Getting Started with the Illinois Campus Cluster

Bruno Abreu, Ph.D.
Research Scientist
babreu@illinois.edu

February 20, 2024

National Center for Supercomputing Applications
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Outline

Illinois Campus Cluster Program
AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Block A
~ 20 min

- What is the ICC?
- Who is it for?
- Examples of research applications
- Why should I use it?

Block B
~ 55 min

- How can I access it?
- How can I run my codes and applications?
- What else is available at ICC?
- Support material and further learning

- Q&A session *(~ 15 min)*

What is the ICCP?

https://campuscluster.illinois.edu/

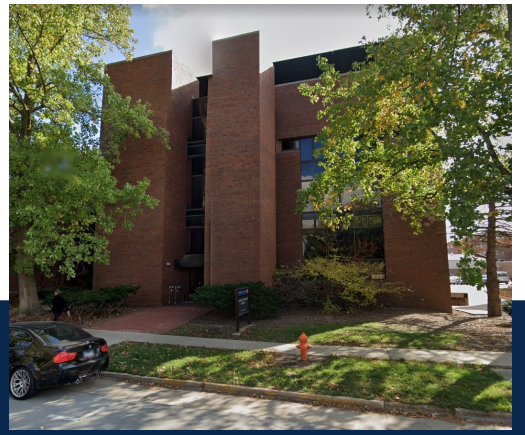
Illinois Campus Cluster Program
AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Technical resource that enables **advanced computing**, **data**, and **networking** for researchers on campus

- Highly trained professional staff from NCSA
- Continuous deployment model: add/retire hardware as needed
- It is on the smaller scale of computation clusters, but you can have a lot of memory
- Located at *1011 W Springfield Avenue* – **Advanced Computation Building**

ILLINOIS
Technology Services

ILLINOIS
NCSA | National Center for Supercomputing Applications



Who is the ICC for? Why ICC?

- Any individual, research team, or campus unit can invest in compute nodes or storage disks or pay a fee for on-demand use
- Are you a researcher that doesn't want to manage your own cluster? Want to move workstation from your office? Here you go!

- Provides access to fast and efficient computing and data storage resources
- Frees you from the hassle of administering your computer cluster/workstation
- Specialized or custom nodes can be prepared with help of highly trained and professional staff
- Centralized resources that are locally accessible
- Physical infrastructure: optimal cooling, power management
- Data security and integrity (backups) services are available
- All equipment and infrastructure costs are approved to be charged to grants
- Focus on your **research** rather than managing your computer hardware

ICC Resources: Hardware

Traditional CPU Nodes

Intel Xeon 6248
- 20 cores / 40 threads processor
- 2.5 GHz base frequency
- 192 to 384 GB of RAM

AMD EPYC 7702
- 64 cores / 128 threads processor
- 2.0 GHz base frequency
- 256 GB to 4 TB GB of RAM

Mixed CPU/GPU Nodes

Traditional CPU Node type capable of accepting 1-4 GPUs

- **NVIDIA GeForce RTX**
- **NVIDIA A6000, A40** and **A100**
- **NVIDIA Tesla T4**

16 to 48 GB GPU Memory
2,560 to 10,752 CUDA cores

Storage

- **DDN SFA 14KX**
420 x 8 TB SAS (2.5 PB usable)
- **Samsung PM1725**
12 x 3.5 TB NVME (19.2 TB usable for metadata)
- 6 x **Dell R740**: I/O Core Servers
- 3 x **Dell R6515**: Export Servers

