Session 1: Everything is Somewhere

Navigating the directory structure
Moving files around

Session 2: The computer as a nut - using Unix commands

Summary
Unix cheat sheet

Everything is *somewhere*, and everything happens somewhere. When you are using a Unix system (including Linux systems and HPC clusters), there are two kinds of somewhere:

1. What *host* (computer) am I on?
2. What filesystem and directory am I in?
The host determines what CPU and memory resources you have available, with whom you are sharing them and which filesystems (and therefore files and programs) are accessible.

The filesystem and directory (location within the filesystem) determine what files and programs are immediately visible, to you and to any program you run.

In fact, the same is true of a typical workstation, but on a workstation you have one host and one filesystem, shared with nobody, and the details of your location within the filesystem are hidden by the Windows or OSX GUI.

The diagram below is a simple map of some key places you move between when using an NYU HPC system (Tip: click the image for a closer view):

- From your workstation, you can access the Internet, and through it, the bastion host gw.hpc.nyu.edu. The term bastion host refers to a server which stands between a private network - in this case the NYU HPC network - and a public network - the Internet. It protects the private network from attack from the outside. As HPC users, the significance of the bastion host is that we cannot directly access the HPC clusters.

**Exercise**
Try it. Open a terminal on your workstation and attempt to log in to bastion host with:

```
$ ssh NetID@gw.hpc.nyu.edu
```

If you are using Windows, you'll do this via PuTTY instead of a terminal. Click here to expand the instructions for PuTTY
Pri nce (HP C)
Du mbo (Ha do op)
Bro oklyn (Op en Sta ck)
Dal ma (NY U Ab u Dh abi)

Transfe rring data to/from the clusters
Transfe rring data to/from Prince cluster using Globus
Submit ting jobs with sbatch
Availabl e softwar e
Running jobs on the Prince Cluster
Accessing the Prince Cluster

From Windows workstation

From Mac workstation

Software and Environment Module

Job script and resource request

Introduction to job scheduling

Submitting jobs with sbatch

Requesting resources

Using computing nodes interactively

Monitoring batch jobs

Monitoring batch jobs - queue
To log in you will need an SSH client. We recommend PuTTY SSH but others are also available. After using PuTTY to get to `gw.hpc.nyu.edu` you will use the program `ssh` from a terminal window. If you are unfamiliar with the command line interface this may seem daunting - relax, it's easy and vastly more powerful than point-and-click. We have a basic tutorial here.

If you wish to use any software with a graphical interface, your Windows workstation must have an X server installed and running. This is described further in Preparing your Windows workstation below.

For easier access and to transfer files between your workstation and the clusters, you will eventually want to set up SSH tunneling.

Preparing your Windows workstation for X

If you wish to use any software with a graphical interface, you will need an X server. This is a software package that draws on your local screen windows created on a remote computer (such as an NYU HPC cluster). There are a couple of options out there:

- One option is Xming. Installation instructions can be found on its web site. To use with PuTTY, we recommend Xming. Before starting PuTTY you will need to launch Xming by double-clicking the Xming icon on your desktop or start it from the Start Menu.

- Another option is Cygwin/X. Instructions for downloading and installing it can be found here. Before starting PuTTY you will need to have the X server running, by double-clicking the "XWin Server" shortcut under Cygwin-X on the Start Menu. You may wish to add this to your Windows Startup folder so it runs automatically after starting Windows.
  
  Note: There are complications since X server 1.17. See this Cygwin/X FAQ, and the superuser page for how to populate the PuTTY field 'X authority file for local display' (somewhere in the middle to bottom of the page).

You will also need to download and install PuTTY SSH if you have not already.

Logging In - the easy, primitive way

This method is useful as a first-time login, to ensure you are able to connect. For ongoing use however, we recommend setting up and using SSH tunneling instead.
When you installed PuTTY you will have added to your Desktop and also the the Start Menu and icon labeled “putty”. Double-clicking this icon will launch PuTTY, with the dialog initially in the “Session” category as in the screenshot below:

If this is not your first time through here, and you already saved a setup previously, you can select the saved session and hit “Load”. This will do the following steps for you so you can just hit “Open”. Otherwise, read on...

Enter the as Host Name "gw.hpc.nyu.edu", leave the Port at its default value of 22 and select the “SSH” Protocol. **Don’t** hit “Open” just yet!

Eventually, if not now, you will probably want to use some graphical applications. For the procedure described here, you don’t need to check the “Enable X11 forwarding”. You may disable it by un-selecting Connection -> SSH -> X11 from the Category menu on the left:

Again, **don’t** hit “Open” just yet...

You can save these settings to use again next time. Return to the “Session” category and give this setup a name in the “Saved Sessions” box (perhaps NYU HPC”). Now click “Save”.

