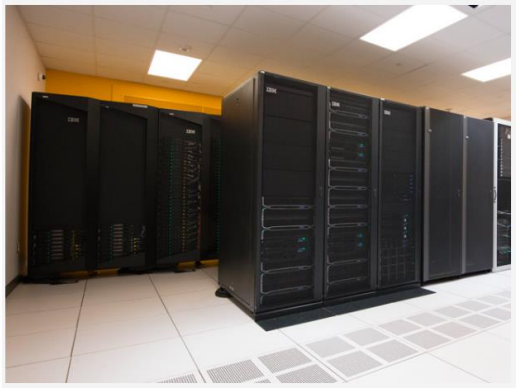
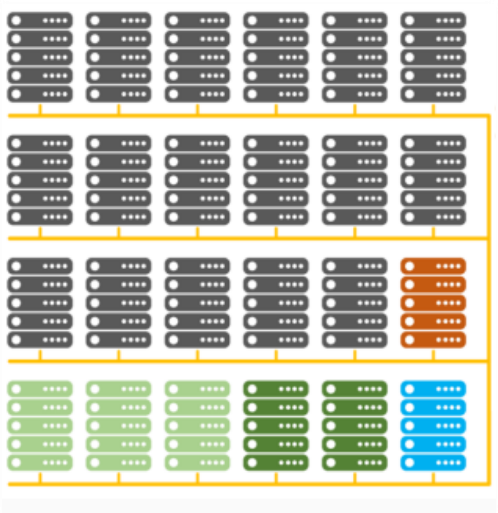
Center for Integrated Research Computing (CIRC) & BlueHive Workshop

Joshua Lawson | Simon Business School, University of Rochester | August 28, 2023

BlueHive

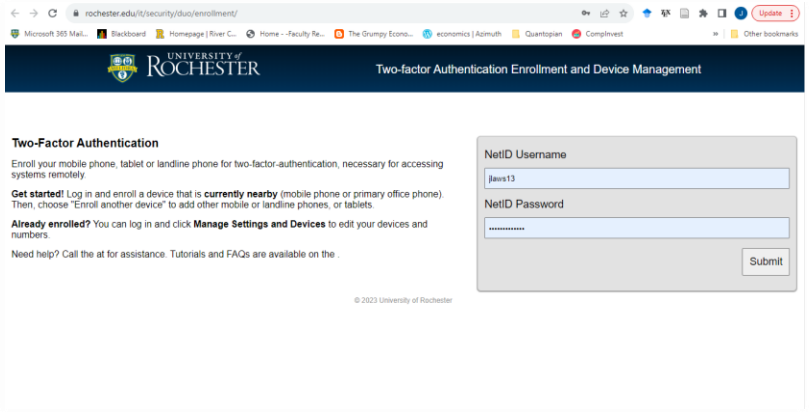
- Partition → Node → Core
- 480 Nodes, +1,400 CPU Cores = ~72TB RAM



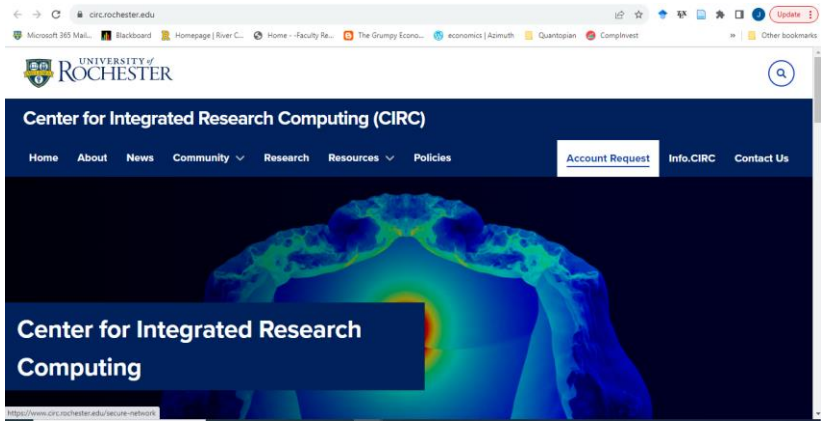


Accessing BlueHive: Step 0

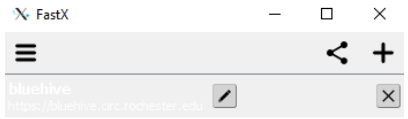
1. Enroll in DUO
    1. https://www.rochester.edu/it/security/duo/enrollment/