ICC: node purchase

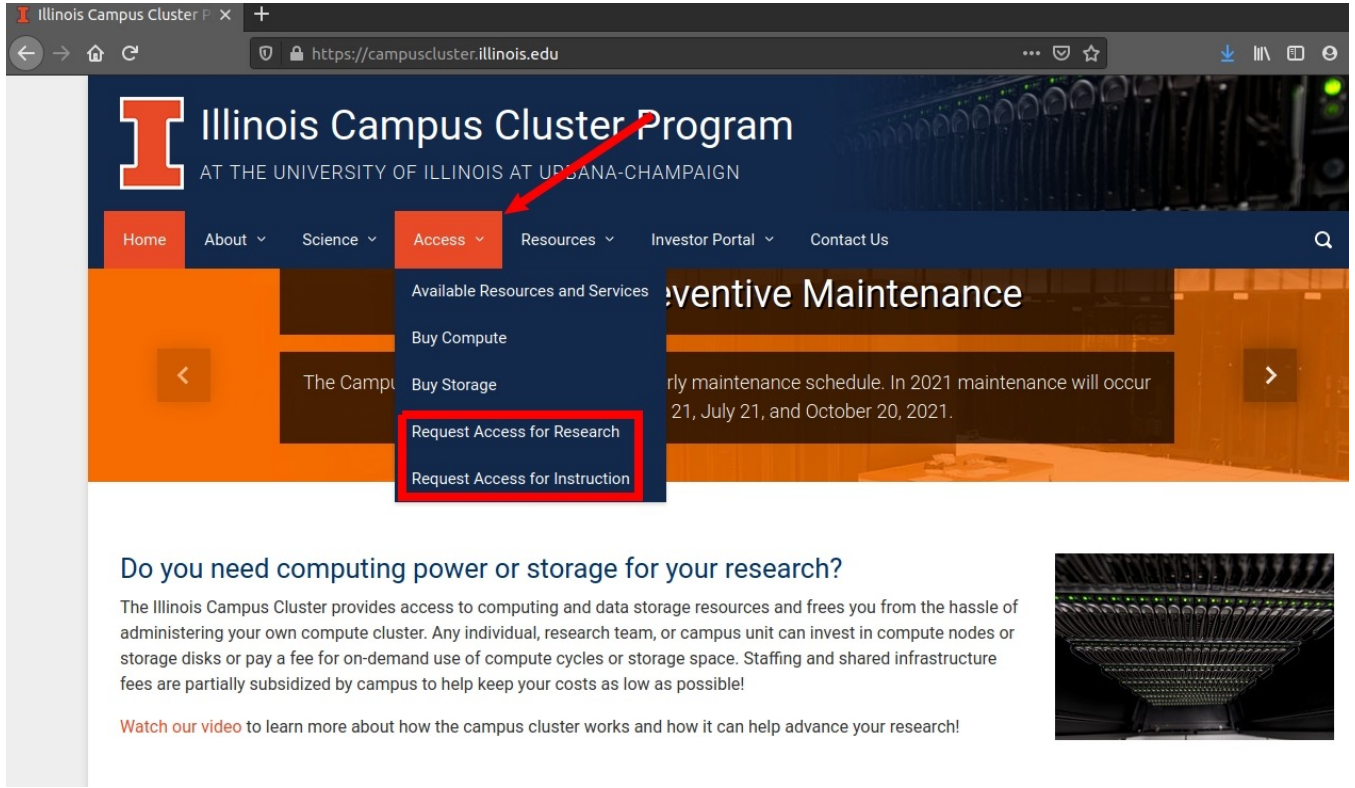
Invest in Compute Hardware

- Minimum buy-in is **one** compute node
- Standard configuration: dual CPUs and 1 TB SATA disk drive
- All equipment and infrastructure costs are approved to be charged to grants
- Infrastructure fees are charged monthly to grants
- Buy-in prices include fees covering 5 years of shared cluster infrastructure

Place an Order https://campuscluster.illinois.edu/access/buy-compute/

ICC: department or group investment

Get access through your department or group investment



List of Investors:
https://campuscluster.illinois.edu/about/investors/

- ***Access for Research*** is likely to be your option

1. You will be redirect to **Shibboleth**
2. Authenticate with your NetID and password
3. Fill out the form
   - Info about you
   - Info about Professor/Advisor
     - Name, NetID and queue

ICC: Illinois Computes

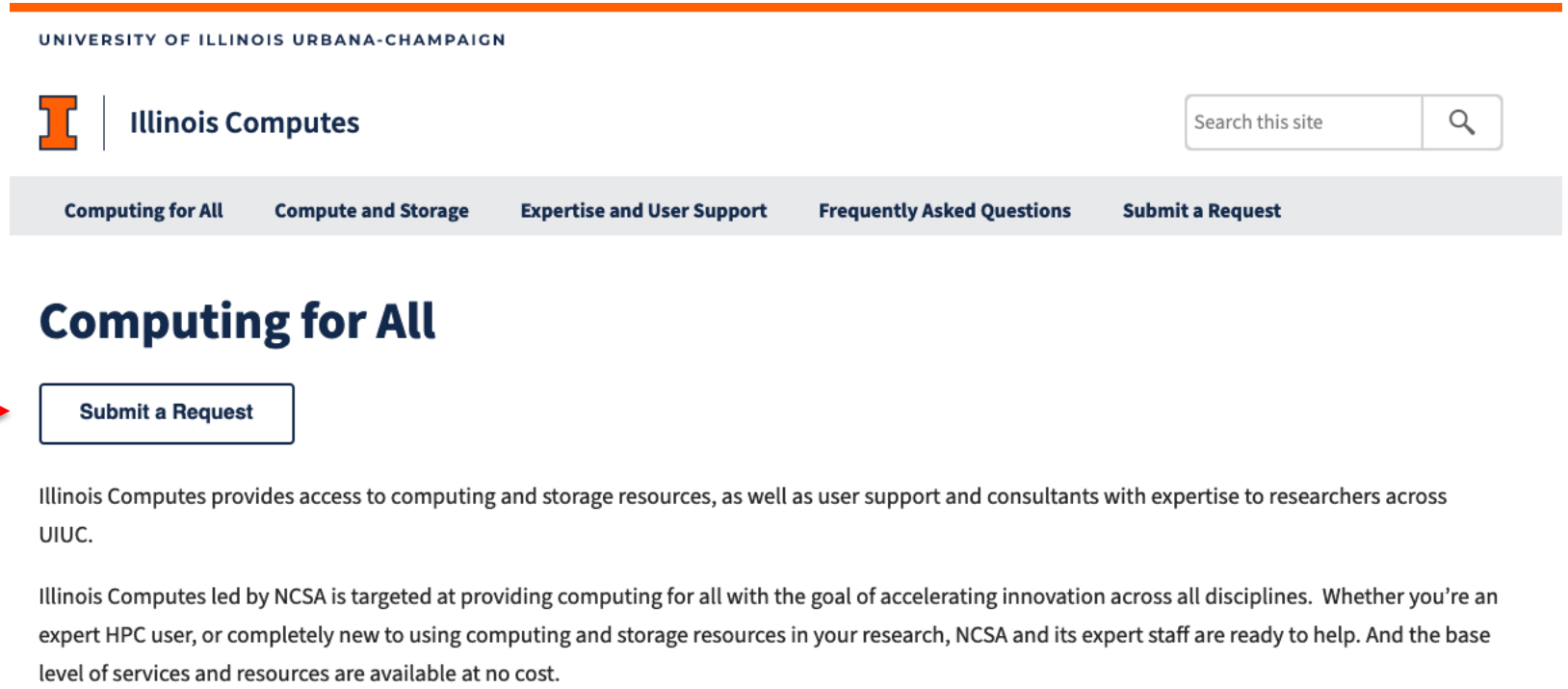
Get access at no-cost through Illinois Computes

https://computes.illinois.edu/

16 AMD EPYC nodes with 128 cores and 512 GB of RAM

4 nodes with 4x NVIDIA A100 80 GB

5 TB of local storage



Examples of research applications

Dark Energy Survey maps the skies

Dr. Felipe Menanteau and the DES team uses ICCP to try to understand where everything came from and if we are alone.

What makes us humans?

Drs. Tandy Warnow and Michael Nute are researching what makes us different from other species and from one another by producing biological sequence analyses as it relates to genetic trees.

