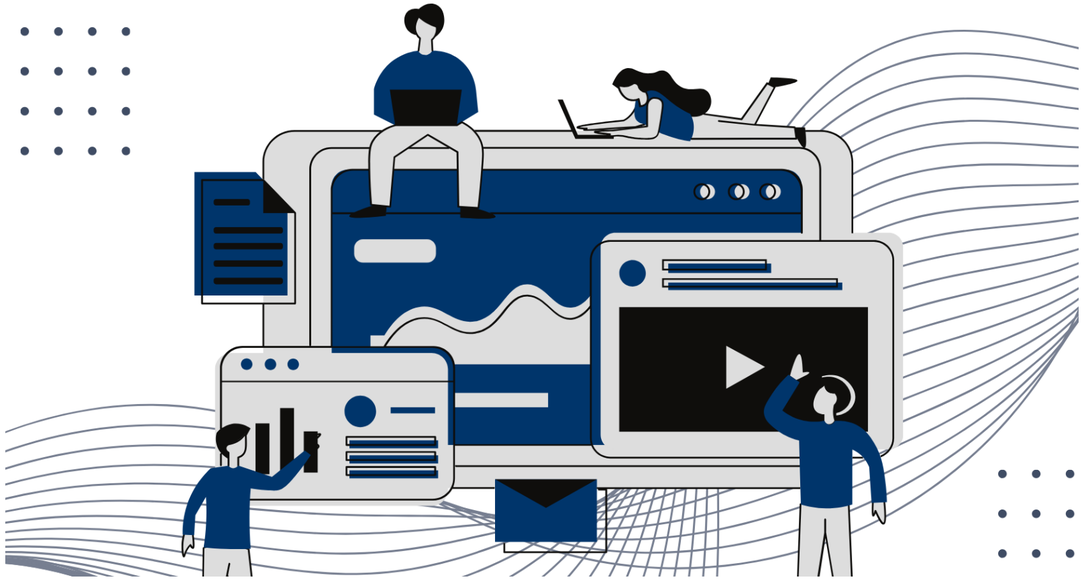
Yale SCHOOL OF PUBLIC HEALTH

Data Science and Data Equity



Information Session on High Performance Computing, and Yale's AI/Clarity Platform

Speakers For Today

Aya Nawano, PhD
Computational Research Support Analyst
Yale Center for Research Computing

Weis Rafi, PhD
Associate CIO, Health Sciences
Health Science IT

Hadar Call
Associate. CIO
Yale IT

Yale SCHOOL OF PUBLIC HEALTH
Data Science and Data Equity

Yale *Center for Research Computing*

Research Computing Support

Aya Nawano, PhD

Computational Research Support Analyst

Yale Center for Research Computing

- Independent center under the Provost's office
- Created to support research computing
- Focus on high performance computing, storage
- Engineering staff operates and maintains our resources
- Application and research support specialists
    - Consult, collaborate, educate researchers

