

Biowulf

A High Performance Computing (HPC)
Resource for the IRP

staff@hpc.nih.gov

Wolfgang Resch, Support scientist

Accelerate biomedical research at the NIH by providing convenient access to large scale **computational resources** and **scientific software** (and the **know-how** to use them).

Computational resources: The Biowulf system

~100k cores / ~200k CPUs

~900 TB memory

~50 PB shared high performance storage

5 PB of object storage

~900 GPUs (K80 - A100)

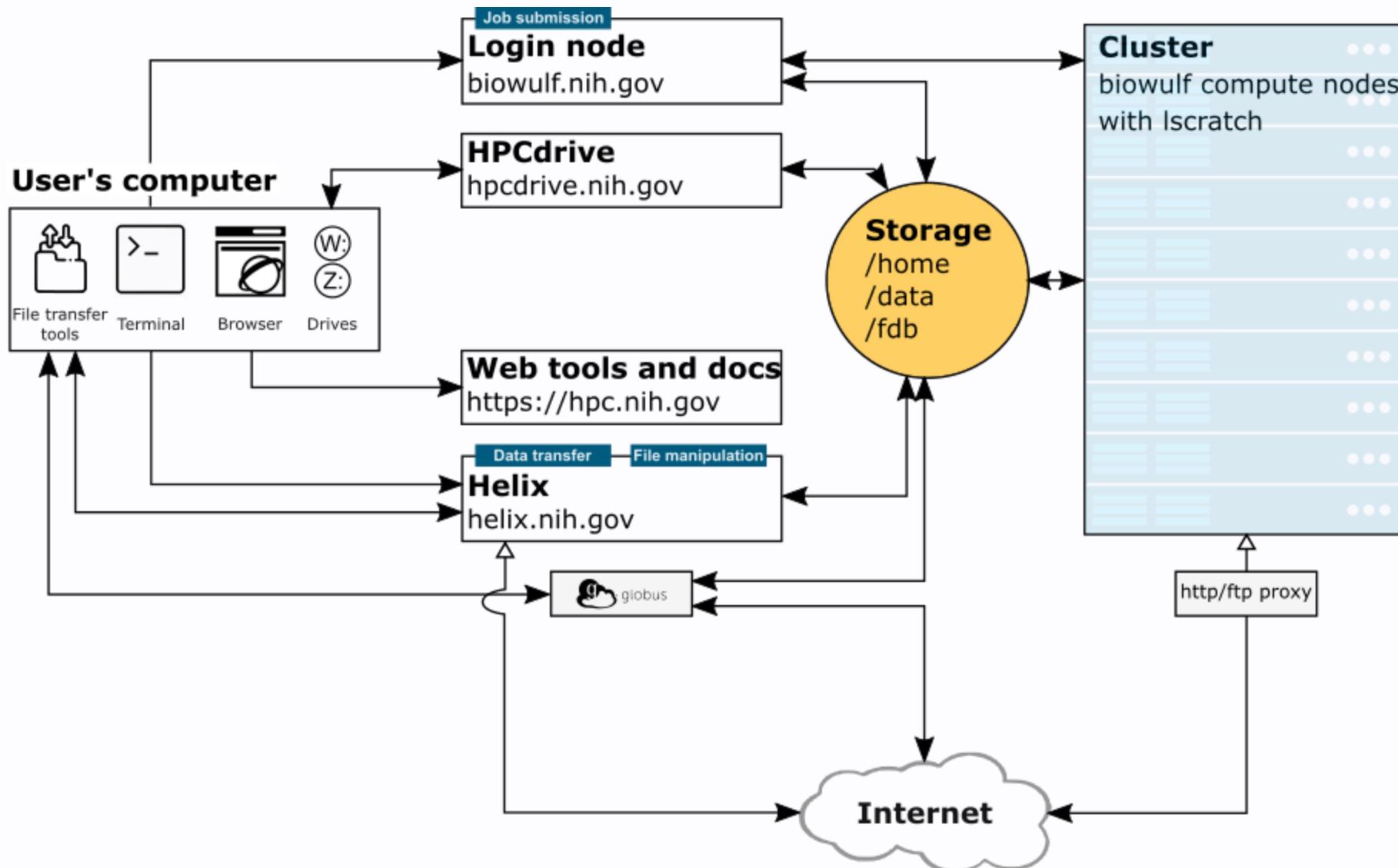
high speed / low latency networks

- + 12k CPUs
- + 160 A100s
- + 16 3TB largemem nodes

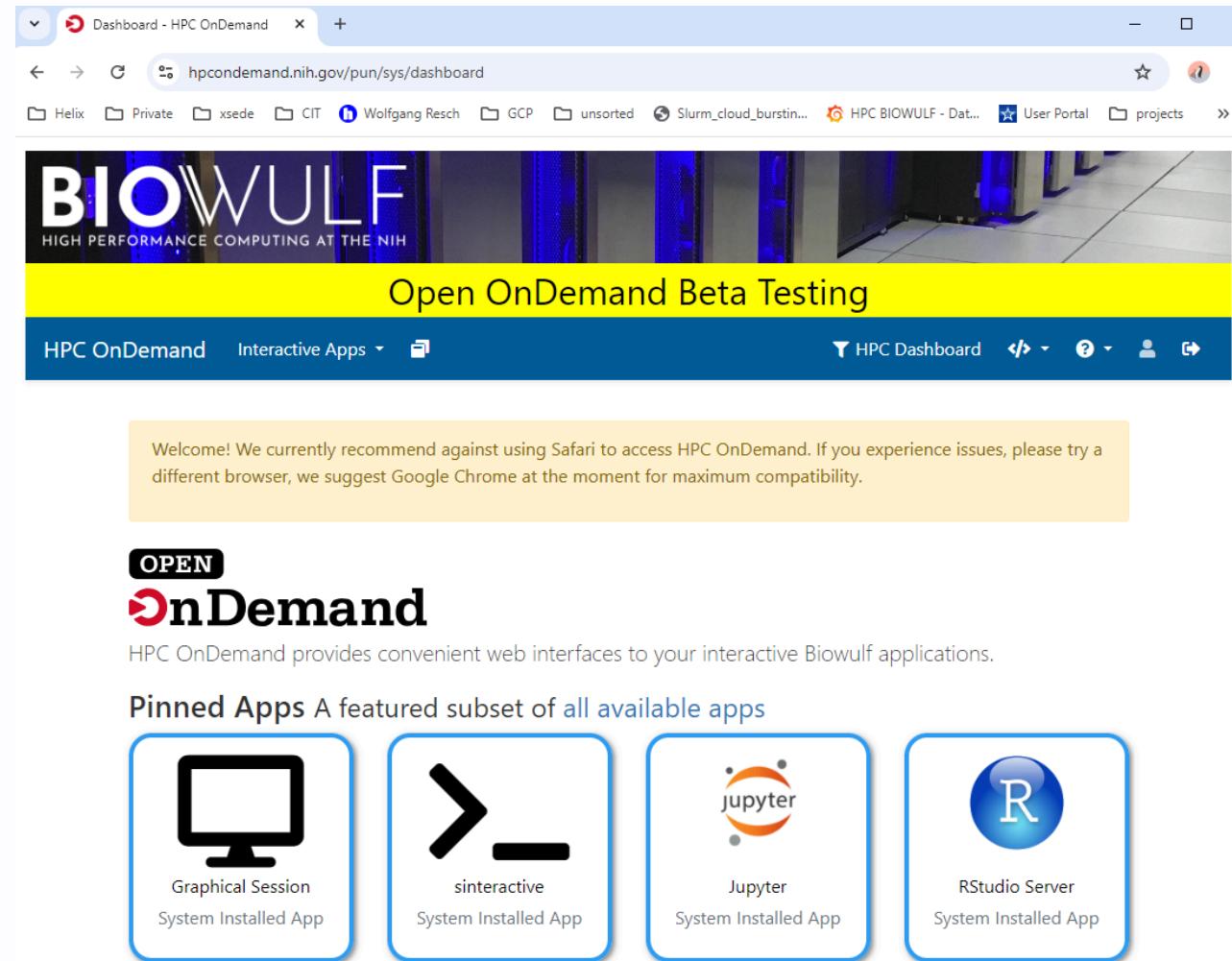
Node classes

- Single node, CPU (56-128 CPUs, 240-500GB)
- Multinode, CPU (56 CPUs, 240GB)
- GPU (K80, P100, V100, A100)
- large memory (1.5-3TB)
- visualization
- buy in nodes