Unraveling the Mysteries of the Brain

Dr. Xiaohui Chen and his team are mining data to help neuroscientists find the mechanism that makes neurodegenerative diseases tick.

The Future of Water

Dr. Maria L. Chu studies how a variety of factors alter how water moves in the watershed, affecting its quantity and quality.

… and **many** other scientific applications: https://campuscluster.illinois.edu/science/teams/

Outline

Illinois Campus Cluster Program
AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Block A
~ 20 min

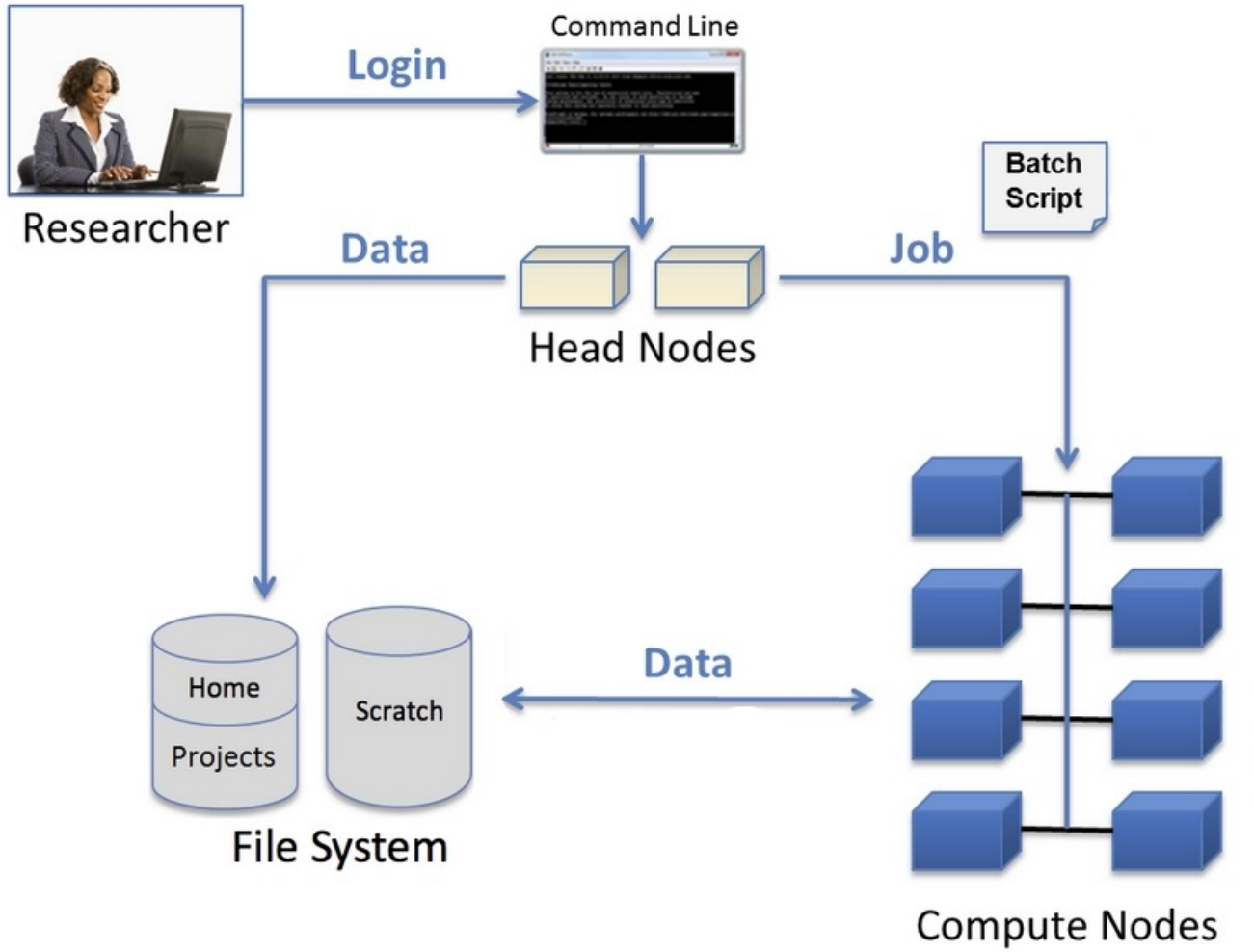
- What is the ICC?
- Who is it for?
- Examples of research applications
- Why should I use it?

Block B
~ 55 min

- How can I access it?
- How can I run my codes and applications?
- What else is available at ICC?
- Support material and further learning

- Q&A session *(~ 15 min)*

ICC Access and Usage Overview



We will discuss each one of these components with 8 hands-on exercises:

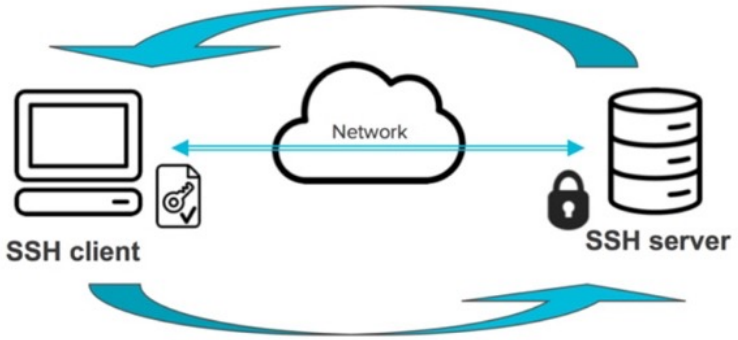
1. Accessing via SSH
2. The file system
3. Copying and moving file
4. Compiling code
5. Using modules
6. Job scripts
7. Keeping track of your jobs
8. Transferring files

SSH-client access

SSH access is usually done via *command line*, so:

SSH? Secure Shell

Provides a secure channel over an unsecured network by using *client-server* architecture



1. Open your Terminal

Mac OS

- Use Spotlight Search (Command+Space)
or
- Use Finder

to search for "Terminal"