research.computing.yale.edu/

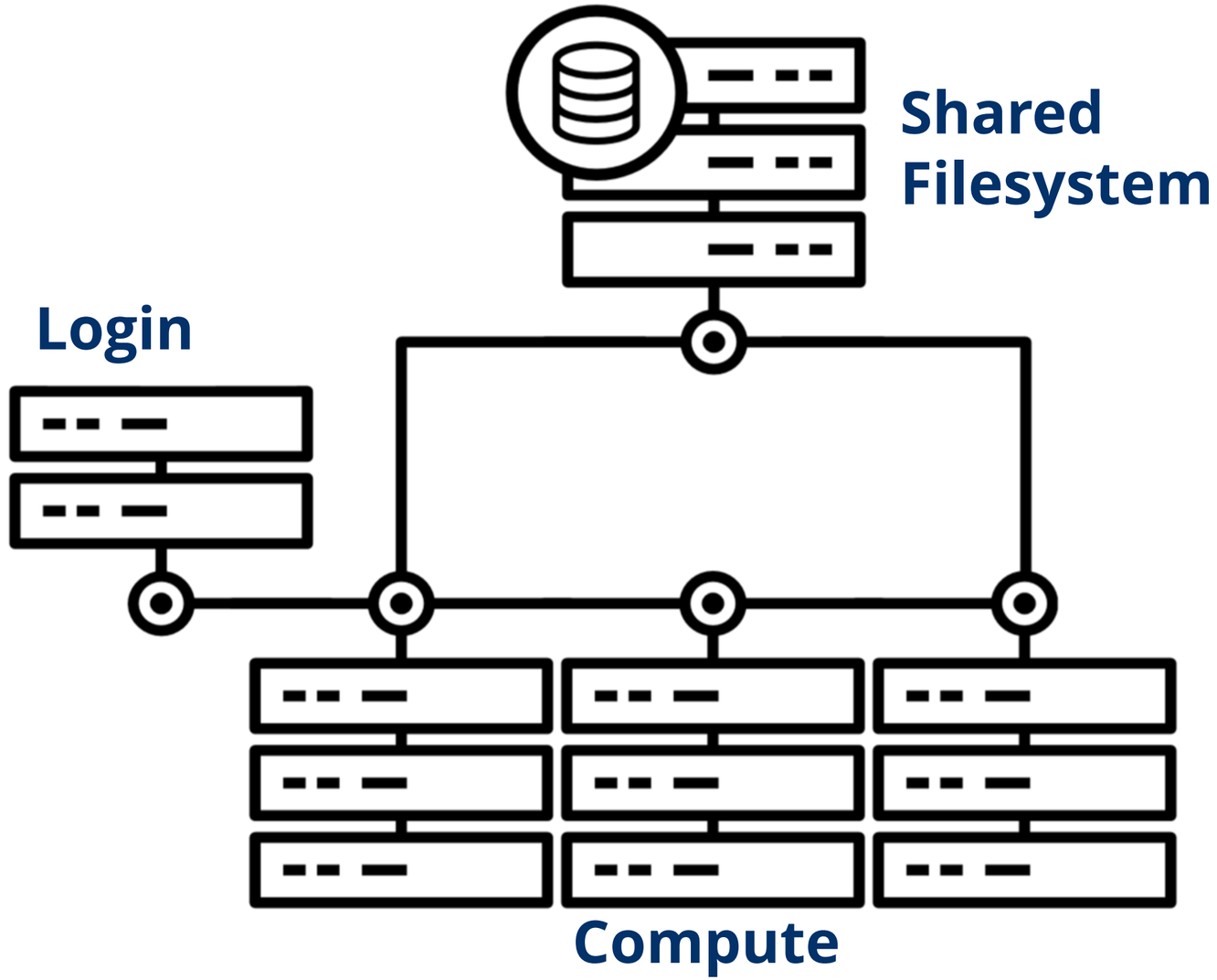
What is a Cluster?

Clusters are modern day supercomputers

- 100s-1000s of rack-mounted computers
- Networking
- Storage

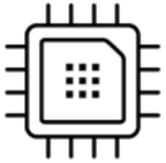
Abstract Cluster Diagram

- Access via the login nodes
- Shared filesystem presents data across all nodes
- Submit jobs scheduled to run on compute nodes



Compute Node

CPU



Many Cores
~24-64

RAM

Memory
~90GiB-4TiB

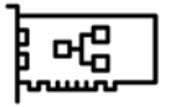
HDD

Temporary, per-node
/tmp

GPU

Accelerates some jobs
Request specially
Not on all nodes

NIC



High bandwidth
Low latency
Tuned for node to node

Why Use a cluster?

- Don't want to tie up your own machine
- Have many long running jobs to run
- Want to run in parallel to get results quicker
- Need more disk space
- Need more memory
- Want to use GPUs
- Want to use software installed on the cluster
- Want to access data stored on the cluster
- YCRC support

Yale Clusters

Any faculty's group may request accounts:

- 125 GiB/person in home
- 1- 4 TiB/group in long-term project
- 10TiB/group in short-term scratch
- Free access to GPUs and CPUs, Large Memory nodes

Private servers & additional storage for purchase

Priority access available for cost

Your PI may already have a group on a cluster

Yale Clusters

Cluster	CPUs	Purpose
Grace	29,000	General
McCleary	13,000	Medical/life science, YCGA
Milgram	2,400	HIPAA
Misha	~2,000	Wu Tsai Institute
Bouchet	3,800	Future general purpose, MPI
Hopper	6,000	Available in summer 2025, NIST 800-171+ePHI

Limitations of HPC Clusters

- Cannot run Windows or macOS programs
- Not for persistent services (DBs or web servers)
- Not ideal for some graphical tasks
- Bad for jobs that run for weeks (unless checkpointed)

Accessing the Clusters

Only reachable from Yale campus network:

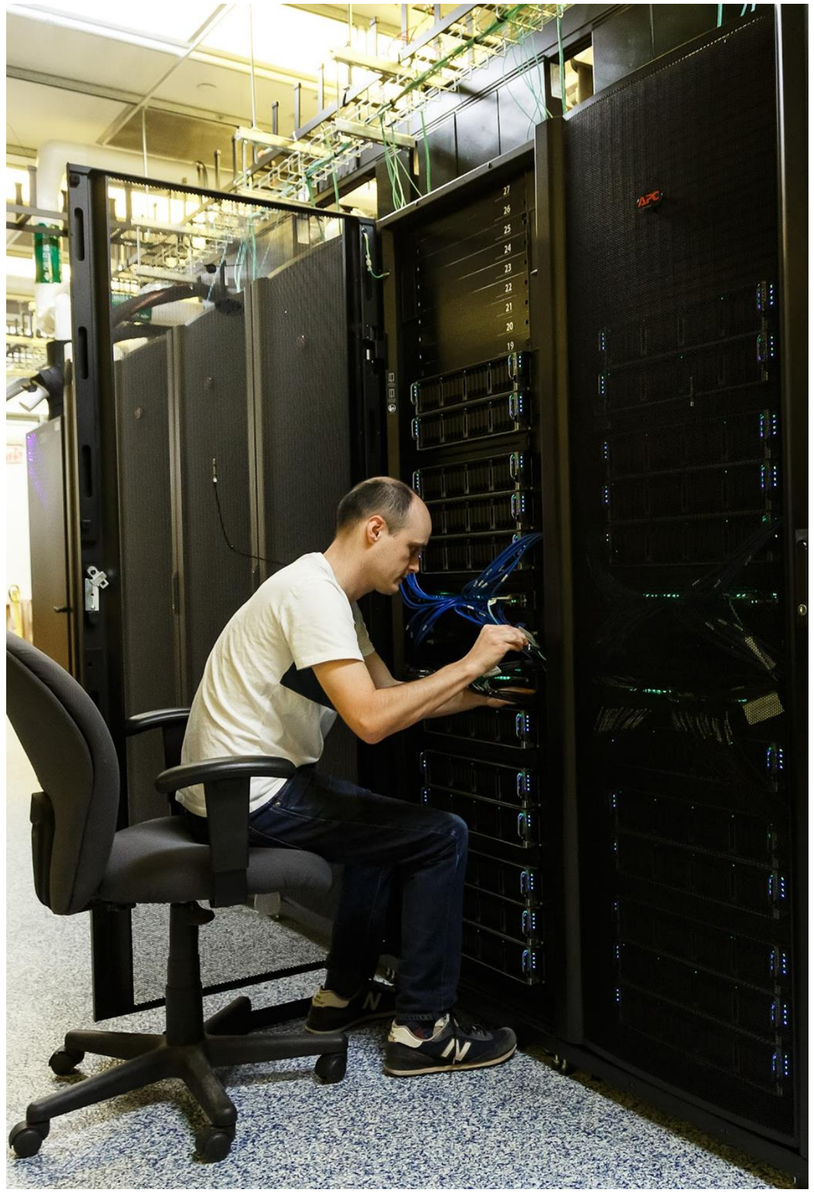
- YaleSecure
- Ethernet (on campus)
- Yale VPN, if off campus

Three general methods to login:

- Open OnDemand, web-based login
- Command line ssh (easiest from linux or macOS)
- Graphical ssh tool (MobaXterm)

Last two methods use `ssh` and require an ssh key.

See also: docs.ycrc.yale.edu/clusters-at-yale/access



Open OnDemand

Open OnDemand is web-based, and handles:

- Direct login (shell) access
- File transfer to and from your local machine
- Edit text files
- Virtual desktop on cluster
- Jupyter notebook, RStudio, MATLAB

ood-grace.hpc.yale.edu

OPEN OnDemand

Production Deployments

Software

Over 400 software packages are pre-installed on the clusters.

- Available as software “modules”
- Commonly used software packages, such as
  - MATLAB
  - Miniconda (for Python)
  - R
  - Compilers and common math libraries

If something you need is missing, let us know and we can add it.

See also: docs.ycrc.yale.edu/clusters-at-yale/applications/modules/

Interactive vs. Batch

Interactive jobs:

- Like a remote session
- Requires an active connection
- For development, debugging, or interactive environments like R
- One or a few jobs at a time

Batch jobs:

- Non-interactive
- Can run many jobs simultaneously
- Usually your best choice for production computing

Partitions

Compute Nodes are divided into separate “partitions” which group the nodes by use case and capabilities.

