Parallel Machine Learning and Artificial Intelligence

Dr. Handan Liu

h.liu@northeastern.edu

Northeastern University

Content

- Connecting to Discovery
- Cluster File System
- Partitions and Compute Nodes
- Data transfer
- Loading Module
- Using Slurm
- Running jobs: interactive mode and batch mode; Job scripts
- Linux Fundamentals for Discovery Cluster
- Learn how to compile and run OpenMP and MPI programs via interactive mode and batch mode

Discovery and MGHPCC

- Discovery is a high-performance computing (HPC) resource for the Northeastern University research community. The Discovery cluster is located in MGHPCC.
- The Massachusetts Green High Performance Computing Center in Holyoke, MA houses computing resources for five institutions: Northeastern, BU, Harvard, MIT, and UMass.

http://www.mghpcc.org

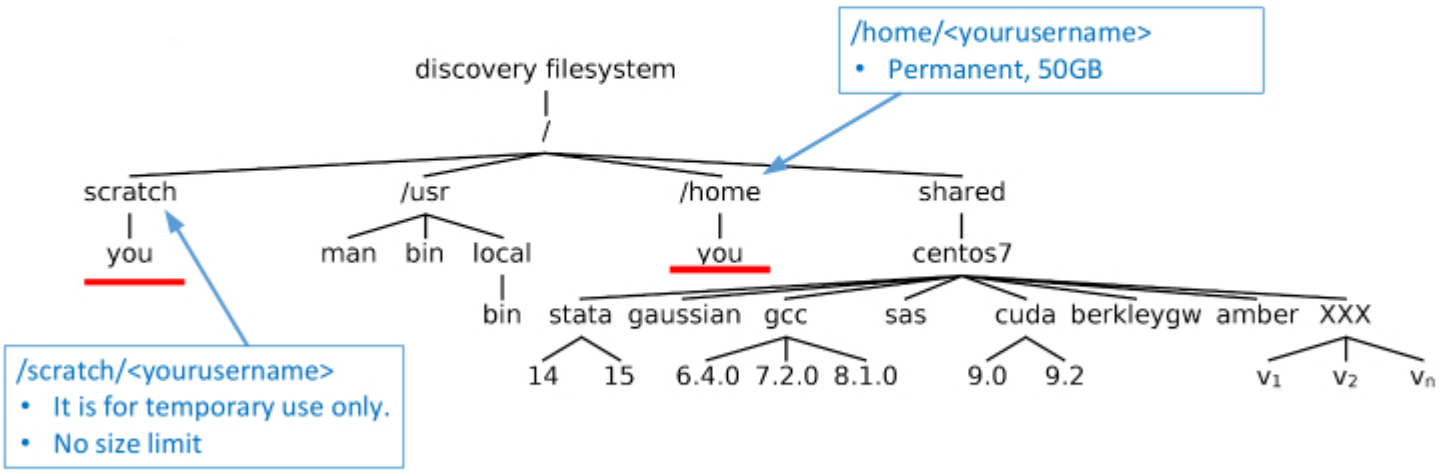
Connecting to Discovery

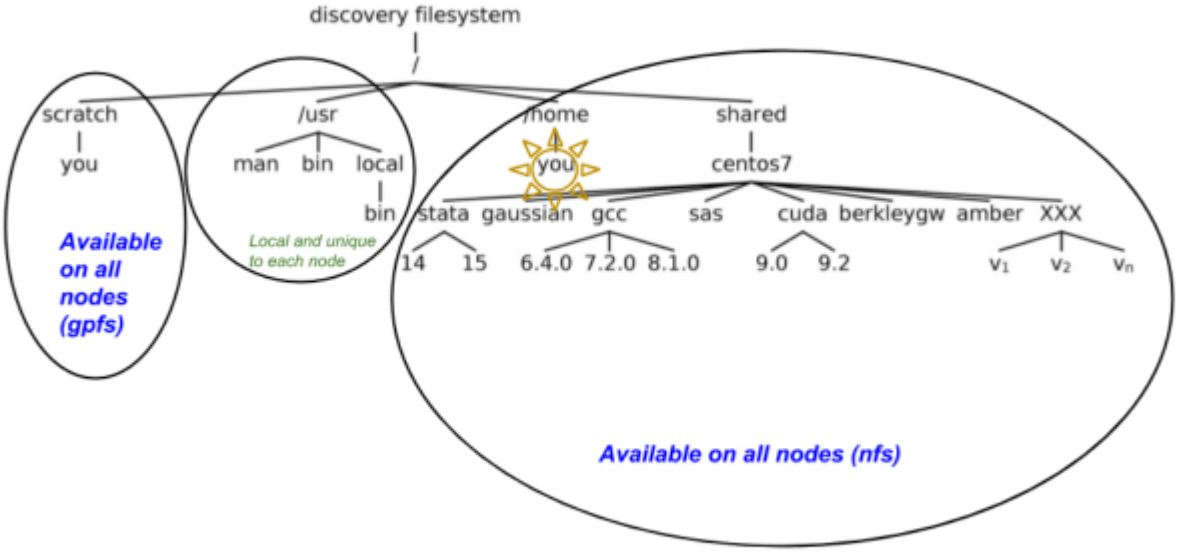
- Secure Shell (SSH): https://www.ssh.com/ssh/protocol/
  - Linux/Mac: Terminal. For GUI, use x11 forwarding.
    - ✔ Mac: use XQuartz https://www.xquartz.org/
  - Windows: Putty (Win&Mac), MobaXTerm (recommend)
    - ✔ Putty: https://www.puttygen.com/download-putty
    - ✔ MobaXTerm: https://mobaxterm.mobatek.net/download.html

$ ssh -Y <username>@login.discovery.neu.edu

- Here, username is your Northeastern username
- -Y is used for x11 forwarding (GUI)
- ==> [username@login-00 ~] [username@login-01 ~]

Discovery File System





Partitions – CPU nodes

Name	Requires approval ?	Time limit (default/max)	Running jobs	Submitted jobs	Core limit	RAM limit
debug	No	20 minutes/20 minutes	10/25	5000	128	256GB
express	No	30 minutes/60 minutes	50/250	5000	2048	25TB
short	No	4 hours/24 Hours	50/500	5000	1024	25TB
long	**Yes**	1 day/5 Days	25/250	1000 per user/5000 per group	1024	25TB
large	**Yes**	6 hours/6 Hours	100/100	1000 per user/5000 per group	N/A	N/A

Partitions – GPU nodes

Name	Requires approval?	Time limit (default/max)	Running jobs	Submitted jobs	GPU per job limit	GPU per user limit
gpu	No	4 hours/8 Hours	25/250	50/100	1	8
multigpu	**Yes**	4 hours/24 Hours	25/100	50/100	12	12

View partitions:

- $ sinfo -a
- $ sinfo --help
- $ sinfo -p express
- $ scontrol show partition express

Compute Nodes: CPU nodes

CPU Type	Cores per Node	Number of Nodes	Total Cores	RAM per node
E5-2680v2@2.8 GHz	20	76	1520	64GB
E5-2690v3@2.6 GHz	24	184	4416	128GB
E5-2680v4@2.4 GHz	28	408	11424	256GB
Platinum 8276@2.2 GHz	56	128	7168	192GB

Compute Nodes: GPU nodes

GPU Type	Number of nodes/GPUs	CPU Type	RAM per node
k20m	23 nodes with 1 GPU each	E5-2650@2.00GHz	128GB
k40m	16 nodes with 1 GPU each	E5-2690v3@2.60GHz	128GB
k80	8 nodes with 8 GPUs each	E5-2680v4@2.40GHz	512GB
p100	12 nodes with 4 GPUs each	E5-2680v4@2.40GHz	512GB
v100	4 nodes with 2 GPUs each	AMD EPYC 7351@2.60GHz	480GB
v100-sxm2	24 nodes with 4 GPUs each	Intel Gold 6132@2.60Ghz	187GB

Move to a compute node

- Strict Policy: DON'T RUN JOBS ON LOGIN NODES!

- Move to a compute node:
  - using srun command: `$ srun -p debug --pty /bin/bash`
  - using the sbatch specify the resources: `$ sbatch batch_script_file`

Configuration

On Login node:

- $ sinfo -Nle -p partition_name

On compute node:

- $ lscpu
- $ lsmem
- $ cat /proc/cpuinfo
- $ cat /proc/meminfo

Data Transfer

- Discovery has a dedicated transfer node that you must use to transfer data to and from the cluster.
- You are not allowed to transfer data from any other node.
- The node name is
  - <username>@**xfer.discovery.neu.edu**:
  - where <username> is your Northeastern username.

- Graphically, FileZilla and MobaXTerm

Using Modules

Module Command	Function
module avail	View a list of all of the available software packages on Discovery that you can load
module list	Displays a list of the software packages currently loaded in your path
module show <module name>	View the details of a software package
module load <module name>	Load a software package into your environment
module unload <module name>	Remove a single software package from your environment
module purge	Removes all of the loaded software packages from your environment

- Stay safe!
- See you next class!

Next Lecture will Continue:

Linux Essentials