Windows OS

Start -> Windows System -> Command Prompt

Linux OS

- Ctrl+Alt+T
or
- Search for "Terminal" in Applications

2. Connect using *ssh*

```
ssh -l <YourUsername> cc-login1.campuscluster.illinois.edu
```

Exercise 1: Access via SSH client

a) Open your Terminal

"Terminal" on *Spotlight Search* or *Finder*

Start -> Windows System -> Command Prompt

Ctrl+Alt+T

b) Connect to ICC

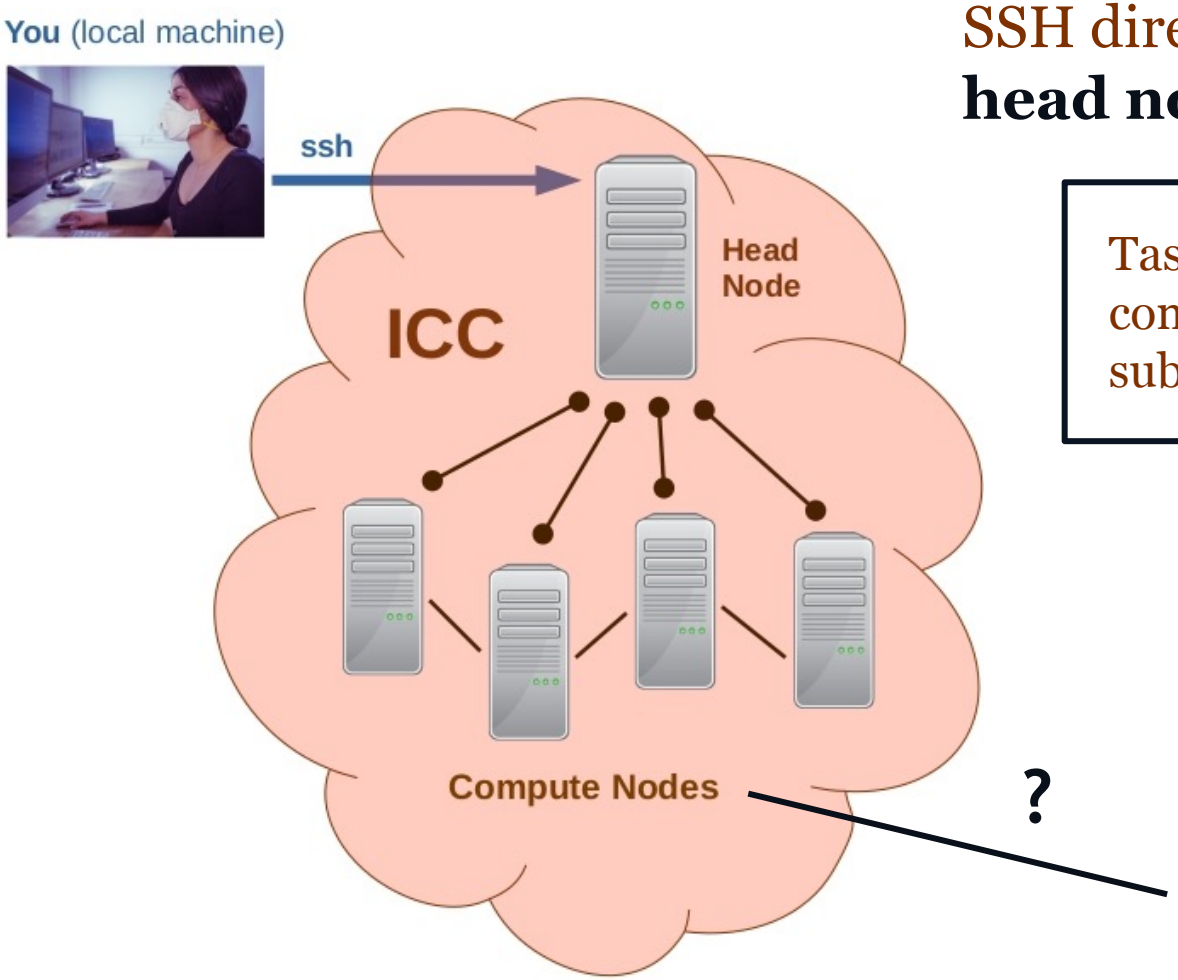
```
ssh -l YourUsername cc-login1.campuscluster.illinois.edu
```

You may get:

The authenticity of host 'blah.blah.blah (10.10.10.10)' can't be established.
RSA key fingerprint is a4:d9:a4:d9:a4:d9a4:d9:a4:d9a4:d9a4:d9a4:d9a4:d9.
Are you sure you want to continue connecting (yes/no)?

YES

Head and Compute Nodes



SSH directs you to one of the cluster's **head nodes** by default

Tasks such as file editing, code compilation, data backup, job submission, and tracking

! **Do not run your applications here**

Once you are here, your Terminal is essentially a Bash Shell in the cluster environment (Linux)