“Common” public partitions available to any researcher

- devel, day, week, mpi, gpu, scavenge

Priority partitions for priority access

Private group partitions

Get Started

Request an account

research.computing.yale.edu/support/hpc/account-request

If you are not faculty, you need a faculty sponsor

Getting Started Information

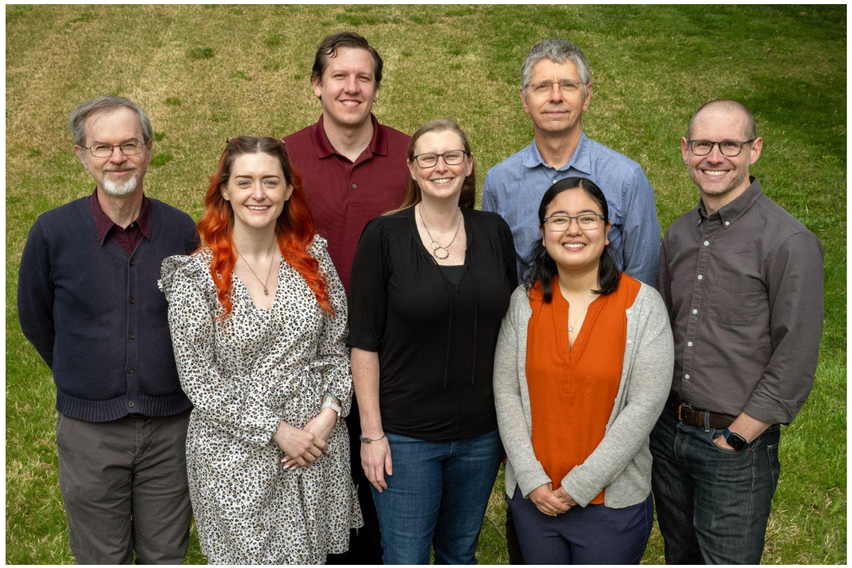
Getting Started Guide: docs.ycrc.yale.edu/clusters-at-yale/

Intro to HPC workshop: YouTube video, in-person

Research Support

Our Team is available via email (research.computing@yale.edu) and Zoom

Specialist	Cluster(s)	Areas of Focus
Kathleen McKiernan	All	Getting Started
Rob Bjornson, Ph.D.	McCleary	Life Sciences, Bioinformatics, Python, R
Sam Friedman	McCleary	Life Sciences, Bioinformatics, Python, R, Snakemake, Nextflow
Tom Langford, Ph.D.	Grace/Bouchet/Milgram	Physics, EPS dept, Python, MPI
Aya Nawano, Ph.D.	Grace/Bouchet	Molecular Dynamics, Matlab, C/C++, MPI
Kaylea Nelson, Ph.D.	Grace/Milgram	Astronomy, MPI, Python
Mike Rothberg, Ph.D.	Grace/McCleary	AI/ML, Computational Chemistry, Python, Matlab
Chuck Sindelar, Ph.D.	McCleary	Life Sciences, Cryo-EM/Structural Biology, Python, MPI, C/C++, Matlab
Michael Strickler, Ph.D.	McCleary	Life Sciences, Structural Biology
Ping Luo	Misha/Milgram	Wu Tsai Institute, Psychology dept, Open OnDemand
Misha Guy, Ph.D.		SRSC Software and Mathematica (email at mikhael.guy@yale.edu for appointment)



YCRC Office Hours

The YCRC Research Support Staff host weekly office hours over Zoom

Every Wednesday at 11am-12pm EST:

https://yale.zoom.us/my/ycrcsupport

No appointments are necessary.

Alternatively, to schedule an appointment, email us:

research.computing@yale.edu

Training Workshops

Taught by YCRC

Introduction to HPC
Practical HPC
Version Control with Git
Data Analysis Tools for Python
Tidying Data
High Performance Python
Parallel Programming with Python
Shareable Reproducible Containers
Writing Efficient R Code
Parallel Programming with R
Access to ACCESS
Youtube Channel

External and Vendor Training

Satellite Site for HPC Workshops by Pittsburgh Supercomputing Center
MATLAB Workshops by MathWorks
Cloud providers : Google, AWS, Microsoft
LinkedIn Learning recommended by YCRC

- Python
- R
- Mathematica
- MySQL
- Git and GitHub
- Linux

Questions?

Website: ycrc.yale.edu

Documentation: docs.ycrc.yale.edu

Contact YCRC

Send an email to research.computing@yale.edu

YCRC at 160 St Ronan Street

Yale's Clarity Platform

Yale SCHOOL OF PUBLIC HEALTH
Data Science and Data Equity

Mr. Nazmul Islam

Inaugural Director of Research Systems

Mr. Islam will lead a newly created Research Systems team within Health Sciences IT composed of 3 core units:

1. Research Compute Infrastructure (RCI)
2. Lab Solutions
3. Specialty Systems

This team provides services and support for secure/high-risk research computing environments, implementation of proprietary lab equipment to computing systems, customizing scientific software and workflows, research data storage and transfer needs, and niche research computational services for the Health Sciences campus.

Yale SCHOOL OF PUBLIC HEALTH
Data Science and Data Equity

Questions?

sph.yale.edu/dsde

@YaleSPH

Public Health Data Science and Data Equity
Yale School of Public Health
60 College Street, New Haven, CT 06510

Yale SCHOOL OF PUBLIC HEALTH