Finally, hit “Open”. You will get a command terminal for gw.hpc.nyu.edu. This is a deliberately restricted environment from which you can do little other than to log in to one of the HPC clusters.

Users do not automatically get accounts on NYU HPC Prince cluster, and dumbo Hadoop cluster. For information on how to get an account, please visit [Getting or renewing an HPC account](#) page.

After you click “Open”, you will get a terminal window where you will be prompted to login using your NYU NetID and password. The window will display the following:
Now, you are on the NYU Network.

Therefore, log into a cluster using one of the following commands:

```
$ ssh NetID@prince.hpc.nyu.edu
$ ssh NetID@dumbo.hpc.nyu.edu
```

Suppose, I wish to login into the prince cluster, the window will display the following:

```
ssb536@hpc-bastion1->$ ssh ssb536@prince.hpc.nyu.edu
ssb536@prince.hpc.nyu.edu's password:
Last login: Wed Aug 30 17:10:28 2017 from psd01a-0046s.cfs.its.nyu.edu
[ssb536@log-1 ~]$  
```

In the boxes below and elsewhere in this wiki, the symbol "$" at the beginning of a line represents the command prompt - don't type the "$", type only the remainder of the line following the "$". Also wherever NetID appears, replace it with your NYU NetID.

On Prince you may notice that you are now on a host named "log-0" or "log-1" or something similar. These clusters use multiple login nodes and which one you get depends how busy each is at the time. The login nodes are configured identically and see the same filesystems, so the specific node you are logged in to is not important.

Setting up SSH Tunneling
Step 1: Creating the tunnel

1. First open PuTTY and prepare to log in to gw.hpc.nyu.edu, as you did in Logging in from Windows - primitive way. If you saved your session during that process, you can load it by selecting from the "Saved Sessions" box and hitting "Load". Don't hit "Open" yet!
2. Under "Connection" -> "SSH", just below "X11", select "Tunnels"

3. Write "8026" (the port number) in the "Source port" box, and "prince.hpc.nyu.edu:22" (the machine you wish to tunnel to - 22 is the port that ssh listens on) in the "Destination" box.
4. Click "Add". You can repeat step 3 with a different port number and a different destination. If you do this you will create multiple tunnels, one to each destination.
5. Before hitting "Open", go back to the "Sessions" page, give the session a name ("hpcgw_tunnel") and hit "Save". Then next time you need not do all this again, just load the saved session.
6. Hit "Open" to login in to gw.hpc.nyu.edu and create the tunnel. A terminal window will appear, asking for your login name (NYU NetID) and password (NYU password). Windows may also ask you to allow certain connections through its firewall - this is so you can ssh to port 8026 on your workstation - the entrance to the tunnel.

Using your SSH tunnel

To log in via the tunnel, first the tunnel must be open. If you've just completed Step 1, it will be open and you can jump down to "Step 2: Logging in via your SSH tunnel". If you completed Step 1 yesterday, and now want to re-use the tunnel you created, first start the tunnel:

You can add other NYU hosts to the tunnel by adding a new source port and destination and clicking "Add". For example, you could add "Source port = 8025" and "Destination = dumbo.hpc.nyu.edu:22", then press "Add". You would then perform Step 2 (below) twice - once for prince on port 8026 and once for dumbo on port 8025.
Starting the tunnel

During a session, you need only do this once - as long as the tunnel is open, new connections will go over it.

1. Start Putty.exe (again, if necessary), and load the session you saved in Setting up SSH Tunneling
2. Hit “Open”, and log in to the bastion host with your NYU NetID and password. This will create the tunnel.

Step 2: Logging in via your SSH tunnel

1. Start the second Putty.exe. In the “Host Name” box, write “localhost” and in the “Port” box, write “8026” (or whichever port number you specified when you set up the tunnel in Setting up SSH Tunneling)
   We use “localhost” because the entrance of the tunnel is actually on this workstation, at port 8026.
2. Go to "Connections" -> "SSH" -> "X11" and check "Enable X11 forwarding".

3. Optionally, give this session a name (in "Saved Sessions") and hit "Save" to save it. Then next time instead of steps 1 and 2 you can simply load this saved session.

4. Hit "Open". You will again get a terminal window asking for your login (NYU NetID) and password (NYU password). You are now logged in to the HPC cluster!

- From gw.hpc.nyu.edu you can access the Prince login nodes.