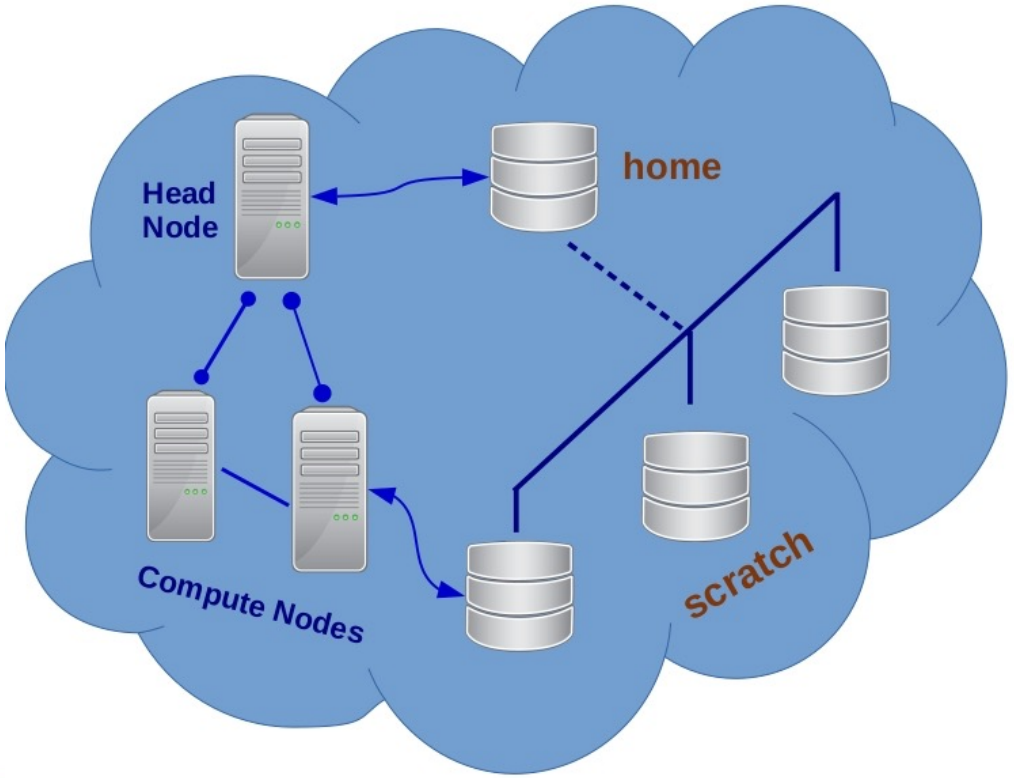
The **compute nodes** are where your applications are going to run - *we will get there soon!*

File System

- Nodes perform **computing work** with light (*Head*) or heavy (*Compute*) loads

↓

data manipulation



Main Storage Areas

Name	Quota	Purge Policy	Snapshots
home	5 GB/user	No Purge	Daily for 30 days
scratch	10 TB/user	Daily Purge for files older than 30 days	No

! Files older than 30 days in the **scratch** area are ***automatically purged***.

Backup consistently!

Exercise 2: Navigating the file system

a) Use **pwd** to verify your current folder/area

b) Use **ls** to see what files/folders you have there

c) Use **cd** to change current directory to **scratch.** Repeat a, b.

```
cd <target_directory>
```

d) Use **quota** to verify your usage in each area

Symbol	Target
.	Current directory
..	Parent directory
~	Home directory
-	Previously accessed directory

Exercise 3: Copying and moving files

a) Use **`mkdir`** to create a new folder in your home area named ***`HelloWorld`***

```
mkdir <new_folder_name>
```

b) Use **`cp`** to copy the ***`HelloWorld.cpp`*** file located at ***`/home/babreu/ICCPWorkshop`*** to your home folder

```
cp <original_file_location> <copy_location>
```

c) Use **`mv`** to move this file to the folder you created in a)

```
mv <current_file_location> <target_file_location>
```

d) Use **`cp`** to copy the file that you just moved from ***`HelloWorld`*** to home

e) Verify that copies are indeed in both folders using **`ls`**

f) Remove the copy in your home folder using **`rm`**

```
rm <file_location>
```

Tip: after each item, use **`ls`** to check if it worked!

Compiling code

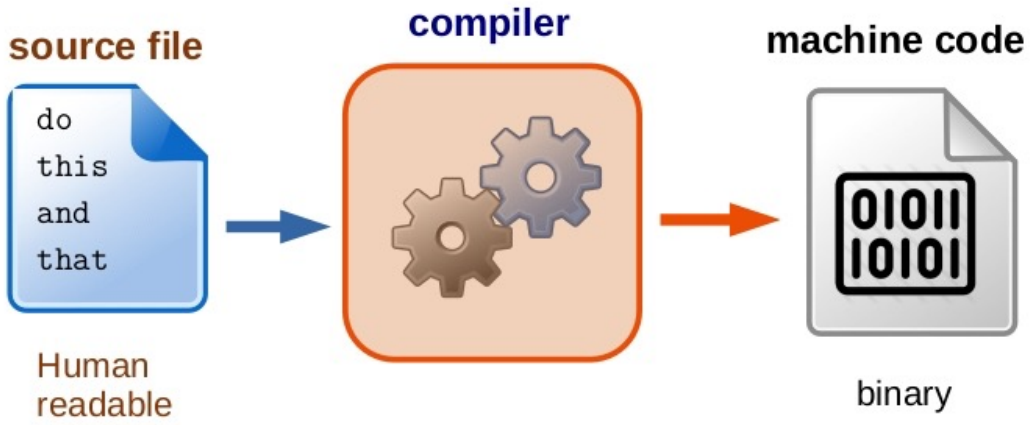
GNU compilers are by default available when you log in the **Head Nodes**

- **gcc**: compiles C code
- **g++**: compiles C++ code
- **gfortran**: compiles Fortran code

General compiling syntax is

```
compiler <source_code_file> -o <binary_file>
```

Language	GNU compiler command
C	`gcc MyProg.c -o MyProg.exe`
C++	`g++ MyProg.cpp -o MyProg.exe`
Fortran	`gfortran MyProg.f -o MyProg.exe`



Many other compilers are available in ICC, but if you want to use them, you need to **load** them – *we will cover that too!*

Exercise 4: Compiling code

a) Go to the directory where the file from the previous exercise is located

b) Check the version of the GNU compiler that is available to you in the Head Node

```
gcc --version
```

c) Compile *HelloWorld.cpp* using **g++**

```
g++ HelloWorld.cpp -o HelloWorld.exe
```

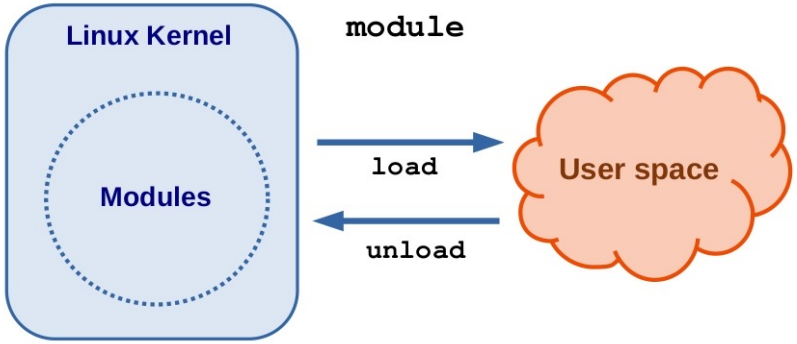
d) Verify that the binary has been generated in your folder

Working with modules

- What if you want to use a different compiler, or even a different version of GCC?