Architecture



Open OnDemand beta testing



<https://hpcondemand.nih.gov>

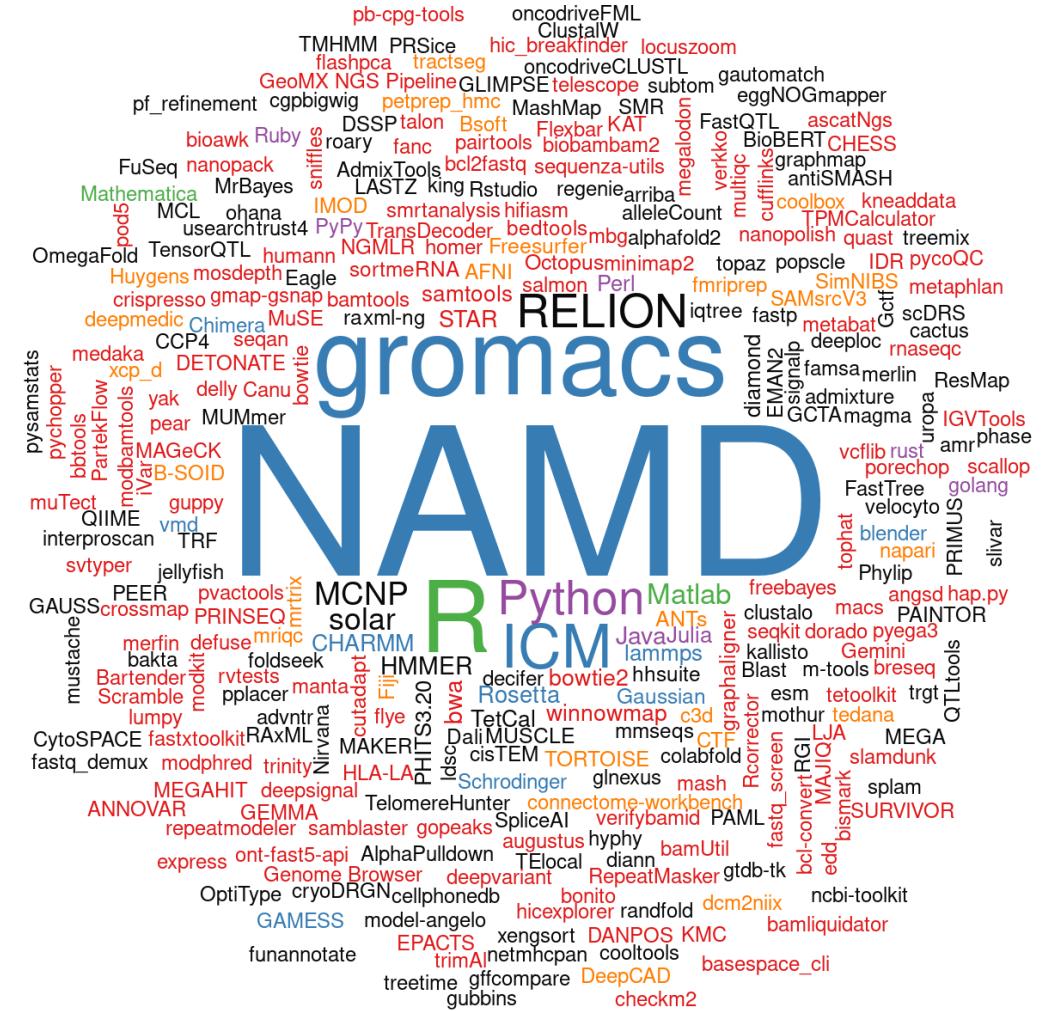
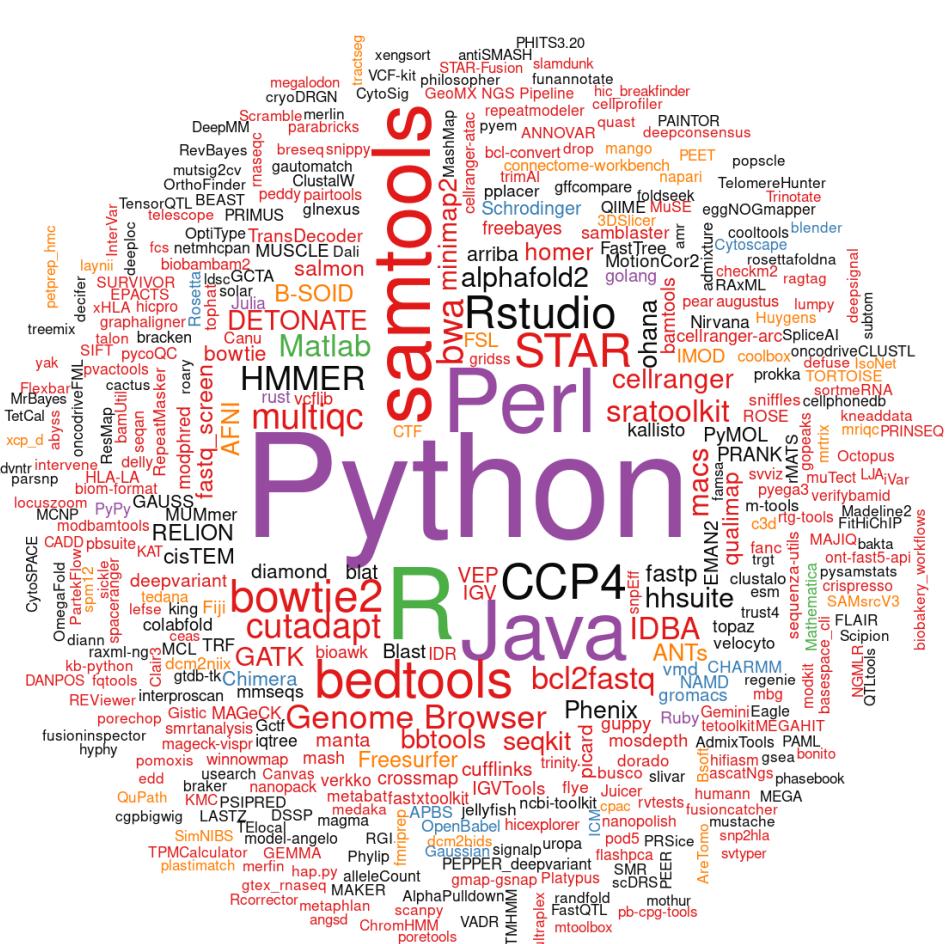
In 2023 **~2400** active users from **~75%** of intramural research groups used **1.1B** CPUh, **4.8M** GPUh and **30PB** of storage