**Exercise**

Log in again, this time via prince.hpc.nyu.edu:

```
$ ssh prince.hpc.nyu.edu
```

Again, Windows users will use PuTTY for the first step and for the second, ssh from the terminal PuTTY will start. Here's that Wiki page again...
Dalma (NYU Abu Dhabi)

Transferring data to/from the clusters

Transferring data to/from Prince cluster using Globus

Submitt ing jobs with sbatch

Available software

Licensed Software Available on the HPC Cluster

Building Software

Slurm Tutorial

Tutorials

FAQs
Running jobs on the Prince Cluster

Accessing the Prince Cluster

From Windows workstation

From Mac workstation
<table>
<thead>
<tr>
<th>Software and Environment Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job script and resource request</td>
</tr>
<tr>
<td>Introducing job scheduling</td>
</tr>
<tr>
<td>Submitting jobs with sbatch</td>
</tr>
<tr>
<td>Requesting resources</td>
</tr>
<tr>
<td>Using computing nodes interactively</td>
</tr>
<tr>
<td>Monitoring batch jobs</td>
</tr>
<tr>
<td>Monitoring batch jobs - queue</td>
</tr>
<tr>
<td>What is running and where? slurmtop</td>
</tr>
<tr>
<td>Canceling your jobs</td>
</tr>
<tr>
<td>Compiling your own software</td>
</tr>
<tr>
<td>Putting all pieces together</td>
</tr>
</tbody>
</table>
To log in you will need an SSH client. We recommend PuTTY SSH but others are also available. After using PuTTY to get to gw.hpc.nyu.edu you will use the program ssh from a terminal window. If you are unfamiliar with the command line interface this may seem daunting - relax, it's easy and vastly more powerful than point-and-click. We have a basic tutorial here.

If you wish to use any software with a graphical interface, your Windows workstation must have an X server installed and running. This is described further in Preparing your Windows workstation below.

For easier access and to transfer files between your workstation and the clusters, you will eventually want to set up SSH tunneling.

Preparing your Windows workstation for X

If you wish to use any software with a graphical interface, you will need an X server. This is a software package that draws on your local screen windows created on a remote computer (such as an NYU HPC cluster). There are a couple of options out there:

- One option is Xming. Installation instructions can be found on its web site. To use with PuTTY, we recommend Xming. Before starting PuTTY you will need to launch Xming by double-clicking the Xming icon on your desktop or start it from the Start Menu.

- Another option is Cygwin/X. Instructions for downloading and installing it can be found here. Before starting PuTTY you will need to have the X server running, by double-clicking the "XWin Server" shortcut under Cygwin-X on the Start Menu. You may wish to add this to your Windows Startup folder so it runs automatically after starting Windows.

  Note: There are complications since X server 1.17. See this Cygwin/X FAQ and the superuser page for how to populate the PuTTY field 'X authority file for local display' (somewhere in the middle to bottom of the page).

You will also need to download and install PuTTY SSH if you have not already.

Logging In - the easy, primitive way

This method is useful as a first-time login, to ensure you are able to connect. For ongoing use however, we recommend setting up and using SSH tunneling instead.

When you installed PuTTY you will have added to your Desktop and also the the Start Menu and icon labeled “putty”. Double-clicking this icon will launch PuTTY, with the dialog initially in the "Session" category as in the screenshot below:

If this is not your first time through here, and you already saved a setup previously, you can select the saved session and hit "Load". This will do the following steps for you so you can just hit "Open". Otherwise, read on...

Enter the as Host Name "gw.hpc.nyu.edu", leave the Port at its default value of 22 and select the "SSH" Protocol. Don't hit "Open" just yet!

Eventually, if not now, you will probably want to use some graphical applications. For the procedure described here, you don't need to check the "Enable X11 forwarding". You may disable it by un-selecting Connection -> SSH -> X11 from the Category menu on the left:
Again, **don’t** hit “Open” just yet...

You can save these settings to use again next time. Return to the “Session” category and give this setup a name in the “Saved Sessions” box (perhaps NYU HPC”). Now click “Save”.

Finally, hit “Open”. You will get a command terminal for gw.hpc.nyu.edu. This is a deliberately restricted environment from which you can do little other than to log in to one of the HPC clusters.

```
Users do not automatically get accounts on NYU HPC Prince cluster, and dumbo Hadoop cluster. For information on how to get an account, please visit Getting or renewing an HPC account page.
```

After you click “Open”, you will get a terminal window where you will be prompted to login using your NYU NetID and password. The window will display the following:

```
login as: ssb536
** NOTICE: NYU Authorized Use Only **

Access and use, or causing access and use, of this computer system by anyone other than as permitted by New York University (NYU) is strictly prohibited by NYU and by law. Such use might subject an unauthorized user, including unauthorized employees, to criminal and civil penalties as well as NYU-initiated disciplinary proceedings. The use of this system is routinely monitored and recorded, and anyone accessing this system consents to such monitoring and recording.
ssb536@gw.hpc.nyu.edu’s password:
Last login: Sun Sep 17 20:03:50 2017 from 24.193.233.128
```

Now, you are on the NYU Network.

