**OVERVIEW OF USING MIDWAY2**

(Further info can be found in the [RCC User Guide](http://docs.rcc.uchicago.edu))

When you initially connect to midway2, from the terminal, you will be located in your home directory /home/$USER

* Note: Your home folder quota is 30GB. There is also a quota for the total number of files you can have in your home directory which is set at 300,000 files.

In addition to any files or folders that you have created in your home directory you will see a soft-link to your scratch directory. This is another storage location that typically a user would use for running their jobs. We will not use the scratch space. You will however use the shared group space as this is a location where some materials and tools will be located and provides a space where you can share data and files with other members of the class.

**Shell Navigational Commands**

**pwd**  The present working directory command (pwd) is used to identify

which directory you are presently located in.

**ls**  The list directory contents command 'ls' will list all files, folders,

and links that are located in your present working directory.

**ls -al**  The list all contents in long format will show more information about

each file and folder in your present working directory. The first column

lists the permissions for users, the second column shows number of files

under a folder, the third column is the ownership of the file/folder, the

fourth column the size of the file/folder, the fifth column the date the

file/folder was last modified, and the last column is the file/folder name.

**cd**  The change directory command 'cd' is used to move around on the file

system. If you type cd without any path, you will change directory to

your home directory.

**cd ../**  To go up a directory use two periods followed by a forward slash ( ../ )

after the cd command.

**cd** dir-path Using the cd command in this manner will change your location to the

'dir-path' path. For example if you have a directory under your present

working directory named lab1, you would:

cd lab1

to descend into the lab1 folder. For example to change your location to the shared

/project2/env\_bootcamp folder, you would type:

cd /project2/env\_bootcamp

**File and Folder Management Commands**

**mkdir**  some-folder To create a folder/subfolder of the name some-folder in the present

directory.

**cp** file1 file2 The cp command will copy file1 to file2

**mv** file1 file2 The mv command will rename file1 to the name file2

**cat** file1 The cat command dumps the contents of file1 to the terminal screen.

**cat** file1 >> file2 The cat command can also be used to append one file to another.

In this example cat is used to append all of the contents of file1 to the end

of file2.

**rmdir** some-folder The rmdir command can only be used on an empty directory. It will remove

the specified directory, in this case 'some-folder'

**rm** file1 The rm command is used to delete a file, in this case the file named 'file1'

**rm -rf** some-folder Be careful with this command. It will remove the folder named

'some-folder' and all the files and other sub-directories in that folder.

**Other Useful System Commands**

**which** python The which command is used to locate a command or program. In this

example we are trying to locate the python executable.

**du -h** List in human readable format the size of the files and folders in the

current working directory.

**top**  See all the system processes that are currently running on the node.

Ctrl-C to exit from top.

**tail**  <file-name> Print to screen the last few lines of a file named <file-name>.

# **Accessing group shared space**

* Upon connecting to midway2, you will land in your home directory. This directory contains no content by default.
* To access the group shared folder on the HPC cluster, you need to change directory to its location. You can do so by using the ‘change directory’ command cd as follows:

cd /project2/env\_bootcamp

* If you join a research lab, you can obtain a RCC account that is sponsored by your PI by completing the [RCC General User Account Request webform](https://rcc.uchicago.edu/accounts-allocations/general-user-account-request). By having a RCC user account under the PI, you will have access to the PI’s group shared project space, which will be located at:

/project2/<CNET\_ID of PI>

* For example if a user has a RCC account provisioned under the PI Liz Moyer, their shared group space would be accessible at /project2/moyer

**Using Environment Module Installed Software**

When you first log into Midway2, you will be entered into a very barebones user environment with minimal software available. Like other Computational resource centers, we use a module system to manage which software packages are loaded into your environment. The module system is a script based system used to configure the user environment to make available to the user selected software packages. To access software that is installed on Midway, you use the module system to load the corresponding software module into your environment.