Modules allow you to dynamically load and unload components of the operating system as you need them!

The same approach applies to loading a variety of software and libraries, as well as compilers



Useful `module` options

Command	Description
`module` **`list`**	Lists current modules loaded in your session
`module` **`avail`**	Lists all available modules
`module` **`help`** *<module_file>*	Information about *module_file*
`module` **`load`** *<module_file>*	Loads *module_file* to your environment
`module` **`unload`** *<module_file>*	Removes *module_file* from your environment

Exercise 5: Working with modules

a) Verify the modules currently loaded in your shell environment

```
module list
```

b) Check the list of modules that are available to you

```
module avail
```

c) Try to use the Intel compiler to compile HelloWorld.cpp

```
icc HelloWorld.cpp -o HelloWorld_Intel.exe
```

Make sure you are in the right folder!

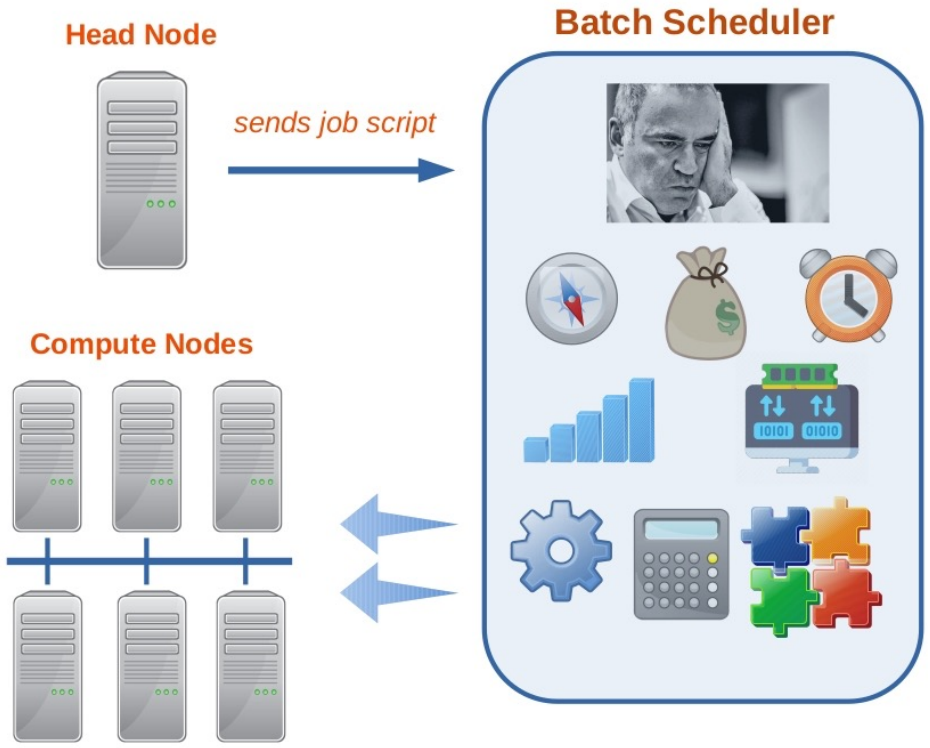
d) Load the intel/18.0 module and repeat c)

```
module load intel/18.0
```

e) Unload the intel/18.0 module

```
module unload intel/18.0
```

Batch schedulers (Slurm)



Now that we know how to

- **Log in to ICC**
- **Navigate file system**
- **Load modules**
- **Compile code**

It is time to use the cluster's power!

Batch processing is "the running of jobs that can run without end-user interaction, or can be scheduled to run as resources permit."

The **batch scheduler** is a computer application that weights in a number of factors and resources to determine *where, how and when* a certain request is going to run in the **compute nodes**

In ICC, this application is SLURM:

https://slurm.schedmd.com/

Batch job scripts

For SLURM to do its magic, we need to provide it with some information…

Job Script

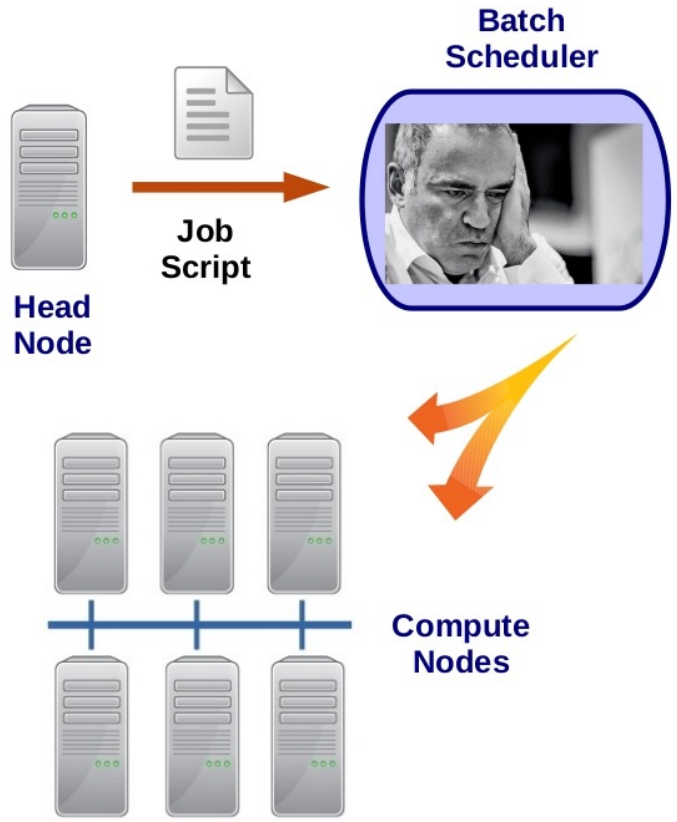
We will send bash commands

```
#!/bin/bash

#SBATCH --time=XX:YY:ZZ
#SBATCH --nodes=N
#SBATCH --ntasks-per-node=M
#SBATCH --partition=Where
#SBATCH --output=OutFile
#SBATCH --error=ErrorFile
#SBATCH --mail-user=Email
#SBATCH --mail-type=When

./HelloWorld.exe
```

