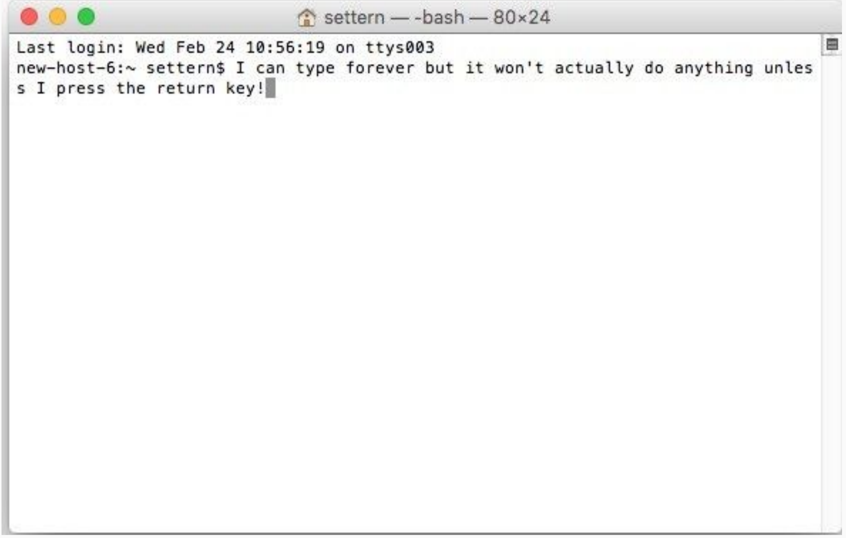
BF528 - Applications in Translational Bioinformatics

Computational Skills Primer

Computer was born in the mind of man, **not the other way around**!!



```
Last login: Wed Feb 24 10:56:19 on ttys003
new-host-6:~ settern$ I can type forever but it won't actually do anything unles
s I press the return key!
```

Goal of this lecture:

- Overcome the fear of black screen (if you have one !!)
- Learn techniques for working on SCC you will need
- Introduce shared computing concepts you will use

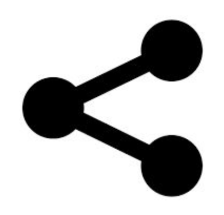
Prerequisites

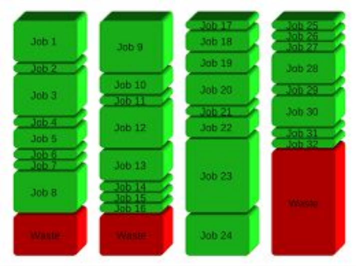
- Patience with self and with each other
- Keep an open mind
- Attitude of collaboration
- It’s OK to not know - we will learn together
- **Rome ne s'est pas faite en un jour** !!!

What is SCC ?

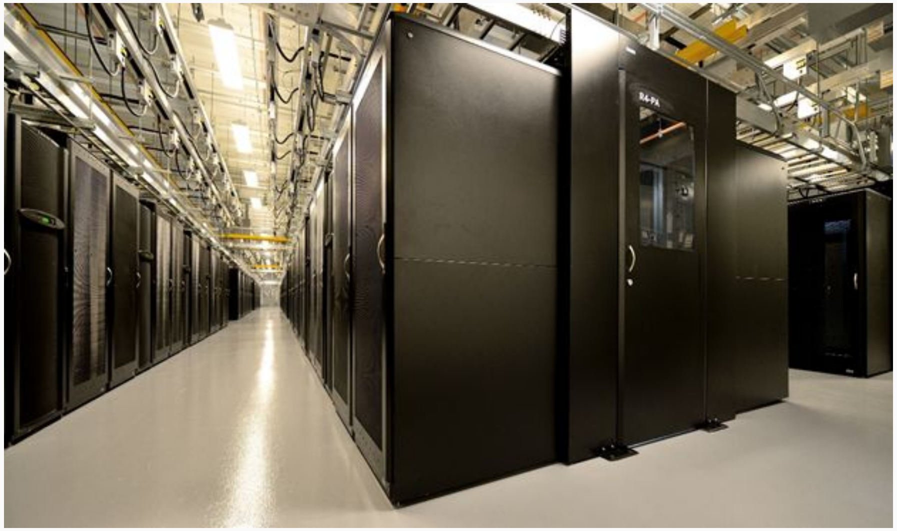
Shared Computing Cluster (SCC)