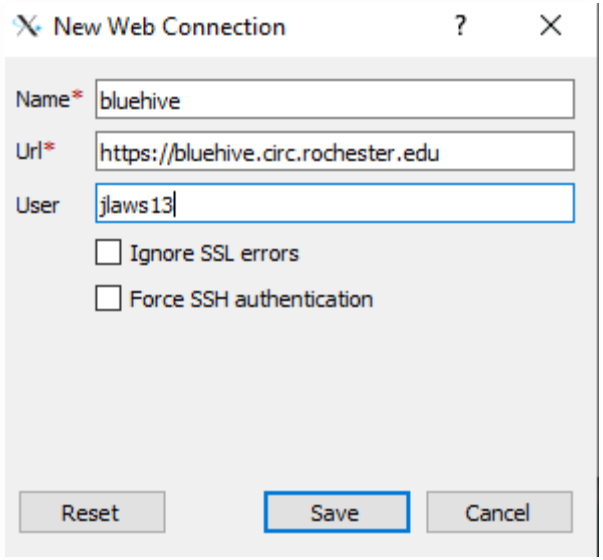
2. Request Account
    1. https://www.circ.rochester.edu/
    2. Click on ‘Account Request’
        1. Fill in information
        2. Select Mitchell Lovett as faculty sponsor group



Accessing BlueHive: FastX

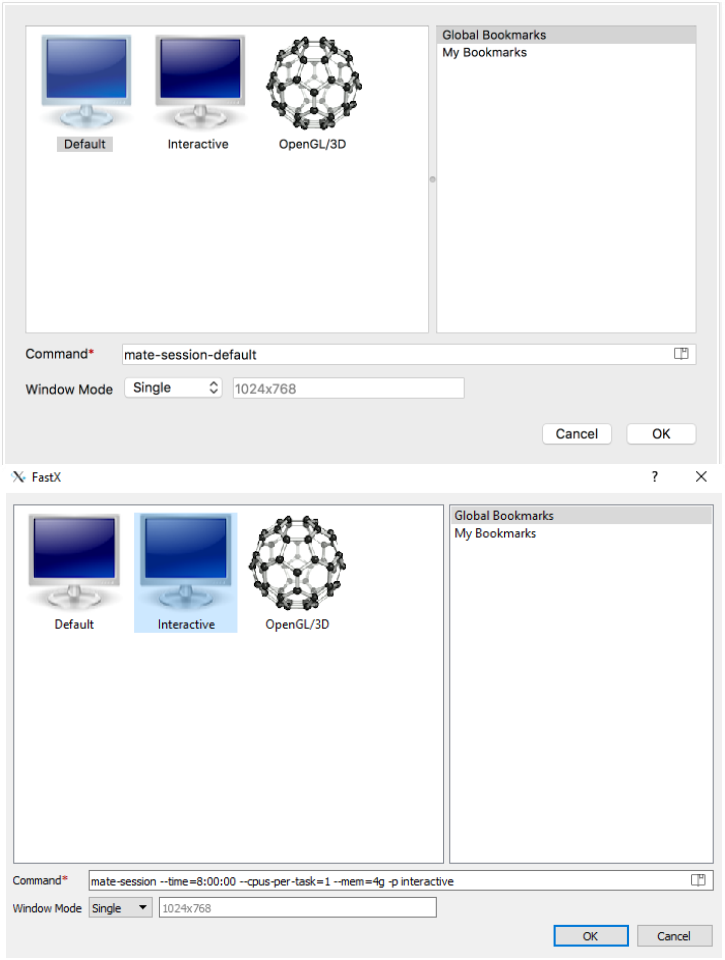


1. Install from: https://gitlab.circ.rochester.edu/circ-docs/downloads/-/wikis/windows-software
   - Under FastX, select installer
   - Complete the installation.
2. Open FastX on your machine.
3. Click the "+" on top right corner.
   - Select web
4. Enter the following:
   1. Name = bluehive
   2. Url = https://bluehive.circ.rochester.edu
   3. User = netid
5. Click Save
6. Double click the connection and use NetID + DUO to login.



Requesting Resources

- Default session
  - 30-day time limit
  - Limited computing power
- Interactive session
  - Limited availability
  - Control number of CPUs, memory, GPU, etc.



Interactive Session Commands

Command list

- --time/-t: wall time for the session
- --cpus-per-task/-c: number of cpus per task
- --mem-per-cpu/--mem: total RAM
- --w: request a particular node within the partition
- --partition/-p: select partition to run on
- --gres=gpu: number of GPUs

Open partitions

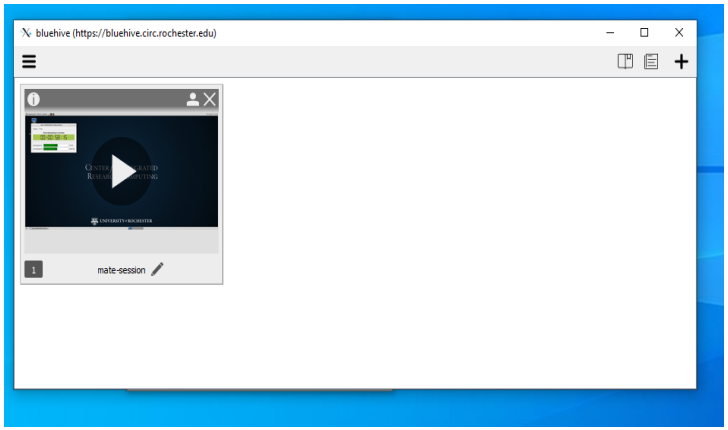
partition	wall time limit	resource limit	running jobs limit	submitted jobs limit
debug	01:00:00	cpu=48, node=2		2000
gpu	5-00:00:00	cpu=120, node=15		2000
gpu-debug	01:00:00	cpu=23, node=1		2000
gpu-interactive	12:00:00	cpu=23, node=1		2000
highmem	5-00:00:00			2000
interactive	12:00:00	cpu=40, node=2		2000
preempt	2-00:00:00	cpu=120		2000
standard	5-00:00:00	cpu=120, mem=3100G		2000
visual	12:00:00	cpu=24, node=1		2000

▼ **gpu** - 33 nodes, 56 max cores, 496G max mem, 1148 total cores, 5926G total mem

nodes	node list	cores	CPU model	mem	GPUs	GPU model	constraints
14	bhg[0004,0012-0018,0020,0022,0024-0027]	24	Intel Xeon CPU E5-2695 v2 @ 2.40GHz	62GB	2	Tesla K20Xm	ib,E52695v2,K20X
8	bhg[0032,0034-0040]	36	Intel Xeon CPU E5-2695 v4 @ 2.10GHz	124GB	4	Tesla K80	ib,E52695v4,K80
1	bhg0033	28	Intel Xeon CPU E5-2695 v3 @ 2.30GHz	124GB	8	Tesla K80	ib,E52695v4,K80
1	bhg0041	36	Intel Xeon CPU E5-2695 v4 @ 2.10GHz	124GB	4	Tesla K80	ib,E52695v2
1	bhg0042	36	Intel Xeon Gold 6140 CPU @ 2.30GHz	176GB	4	Tesla V100-SXM2-16GB	ib,Gold6140,V100
1	bhg0043	32	Intel Xeon Gold 6130 CPU @ 2.10GHz	176GB	4	Tesla V100-SXM2-32GB	ib,Gold6130,V100
7	bhg[0051-0057]	56	Intel Xeon Gold 6330 CPU @ 2.00GHz	496GB	4	A100-PCIE-40GB	ib,Gold6330,A100

Home Screen

- Wall-time limit
- Storage quota
  - /home – 25GB
  - /scratch – 200GB
- Managing files
- Run (limited) software



Transferring Files

Mac OS

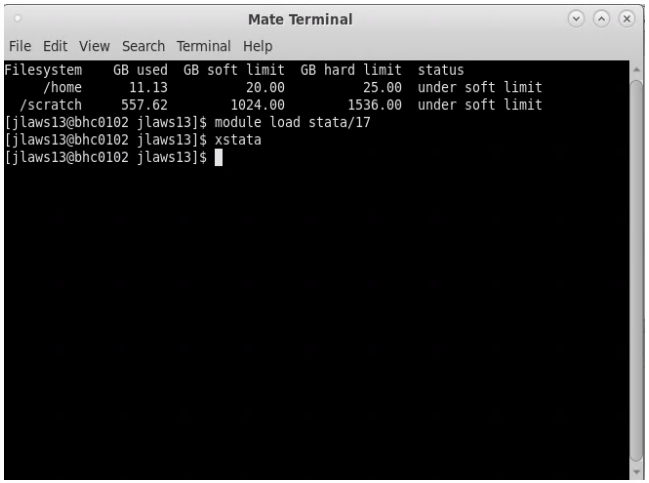
- Download and install Fetch from https://gitlab.circ.rochester.edu/circ-docs/downloads/-/wikis/mac-software
  - Unzip Fetch.tar.gz into Applications directory
  - Enter hostname, NetID, and password
- Get = download from BlueHive
- Put = upload to BlueHive

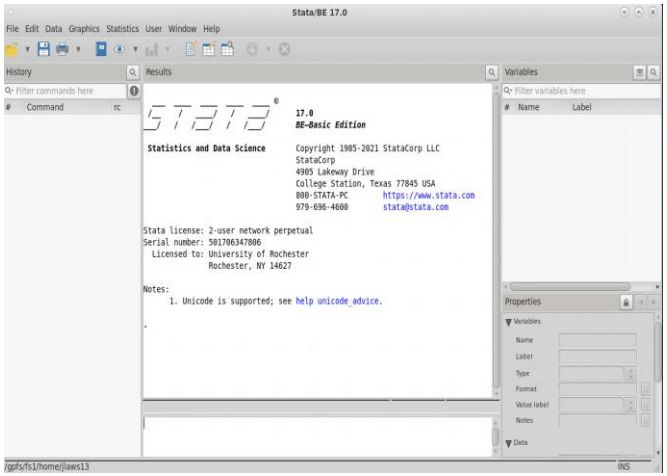
Windows

- Download and install WinSCP from https://gitlab.circ.rochester.edu/circ-docs/downloads/-/wikis/windows-software
- File protocol = SFTP
- Host name = bluehive.circ.rochester.edu
- Port number = 22
- User name = NetID
- After logging in, simply drag and drop files.

Launching Software: STATA

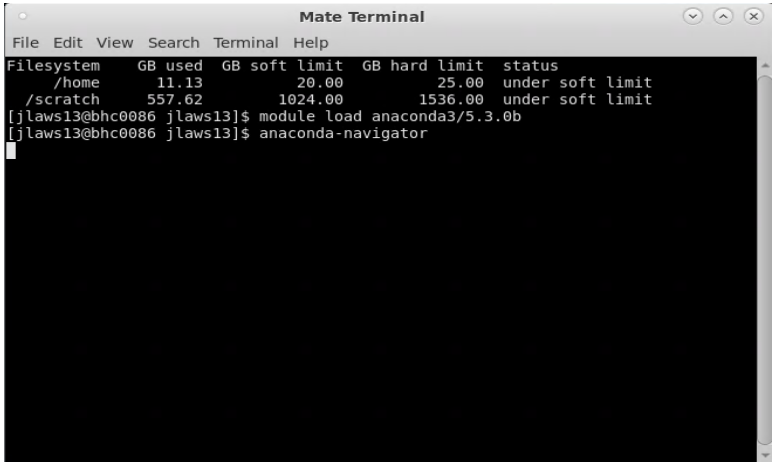
- Open Mate Terminal and type the following:
  - **module load stata/17**
  - **xstata**

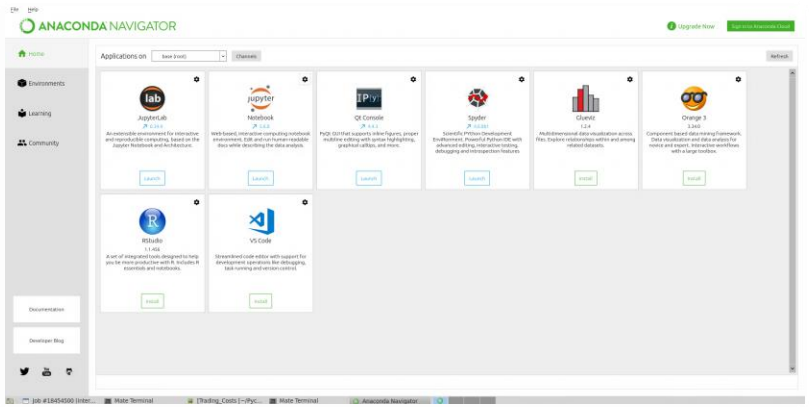




Launching Software: Anaconda3

- In a <u>new</u> Mate Terminal, type the following:
  - **module load anaconda3/5.3.0b**
  - **anaconda-navigator**
    - (might have to repeat this step)
- Now you can access Jupyter, Rstudio, VS Code, etc.
- Setup environments, manage packages

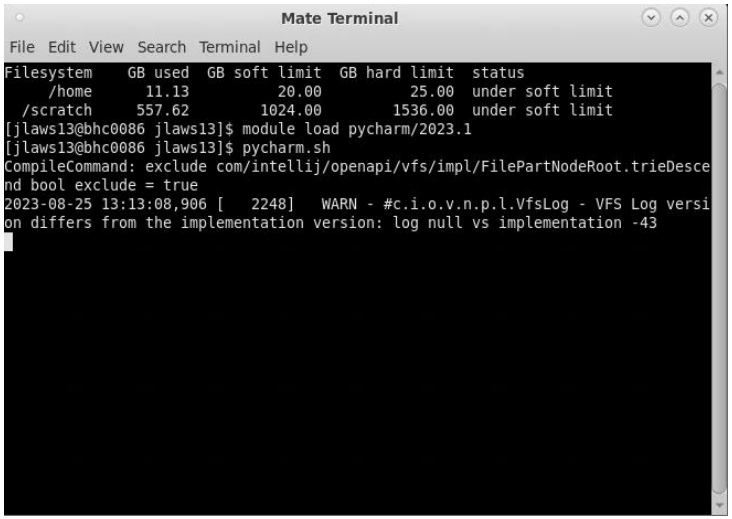




Launching Software: PyCharm

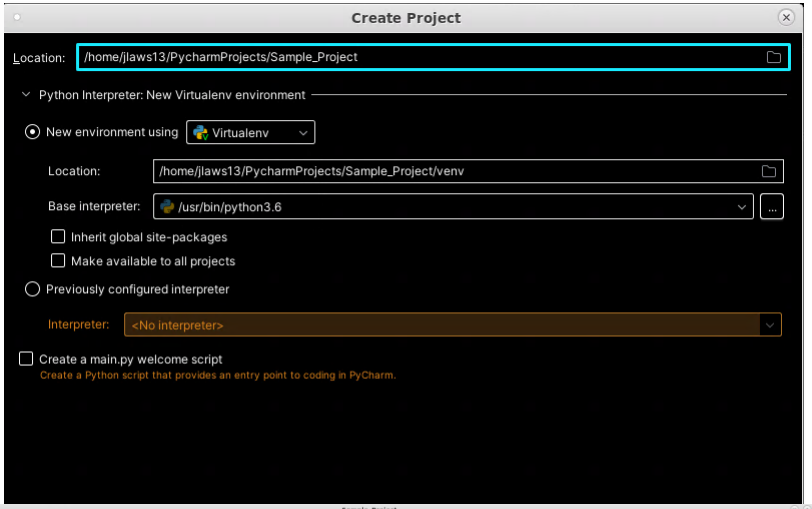
- Applications → Programming Tools → pycharm → 2023.1

- In a <u>new</u> terminal, type the following:
  1. **module load pycharm/2023.1**
  2. **pycharm.sh**



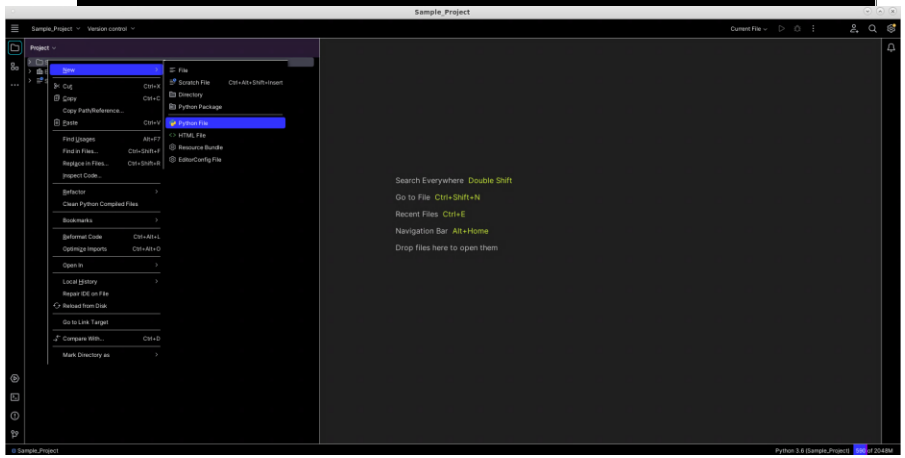
PyCharm Setup

1. Create a new project called ‘Sample_Project’
   - Recommendation: create a new virtual environment



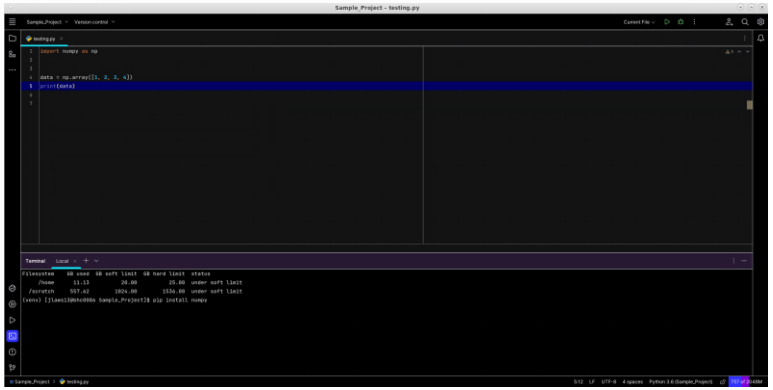
2. Create new .py file

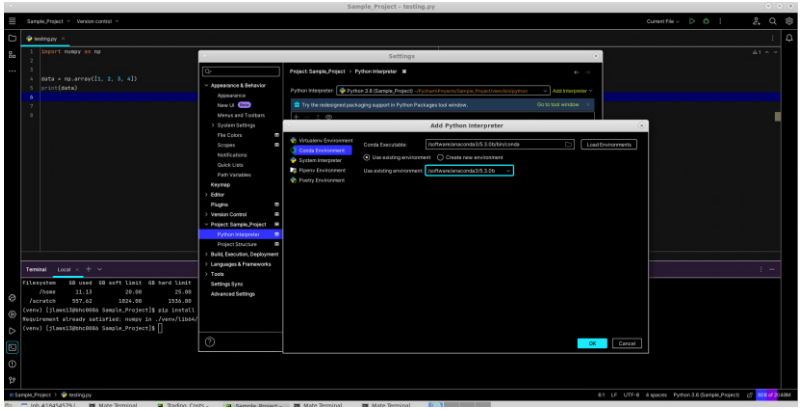
3. Install required packages



Installing Packages

1. Open terminal within PyCharm
    1. **pip install numpy**
        1. *Note: This is being installed in the virtual environment that we just created!*
2. Alternatively: Load Conda Environment
    1. Settings → Project: Sample_Project → Python Interpreter
    2. Add Interpreter → Conda Environment
    3. Path to conda Executable: /software/anaconda3/5.3.0b/bin/conda
    4. Load Environments and select.





Additional Unix Commands – ChatGPT is your friend!

- Change directory – **cd**
- List files in a directory – **ls**
- Connect to WRDS – **ssh <netid>@wrds-cloud.wharton.upenn.edu**
- Transfer files from WRDS to BlueHive **– scp <netid>@wrds-cloud.wharton.upenn.edu:<target/file/location> <destination/on/BlueHive>**

File Commands
ls – directory listing
ls -al – formatted listing with hidden files
cd *dir* - change directory to *dir*
cd – change to home
pwd – show current directory
mkdir *dir* – create a directory *dir*
rm *file* – delete *file*
rm -r *dir* – delete directory *dir*
rm -f *file* – force remove *file*
rm -rf *dir* – force remove directory *dir* *
cp *file1 file2* – copy *file1* to *file2*
cp -r *dir1 dir2* – copy *dir1* to *dir2*; create *dir2* if i doesn't exist
mv *file1 file2* – rename or move *file1* to *file2* if *file2* is an existing directory, moves *file1* into directory *file2*
ln -s *file link* – create symbolic link *link* to *file*
touch *file* – create or update *file*
cat > *file* – places standard input into *file*
more *file* – output the contents of *file*
head *file* – output the first 10 lines of *file*
tail *file* – output the last 10 lines of *file*
tail -f *file* – output the contents of *file* as it grows, starting with the last 10 lines

System Info
date – show the current date and time
cal – show this month's calendar
uptime – show current uptime
w – display who is online
whoami – who you are logged in as
finger *user* – display information about *user*
uname -a – show kernel information
cat /proc/cpuinfo – cpu information
cat /proc/meminfo – memory information
man *command* – show the manual for *command*
df – show disk usage
du – show directory space usage
free – show memory and swap usage
whereis *app* – show possible locations of *app*
which *app* – show which *app* will be run by default

Useful Links

- List of Compute Nodes: https://info.circ.rochester.edu/#BlueHive/Compute_Nodes
- Software: https://info.circ.rochester.edu/#BlueHive/Software/
- Other trainings: https://info.circ.rochester.edu/#Training/Training/
- Previous Workshops: https://info.circ.rochester.edu/#Training/CIRC-Workshop-Simon.pdf