

Intro to High-Performance Computing and the Smithsonian Hydra Cluster

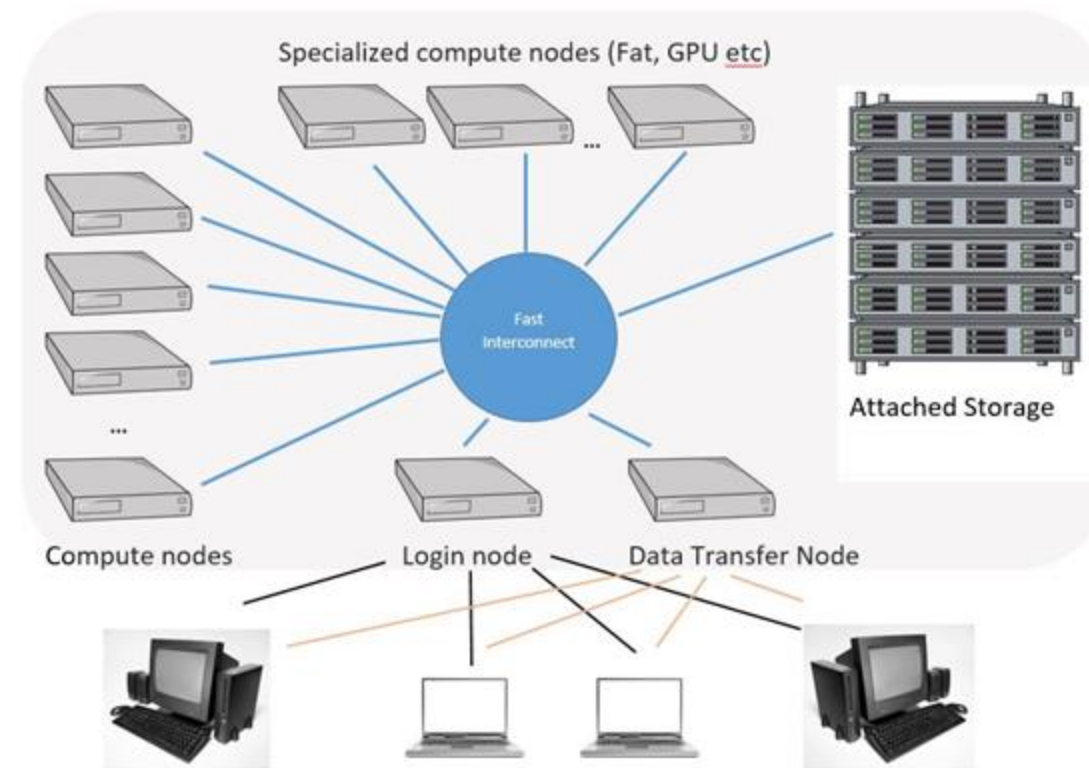
Dan MacGuigan
June 13, 2025

Hydra vulgaris, (c) Proyecto Agua
<https://www.inaturalist.org/taxa/486293-Hydra-vulgaris>



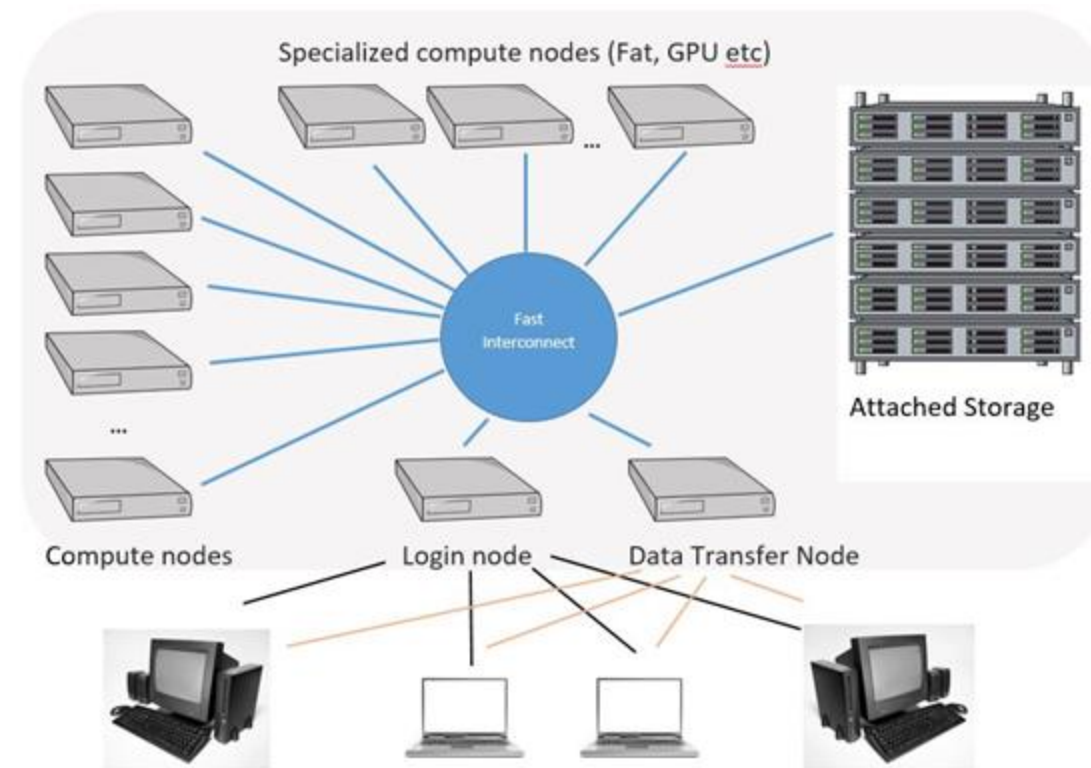
What is a High-Performance Computing (HPC) cluster?