**Basic module commands:**

Command Description

---------------------------------------------------------------------------------------------------------------

**module avail**  lists all available software modules

**module avail** [name] lists modules matching [name]

**module load** [name] loads the named module

**module unload** [name] unloads the named module

**module list**  lists the modules currently loaded for the user

See [here](https://rcc.uchicago.edu/docs/software/modules) a full list of the software modules available on the RCC cluster.

Show the list of all available software modules:

module avail

Show the list of available modules for a specific software you are interested (e.g., RStudio ):

module avail rstudio

Show the list of currently loaded modules:

module list

Show more information about a specific software package:

module show rstudio

Load a specific software package version into your environment:

module load rstudio/1.2.1335

Confirm you now have access to the RStudio software from the command line by issuing the following from the CLI:

which rstudio

This should return the path to where the binary for rstudio is located on the system. If instead you see the message “/usr/bin/which: no rstudio in”, along with a long list of all paths that were searched, it means RStudio is not accessible to you because the module has not been successfully loaded.

You should be able to launch rstudio by issuing the command:

rstudio

If you have any issues viewing the GUI when launching rstudio, then you likely have a problem with X11 forwarding that needs to be sorted. If you are using your own laptop/desktop and it is a mac/linux OS, then please review the information about installing X11 on SSH Client Setup, and the information on enabling X11 forwarding in the [Connecting\_2\_Midway2 document](https://docs.google.com/document/d/1SkQPf_lBz4c1XXPdZVBUaMgqMMfxv8sckqoZ3-jPN_8/edit).

To unload a specific software package because you no longer need it, issue the following:

module unload rstudio

**Slurm Job Scheduler**

When you first log into Midway2 you will be connected to a login node (midway2-login1 or midway2-login2). Login nodes are not intended to be used for computationally intensive work. Instead, login nodes should be used for managing files, submitting jobs, etc. If you are going to be running a computationally intensive program, you must do this work on a compute node by either obtaining an interactive session or submitting a job through the scheduler. However, you are free to run very short, non-computationally intensive jobs on the login nodes as is often necessary when you are working on and debugging your code.

If you are unsure if your job will be computationally intensive (large memory or CPU usage, long running time, etc), get a session on a compute node and work there.

There are two ways to send your work to a Midway compute node.

1.) **sinteractive** - Request access to a compute node and log into it.

For example to start an interactive session on 1 broadwell node requesting

6 compute cores for 4 hours:

sinteractive --partition=broadwl --ntasks=6 --time=04:00:00

The default disposition of the sinteractive command is to provide you access

for 2 hours to a compute node with 1 CPU and 2GB of memory.

2.) **sbatch** - Write a script which defines commands that need to be executed

and let SLURM run them on your behalf.

We will use the sbatch script method to submit jobs to the scheduler. To submit an sbatch

script that is located in your present working directory named *job.sbatch* so that it is

scheduled to run on a compute node, issue the following from the command line:

sbatch job.sbatch

There are slurm commands that can be used to query the status of a job or

interact/adjust a submitted job. The primary slurm command squeue allows the

user to monitor jobs.

**squeue** -- will list all jobs in all queues

**squeue -p** broadwl -- will list all jobs in just the broadwl partition

**squeue -u $USER** -- will list only your jobs

The output of squeue lists jobs in rows. The columns are described below:

**Name Description**

--------------------------------------------------------------------------------------------------

JOBID Job ID # Unique reference number for each job

PARTITION Type of node job is running/will run on

NAME Name for the job, defaults to slurm-JobID

USER User who submitted job

ST State of the job – CF (configuring),PD (pending),R (running)

TIME Time used by the job in D-HH:MM:SS

NODES Number of Nodes utilized

NODELIST(REASON) List of Nodes in use, or reason the job has not started

To cancel a queued or running job you can use the scancel command.

**scancel** <jobid> where <jobid> is the id of the job you wish to cancel.

To get the <jobid> you need to use the squeue command.