Therefore, log into a cluster using one of the following commands:
Suppose, I wish to login into the prince cluster, the window will display the following:

```
$ ssh NetID@prince.hpc.nyu.edu
$ ssh NetID@dumbo.hpc.nyu.edu
```

On Prince you may notice that you are now on a host named "log-0" or "log-1" or something similar. These clusters use multiple login nodes and which one you get depends how busy each is at the time. The login nodes are configured identically and see the same filesystems, so the specific node you are logged in to is not important.

**Setting up SSH Tunneling**
Step 1: Creating the tunnel

1. First open PuTTY and prepare to log in to gw.hpc.nyu.edu, as you did in Logging in from Windows - primitive way. If you saved your session during that process, you can load it by selecting from the "Saved Sessions" box and hitting "Load". Don't hit "Open" yet!
2. Under "Connection" -> "SSH", just below "X11", select "Tunnels".

3. Write "8026" (the port number) in the "Source port" box, and "prince.hpc.nyu.edu:22" (the machine you wish to tunnel to - 22 is the port that ssh listens on) in the "Destination" box.
4. Click "Add". You can repeat step 3 with a different port number and a different destination. If you do this you will create multiple tunnels, one to each destination.
5. Before hitting "Open", go back to the "Sessions" page, give the session a name ("hpcgw_tunnel") and hit "Save". Then next time you need not do all this again, just load the saved session.
6. Hit "Open" to login in to gw.hpc.nyu.edu and create the tunnel. A terminal window will appear, asking for your login name (NYU NetID) and password (NYU password). Windows may also ask you to allow certain connections through its firewall - this is so you can ssh to port 8026 on your workstation - the entrance to the tunnel.

Using your SSH tunnel

To log in via the tunnel, first the tunnel must be open. If you've just completed Step 1, it will be open and you can jump down to "Step 2: Logging in via your SSH tunnel". If you completed Step 1 yesterday, and now want to re-use the tunnel you created, first start the tunnel:

You can add other NYU hosts to the tunnel by adding a new source port and destination and clicking "Add". For example, you could add "Source port = 8025" and "Destination = dumbo.hpc.nyu.edu:22", then press "Add". You would then perform Step 2 (below) twice - once for prince on port 8026 and once for dumbo on port 8025.
Starting the tunnel

During a session, **you need only do this once** - as long as the tunnel is open, new connections will go over it.

1. Start Putty.exe (again, if necessary), and load the session you saved in **Setting up SSH Tunneling**
2. Hit "Open", and log in to the bastion host with your NYU NetID and password. This will create the tunnel.

Step 2: Logging in via your SSH tunnel

1. Start the second Putty.exe. In the "Host Name" box, write "localhost" and in the "Port" box, write "8026" (or whichever port number you specified when you set up the tunnel in **Setting up SSH Tunneling**)
   
   We use "localhost" because the entrance of the tunnel is actually on this workstation, at port 8026.
Prince has a number of filesystems mounted (attached and visible). The diagram above only shows a few important ones.

Some things to note are:

- `/home`, `/scratch` and `/beegfs` are connected to the login nodes and also to the compute nodes
- `/archive` is connected to the login nodes, but not to the compute nodes
  So processes running on the compute nodes cannot see `/archive`
- The connection from `/scratch` to the compute nodes is fatter - indicating faster - than the other connections on the system

On NYU HPC clusters, `/scratch` is configured to provide fast I/O for jobs running on the compute nodes

2. Go to "Connections" -> "SSH" -> "X11" and check "Enable X11 forwarding".

3. Optionally, give this session a name (in "Saved Sessions") and hit "Save" to save it. Then next time instead of steps 1 and 2 you can simply load this saved session.

4. Hit "Open". You will again get a terminal window asking for your login (NYU NetID) and password (NYU password). You are now logged in to the HPC cluster!

You can see which host you are on with the hostname command:

```
$ hostname
log-0
```

Prince currently has two login nodes, so the hostname you see may not be the same as the one above. The login nodes are configured to be equivalent so for most purposes it makes no difference which specific login node you land on.
- /home and /scratch are separate filesystems, in separate places.
  Moving files around within /home or within /scratch is trivial and fast, but to move files between /home and /scratch, the file must travel from one filesystem to the other over a network. Expect this to take longer!

Finally

We go to a host with ssh.. and to leave a host, type "exit".

(actually, "exit" leaves the shell - we'll discuss shells in session 2 - but by leaving the shell that ssh started upon your arrival, you will also leave the host)

Next: Navigating the directory structure