- **Shared:** Multi-user environment.
- **Computing:** Calculation, file manipulation, etc
- **Cluster:** Many individual computers connected together working in coordination





Massachusetts Green High Performance Computing Center



BOSTON UNIVERSITY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HARVARD

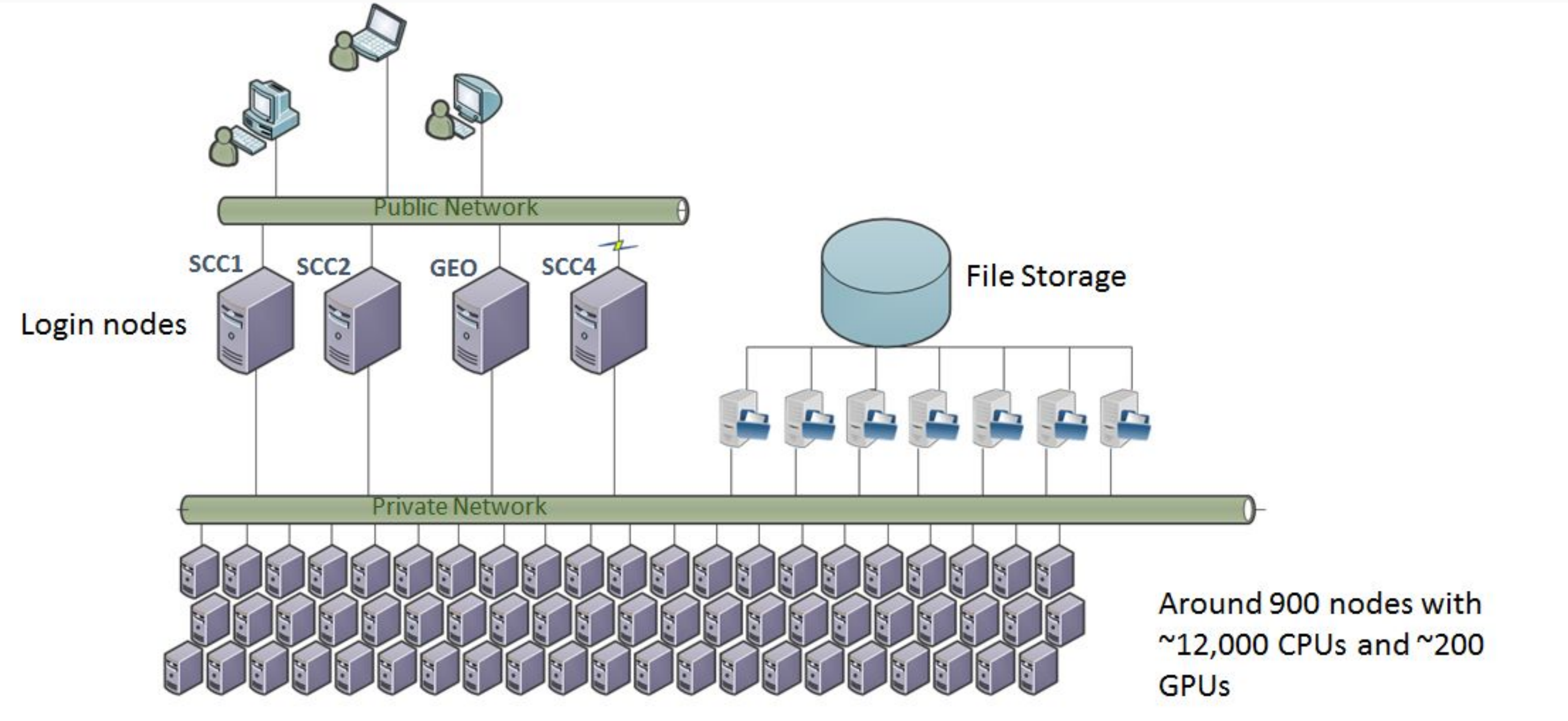
NORTHEASTERN UNIVERSITY

UMASS BOSTON

Terminology

- **Cluster** - connected set of computers
- **Node** - a single computer (on a rack)
  - **Head/login node** - “gateway” computer to cluster
  - **Compute node** - single computer where jobs run
- **terminal** - program used to interact with a computer using text commands
- **ssh** - secure shell, how we connect to cluster
  - NB: both a noun and a verb, e.g. “use ssh to ssh into scc1…”