Scientific Software

Applications

- multiple versions for ~1000 applications available
- Multiple versions of python with ~500 packages each
- Multiple versions of R with about 1600 packages each
- singularity for containerization
- jupyter, rstudio, vscode, desktop sessions

Application usage



Know how - training and
outreach

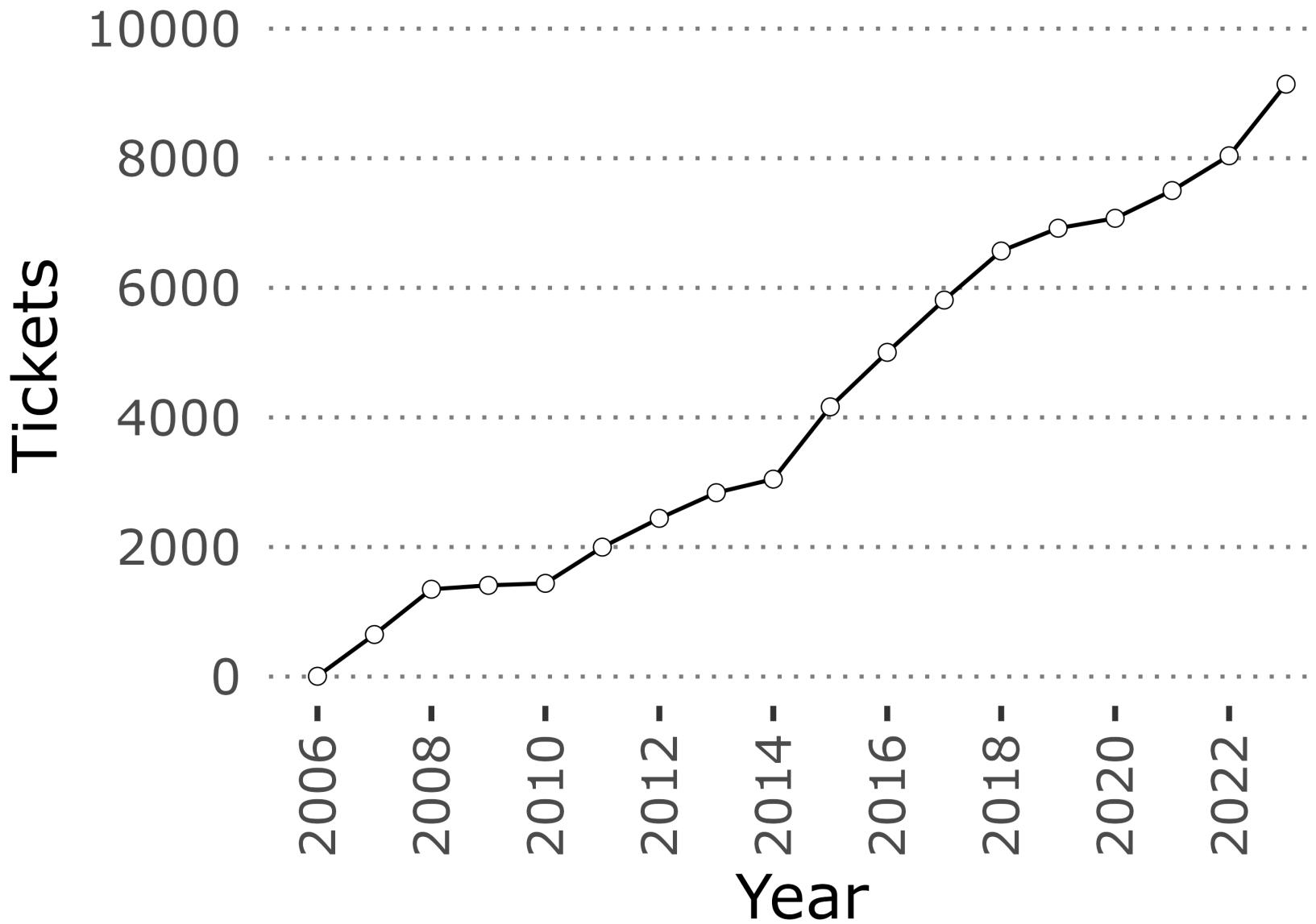
Training

- Online, self-paced classes (intro to biowulf, bash)
- Live/recorded class sessions (e.g. Deep learning by example, R, matlab, snakemake)
- Documentation (<https://hpc.nih.gov>) and tutorials (e.g. https://hpc.nih.gov/training/gatk_tutorial)
- *Assist in other groups' training*
- Student accounts

Outreach

- Monthly *Zoom-in* consult
- Meeting with individual groups or 1:1
- Cranky email from staff if/when you break something
- Let us know if you'd like somebody to attend lab/branch/group meeting