- Required time
- Number of nodes
- Number of tasks per node
- Where to run it
- Output file
- Error file
- Who to email
- When to email
- What to run



Exercise 6: Job script for HelloWorld

a) Copy the file ***HelloWorld.sbatch*** located at ***/home/babreu/ICCPWorkshop*** to your ***HelloWorld*** folder

```
cp /home/babreu/ICCPWorkshop/HelloWorld.sbatch .
```

b) Use **`cat`** to verify the information that is being passed to the scheduler

```
cat HelloWorld.sbatch
```

c) Submit your job scrip to the batch scheduler using **`sbatch`**

```
sbatch HelloWorld.sbatch
```

d) Take note of your job ID

Exercise 7: Keeping track of your job

- Use the **squeue** commands to monitor the status of your job

Useful SLURM commands

Command	Action
`sbatch <script_file>`	Submit job script to the scheduler
`squeue -a`	List status of all jobs in the batch system
`squeue -u <Username>`	List status of all your jobs
`squeue -j <JobID>`	Lists information about a job
`scancel <JobID>`	Kills a job

File transfers using SCP

- **There is a specific node for transferring files over the network**

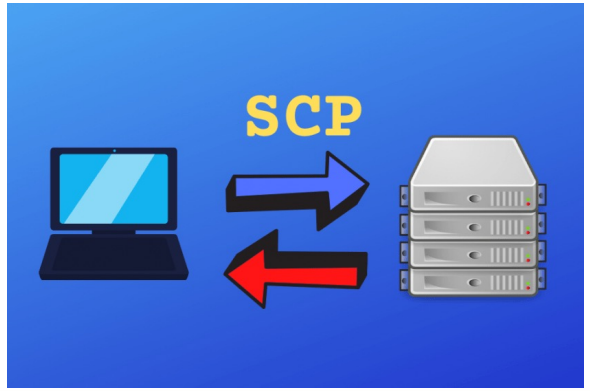
Sending file from ICC to your machine:

```
scp <Username>@cc-xfer.campuscluster.illinois.edu:<path_to_file> <destination>
```

Sending file from your machine to ICC

```
scp <path_to_file> <Username>@cc-xfer.campuscluster.illinois.edu:<destination>
```

Secure Copy Protocol



- Works basically like SSH
- Safe channel to copy/transfer data from or to your local machine
- Command line interface

There are **many** other options available for managing files at ICC

https://docs.ncsa.illinois.edu/systems/icc/en/latest/user_guide/storage_data.html

Exercise 8: Transfer files with SCP

a) Log out of your cluster session using `logout`

b) Use `scp` to copy the ***HelloWorld.out*** file to your local machine

```
scp <Username>@cc-xfer.campuscluster.illinois.edu:~/HelloWorld/HelloWorld.out .
```