SCC Architecture



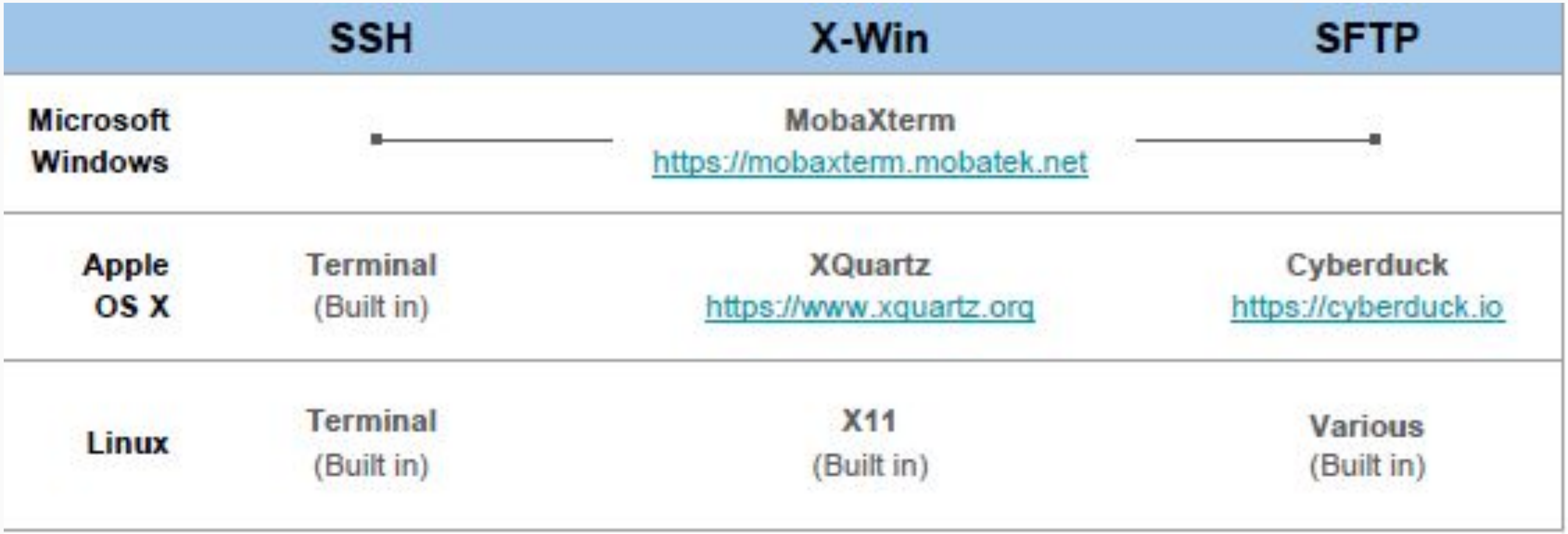
Around 900 nodes with ~12,000 CPUs and ~200 GPUs

Why use SCC? I have a laptop!

- Collaborate on projects
- Run code exceeding workstation capability
- Secured Network
- Fast and easy data sharing
- Analyze protected data (e.g. dbGaP)
- Run code that runs for long periods of time
- Perform many computations in parallel

Software to Access SCC

- **SSH**

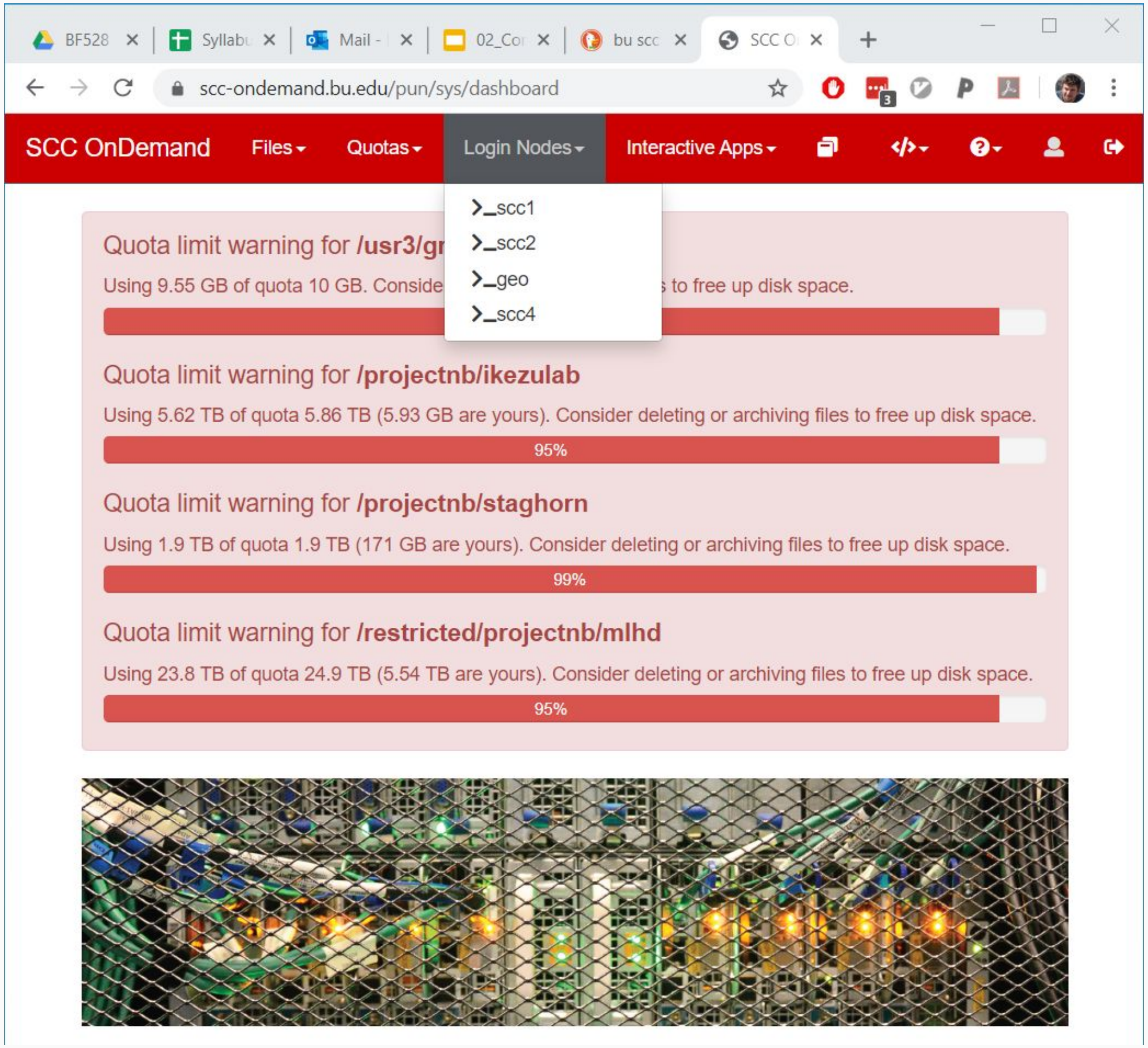


	SSH	X-Win	SFTP
Microsoft Windows	MobaXterm https://mobaxterm.mobatek.net		
Apple OS X	Terminal (Built in)	XQuartz https://www.xquartz.org	Cyberduck https://cyberduck.io
Linux	Terminal (Built in)	X11 (Built in)	Various (Built in)

- **SCC OnDemand**

https://scc-ondemand.bu.edu

How to connect with SCC OnDemand



Quota limit warning for **/usr3/gr**

Using 9.55 GB of quota 10 GB. Conside … s to free up disk space.

Quota limit warning for **/projectnb/ikezulab**

Using 5.62 TB of quota 5.86 TB (5.93 GB are yours). Consider deleting or archiving files to free up disk space.

95%

Quota limit warning for **/projectnb/staghorn**

Using 1.9 TB of quota 1.9 TB (171 GB are yours). Consider deleting or archiving files to free up disk space.

99%

Quota limit warning for **/restricted/projectnb/mlhd**

Using 23.8 TB of quota 24.9 TB (5.54 TB are yours). Consider deleting or archiving files to free up disk space.

95%

SCC OnDemand Demonstration: Connecting

Navigation

Essential navigation commands:

- **pwd** - print current working directory
- **ls** - list files
- **cd** - change directory

The “address” of a file/directory in linux is called a **path**:

- *Absolute path* - starts with /
- *Relative path* - does not start with /

List of useful commands - Part I

Useful options for the “**ls**” command:

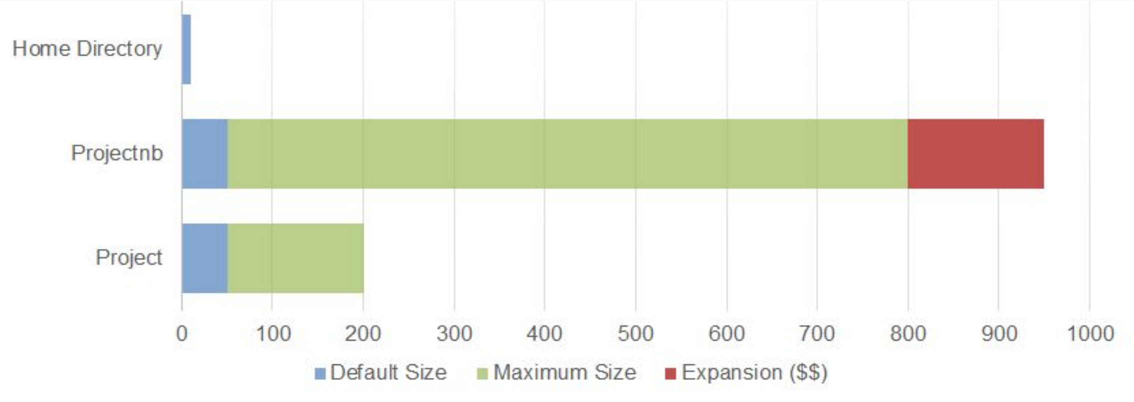
ls -a	List all files, including hidden files beginning with a period “**.**”
ls -ld *	List details about a directory and not its contents
ls -F	Put an indicator character at the end of each name
ls –l	Simple long listing
ls –lR	Recursive long listing
ls –lh	Give human readable file sizes
ls –lS	Sort files by file size
ls –lt	Sort files by modification time (very useful!)

Storage (GB)

- **~/** - Home directory, your personal space (10 GB max)
- **/project** – Source code, files you can't replace.
- **/projectnb** – Output files, downloaded data sets
- Available from all head nodes (scc1, scc2, etc)

Restricted data (protected data, e.g. dbGaP)

- **/restricted/project**
- **/restricted/projectnb**
- Only accessible through scc4.bu.edu and compute nodes.



Permissions

Files Access Control:

- Every file has an owner
- Every file belongs to a group
- Every file has "permissions" controlling access to it:
    - **r** - read
    - **w** - write
    - **x** - execute

```
[kkarri@scc4 ~]$ ls -l newdir
drwxr-xr-x 3 kkarri waxmanlab 512 Jan 21 16:03 newdir
```

chmod

Change the permissions on the directory “newdir” so that members of your group can write to it:

```
[kkarri@scc4 ~]$ ls -l newdir
drwxr-xr-- 3 kkarri waxmanlab 512 Jan 21 16:03 newdir
```

Members of my group don’t have access to this file, change that:

```
[kkarri@scc4 ~]$ chmod g+w newdir
[kkarri@scc4 ~]$ ls -l newdir
drwxrwxr-- 3 kkarri waxmanlab 512 Jan 21 16:03 newdir
```

SCC OnDemand Demonstration: Navigation

List of useful commands

cp [file1] [file2]	copy file
mkdir [name]	make directory
rmdir [name]	remove (empty) directory
mv [file] [destination]	move/rename file
rm [file]	remove (-r for recursive)
file [file]	identify file type
less [file]	page through file
head -n [file]	display first n lines
tail -n [file]	display last n lines
ln –s [file] [new]	create symbolic link
cat [file] [file2…]	display file(s)
tac [file] [file2…]	display file in reverse order

Additional Reading

Assignment: familiarize yourself with the material on basic command line usage found here:

Workshop 0. Basic Linux and Command Line Usage

For in-depth understanding of these concepts go through the following modules on cluster computing and advance command line text editors:

- Cluster Computing Tutorial
- Advanced CLI Tutorial