- Basically, just many individual computers linked together
- **Nodes** \sim individual computers
 - **Login node** = the central node that all users interact with when they log in
 - **Compute nodes** = nodes that run **jobs** (programs, code, scripts, etc) for users
- Each node has its own CPU with some number of cores and RAM



Why use a HPC cluster?

- Access to way more computing power than any individual lab could afford
- Run analyses remotely so they won't bog down your own computer
- Each user can run hundreds of computing jobs in parallel
- A safe environment to learn
 - Each user only has read/write access to their own files
 - Can't accidentally delete someone's data
- Hardware is maintained by professionals
- Tech support is available



The Hydra cluster

- Accessible to all Smithsonian researchers
- Must have an SI account and be on the “si-staff” WiFi or an SI ethernet connection

Hardware

- ~70 compute nodes
- ~6,000 CPUs
- 8 GPUs
- 45 TB RAM

#	Cores/node	Mem/node	Model	Name	Note
1	72	512GB	R750XA	compute-50-??	4x L40s GPUs
16	40	384GB	R640	compute-64-??	
2	32	512GB	R640	compute-64-??	
29	64	512GB	R6515	compute-65-??	
5	128	756GB	R7525	compute-75-??	
2	128	1,024GB	R7525	compute-75-??	
2	192	1,536GB	R7625	compute-76-??	
12	128	1,024GV	R7625	compute-76-??	
2	20	128GB	R790	compute-79-??	2x GV100GL GPUs
1	112	896GB	R840	compute-84-??	
1	64	512GB	R930	compute-93-??	
3	72	760GB	R930	compute-93-??	
1	96	2,048GB	R930	compute-93-??	

<https://confluence.si.edu/display/HPC/Compute+Nodes>

Hydra resources

- [Hydra Wiki](#)
 - [Login to Hydra](#)
 - [Transferring data to/from Hydra](#)
 - [Submitting a job](#)
 - [Hydra hardware](#)
 - [Hydra software](#)
- [Hydra Qsub generation tool](#)
- [Hydra status](#)

Hydra file system

Location	Description	User storage space	Backed up?	Old files deleted?
/home/USER_ID	Your home directory, for your basic configuration files, scripts and job files	512 Gb	Yes, every 4 weeks	No
/pool/public/genomics/USER_ID	For storing large files, fast disk R/W speed, good for analyses	7.5 Tb	No	Yes, after 180 days
/scratch/public/genomics/USER_ID	For storing large files, fastest disk R/W speed, best for analyses	15 Tb	No	Yes, after 180 days

Basic Linux commands

command	description
pwd	Gives the full file path to your current directory.
ls	List the files and directories in my current location
ls -l	Nicer list format
ls -la	Nicer list format, show hidden files
ls -lah	Nicer list format, show hidden files, human-readable file sizes
mkdir myNewDirectory	Make a new directory. Important to avoid spaces or special characters.
cd myNewDirectory	Move into a directory
cd ..	Move up one directory
cd /full/path/to/myNewDirectory	Move to a directory using the full file path
cd ~	Move to your home directory
touch myNewFile.txt	Create a new, empty file
nano myNewFile.txt	Modify a file from the command line using the “nano” text editor
cat my_new_file.txt	Print the contents of a file to the screen
clear	Clean up your terminal window

USE WITH CAUTION

command	description
cp myNewFile.txt fileCopy.txt	Copy a file
cp -rf myNewDirectory directoryCopy	Copy a directory
mv myNewFile.txt newName.txt	Rename a file
rm myNewFileCopy.txt	PERMANENTLY delete a file
rm -r myNewDirectory	PERMANENTLY delete a directory

Many more commands:
<https://phoenixnap.com/kb/linux-commands-cheat-sheet>

Tips for navigating on the command line

Shortcut	Description
<code>~</code>	Your home directory
<code>.</code>	Your present directory
<code>..</code>	One directory above your current directory
<code>ctrl + arrow key (PC)</code> <code>opt + arrow key (Mac)</code>	Move the cursor one word left or right
<code>fn + arrow key left (PC)</code> <code>ctrl + a (Mac)</code>	Move the cursor to the beginning of the line.
<code>fn + arrow key right (PC)</code> <code>ctrl + e (Mac)</code>	Move the cursor to the end of the line.
<code>control + c</code>	Cancel whatever is currently running on the command line
<code>tab</code>	Autocomplete the path or file name
<code>up and down arrows</code>	Scroll through your previous commands
<code>exit</code>	Close your connection to the cluster