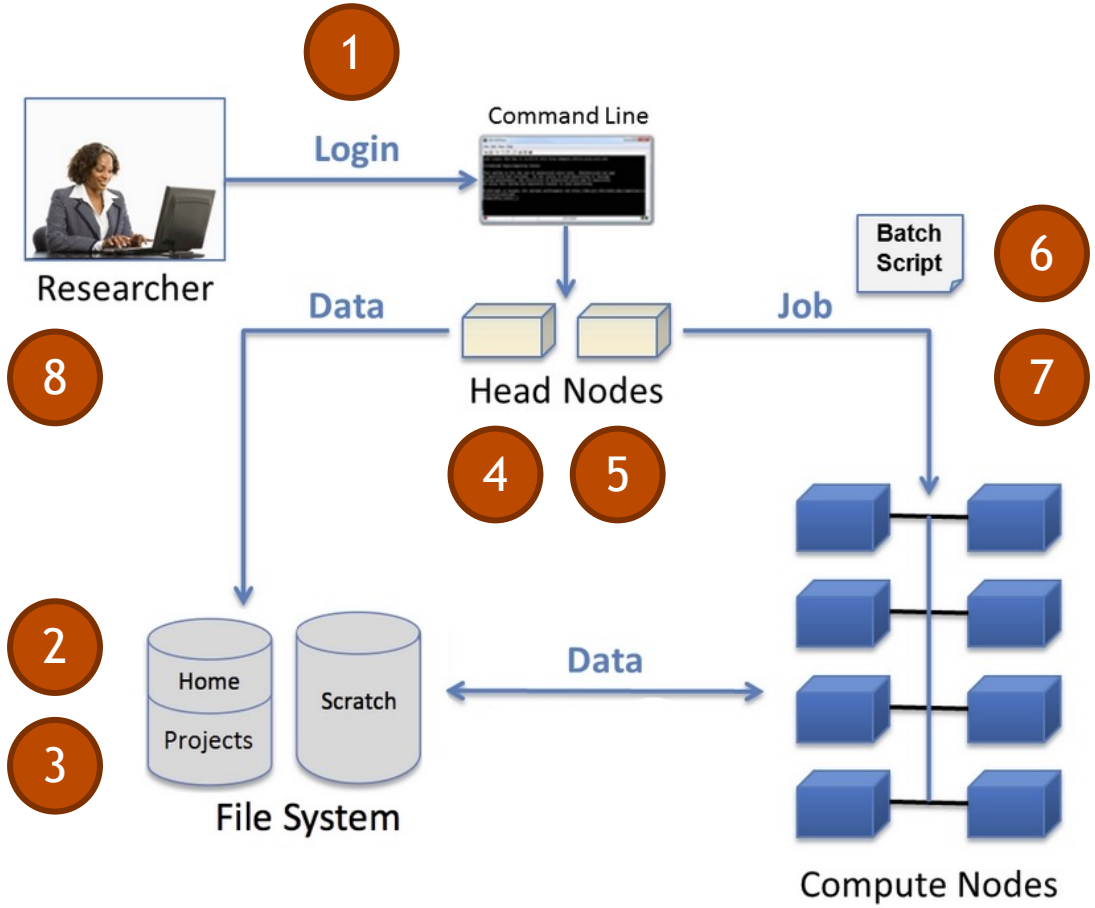
Your cluster /home folder

Where you opened your terminal

c) Look for the file in your machine and try to open it using your text editor

Cooling down

We now have a pretty good idea about navigating the Illinois Campus Cluster…



1. Login using SSH client
2. Understand and use the file system
3. Copy and move files around
4. Compile code
5. Load and unload modules
6. Write and submit a job script
7. Keep track of your jobs
8. Transfer files to your local machine

Stretching

A few other exercises that can help developing the concepts we have just learned:

- Edit the job script to include your email address that will be notified when your job starts running and when it terminates
- Write a simple code in your local machine. Transfer it to the cluster, compile it. Write a job script for your executable and run it in the Compute Nodes. Transfer the output back to your machine and analyze it
- Use one of the other options available in the Storage and Data Guide to manage your files with a graphic interface
- Load and unload different modules: if you have some experience with Python, try to use the console and perform some simple Python 3 tasks.
- Fire up an interactive session on the Compute Nodes using `srun` (Running Jobs Guide)
- Get creative and use your temporary allocation to learn more about the cluster!
- Submit your Illinois Computes request!

Resources

- **ICCP official *Getting Started* webpage**

  https://docs.ncsa.illinois.edu/systems/icc/en/latest/getting_started.html

- **ICCP official *User Documentation***

  https://docs.ncsa.illinois.edu/systems/icc/en/latest/index.html

- **ICCP System Info: Hardware, Storage, Cluster Monitor**

  https://campuscluster.illinois.edu/about/system-info/

- **Self-paced *Getting Started on ICC* – HPC Moodle page**

  https://www.hpc-training.org/xsede/moodle/

Resources

- **SLURM Documentation**

  https://slurm.schedmd.com/quickstart.html

- **Basic Linux commands for beginners**

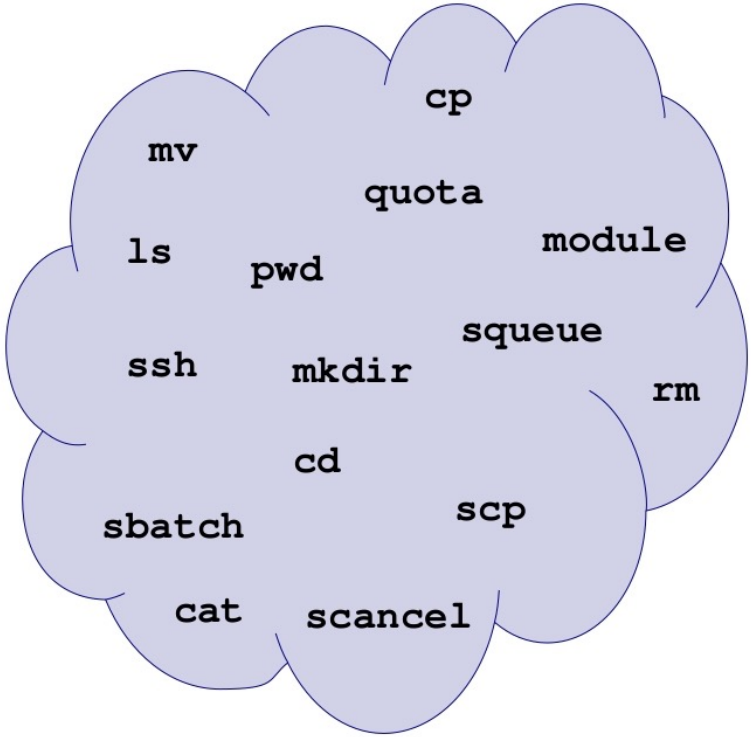
  https://maker.pro/linux/tutorial/basic-linux-commands-for-beginners

- **Environment Modules documentation**

  https://modules.readthedocs.io/en/latest/

- **Examples of Best Practices for running jobs in general**

https://docs.nersc.gov/jobs/best-practices/
https://dccn-hpc-wiki.readthedocs.io/en/latest/docs/cluster_howto/best_practices.html



Getting help

If you are having issues using the Illinois Campus Cluster and cannot find a solution in the official documentation:

- **Getting Started**

  https://docs.ncsa.illinois.edu/systems/icc/en/latest/getting_started.html

- **User Documentation**

  https://docs.ncsa.illinois.edu/systems/icc/en/latest/

Reach out!

help@campuscluster.illinois.edu

Outline

Illinois Campus Cluster Program
AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Block A
~ 20 min

- What is the ICC?
- Who is it for?
- Examples of research applications
- Why should I use it?

- How can I access it?
- How can I run my codes and applications?
- What else is available at ICC?
- Support material and further learning

Block B
~ 55 min

- Q&A session *(~ 15 min)*

Feed us back!

Feedback is vital for our continuous improvement as instructors/lecturers, and for our organization

Please take a couple of minutes to answer this anonymous survey. Any sort of comments are extremely helpful, and all of them are going to be carefully taken into consideration.

Thank you!

https://go.ncsa.illinois.edu/ICCPBasicsFeedback

Q&A

Have further questions about this workshop?

babreu@illinois.edu

Feedback form

https://go.ncsa.illinois.edu/ICCPBasicsFeedback

National Center for Supercomputing Applications
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN