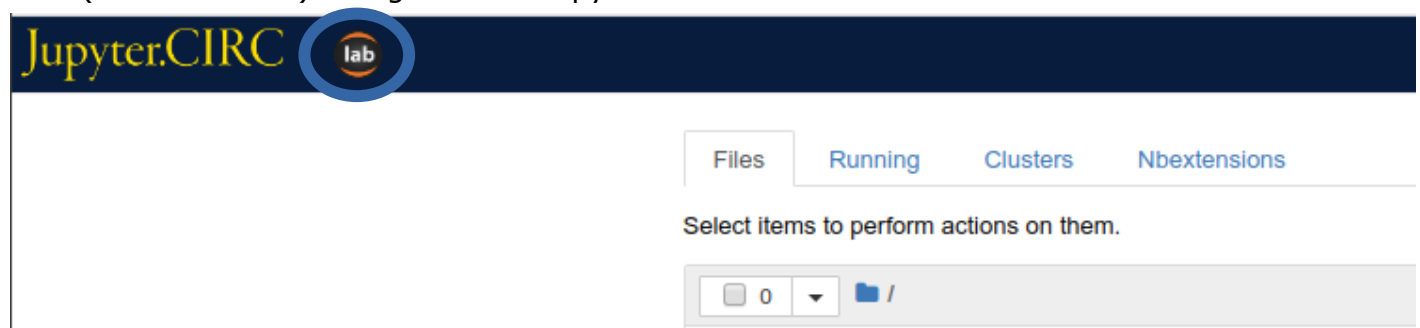
Access Jupyter Lab/Notebook online at bluehive

JupyterHub is installed on the course server: **bluehive.circ.rochester.edu**. To access, follow this link:

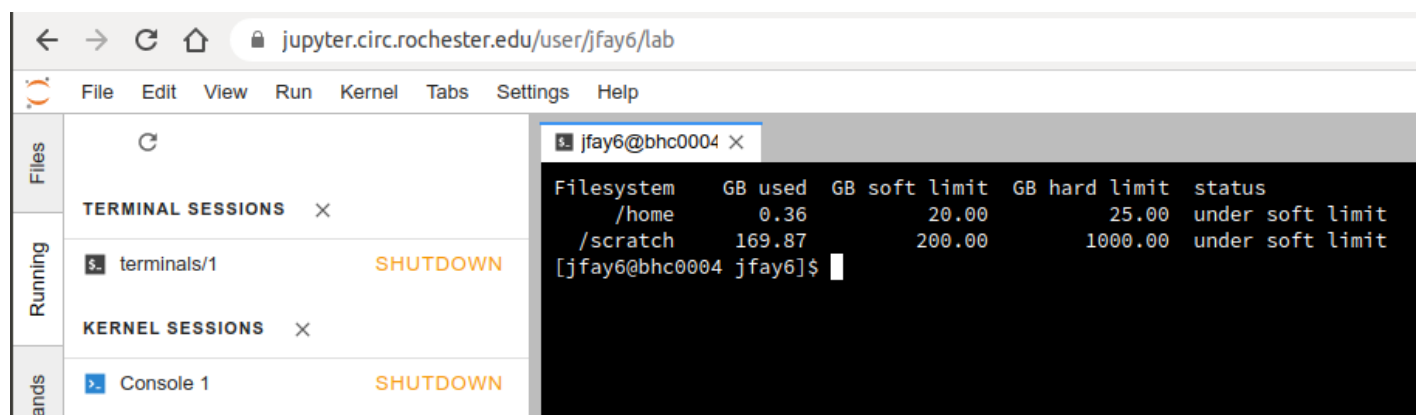
<https://jupyter.circ.rochester.edu>

(remember to approve the dual authentication on your phone to get access)

To login to bluehive, follow the link, use your University NETID and password to login, and click "spawn". There is no need to request more than 1 CPU or additional memory. By default you'll be logged into the older Jupyter Notebook environment. In the upper left, click the "Lab" icon (circled in blue) to log into the Jupyter Lab environment.



To start a bash terminal (when you start Lab 01), scroll down to the bottom of the Launcher window and click on "Terminal" under "Other". This will launch a bash terminal.



Install Jupyter Lab/Notebook on your computer using Anaconda

OPTIONAL: this option enables you to work offline and without having to wait for a bluehive session.

Anaconda is a tool that makes installation of Python and Jupyter easier. Use the following link: <https://www.anaconda.com/>

Use "Get started" or "Products - Individual Edition" to download the installer. Follow the instructions to install, it will install Python and Jupyter Lab/Notebook on your computer. However it will take some time, >500 mb download and 5-10 minutes to install. Follow the installation instructions <https://anaconda.cloud/support-center/installers>. Once installed you should be able to launch a Jupyter Lab session in your computers browser. It should be nearly identical to the bluehive sessions, but it will be run locally on your computer.

Download and install file transfer software (SCP/SFTP)

OPTIONAL: this option may be needed to get Jupyter Notebook and other files needed to complete the lab onto your own computer. These files will also be put on Blackboard, in which case this step is not necessary.

You probably will not need to use this software, but instructions are included here just in case you need them. To either send or retrieve files from bluehive you will need SCP/SFTP file transfer software.

Secure Shell (SSH) is a cryptographic network protocol for communicating securely over an unsecured network. It enables one to login to a remote server's shell environment.

Secure copy protocol (SCP) is a means of securely transferring computer files between a local host and a remote host or between two remote hosts using the SSH protocol. SFTP also enables secure transfer of files using the secure file transfer protocol (SFTP).

Host keys

The first time you login to a server such as fauna, you be told that the server's host key is unknown. You can respond to accept the host key as a trusted host and continue. The host key is simply a way of verifying that the server is who it says it is. The act of "accepting" a public key essentially authenticates that key to that server address. In practice, every connection to that machine thereafter will check against the public key that was originally recorded at first connection. This procedure protects the client-side machine from malicious attacks.

Mac/Windows Setup and Test

For Mac: FileZilla (<https://filezilla-project.org/>) is a popular file transfer client. Cyberduck (<https://cyberduck.io/>) is another population client and can be used as well.

For Windows: WinSCP (<https://winscp.net/eng/download.php>) is the most commonly used, although Cyberduck and FileZilla also offer windows software.

1. Download and install FileZilla, Cyberduck or WinSCP.
2. Upon starting add the following information
 - Host: `bluehive.circ.rochester.edu`
 - Port: 22

- Protocol: SFTP – SSH File Transfer Protocol
- User: Your UR NETID
- Password: Your UR NETID password

Connect and initiate a connection to the server. Accept the public key if needed.

For transferring files with bluehive you can also follow similar instructions here:

https://info.circ.rochester.edu/#BlueHive/Transferring_Files/Transferring_Files/