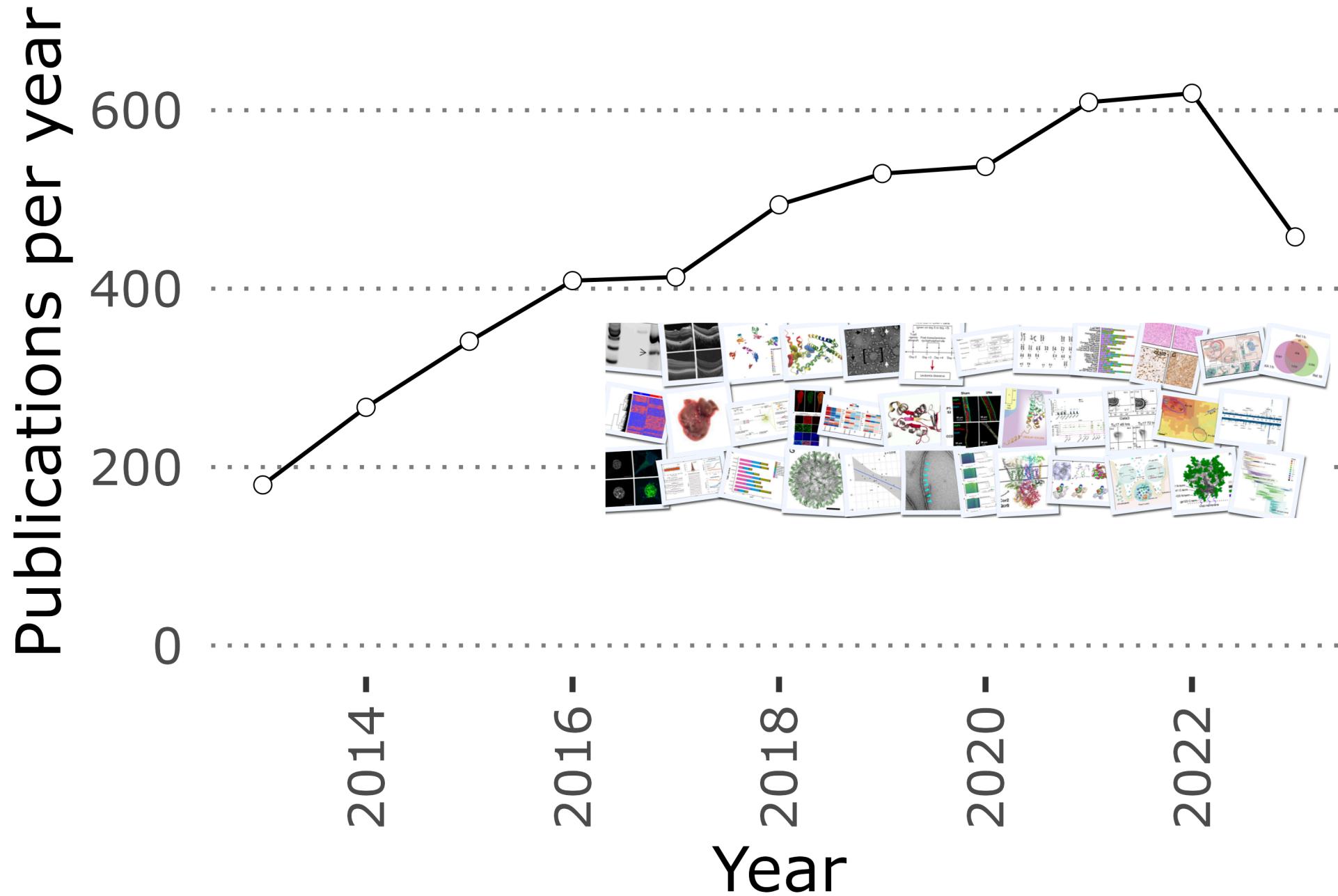
Support - staff@hpc.nih.gov

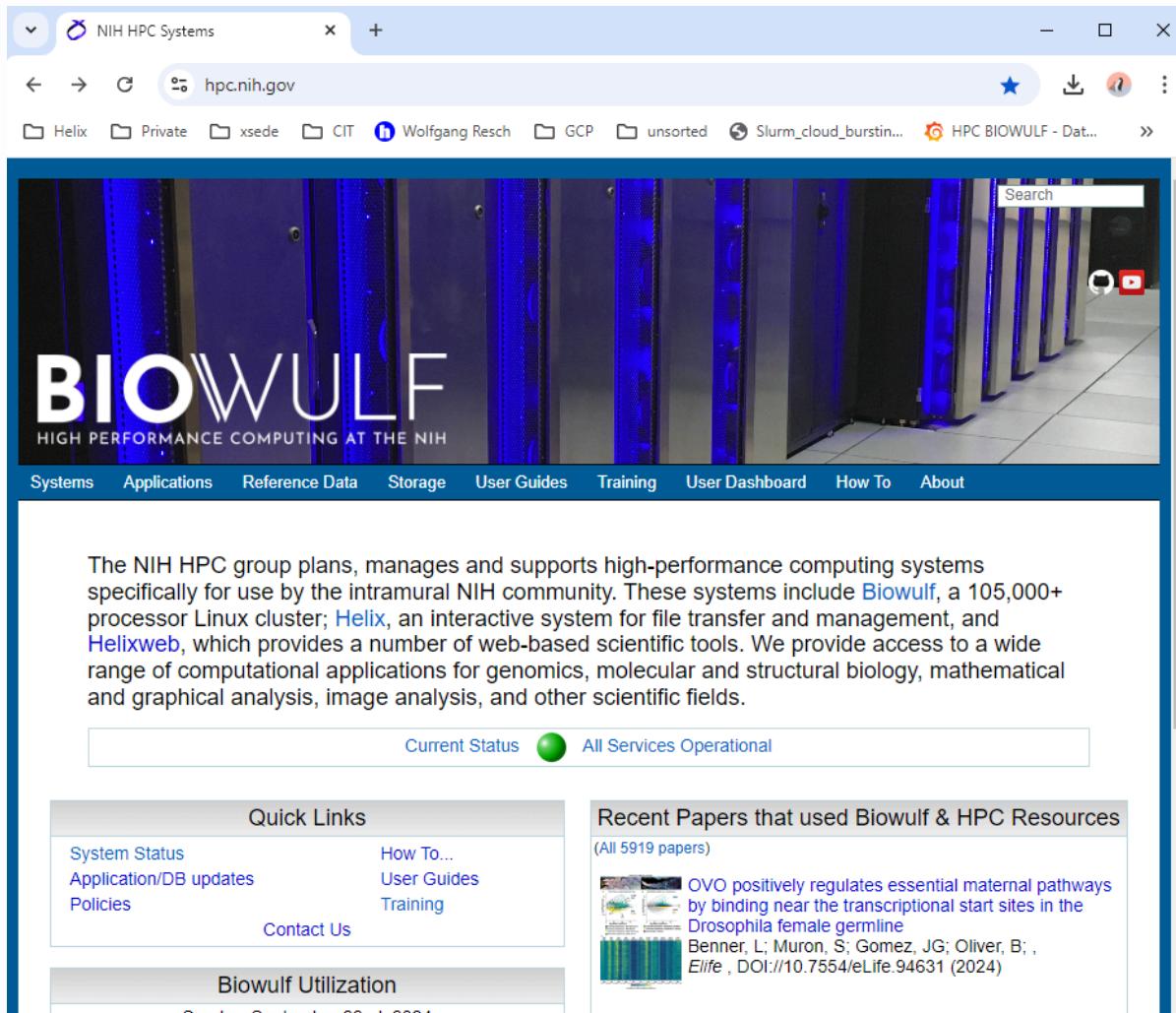


<https://hpc.nih.gov/docs/accounts.html>

Flat fee of \$40 / user / month

Accelerate research





The screenshot shows a web browser window for the NIH HPC Systems website at hpc.nih.gov. The page features a large banner image of server racks with the text "BIOWULF HIGH PERFORMANCE COMPUTING AT THE NIH". Below the banner is a navigation bar with links: Systems, Applications, Reference Data, Storage, User Guides, Training, User Dashboard, How To, and About. The main content area contains a paragraph about the NIH HPC group's mission to support high-performance computing systems for the intramural NIH community, mentioning BioWulf, Helix, and Helixweb. A "Current Status" box indicates "All Services Operational". Below this are "Quick Links" for System Status, Application/DB updates, Policies, How To..., User Guides, Training, and Contact Us. A "BioWulf Utilization" section shows the date as Sunday, September 23rd, 2024. To the right, a "Recent Papers that used BioWulf & HPC Resources" section lists a paper: "OVO positively regulates essential maternal pathways by binding near the transcriptional start sites in the Drosophila female germline" by Benner, L; Muron, S; Gomez, JG; Oliver, B; , *Elife* , DOI://10.7554/eLife.94631 (2024).

- Acknowledge
- Check the box in NIDB
- Let us know directly

Staff - staff@hpc.nih.gov



Steve Bailey



Tim Miller



David Hoover, Ph.D.



Helen Ashdown



Gennady Denisov,
Ph.D.

Picture
unavailable

Afif Elghraoui



Ali Erfani



Andrew Fant, Ph.D.



Beide (Peter) Fu,
Ph.D.



Jonathan Goodson,
Ph.D.



Patsy Jones

Picture
unavailable

Charles Lehr



Jean Mao, Ph.D.



Nitish Narula



David O'Brien

Picture
unavailable



Ifeanyi Okoye



Mark Patkus

Wolfgang Resch,
Ph.D.



Antonio Ulloa, Ph.D.

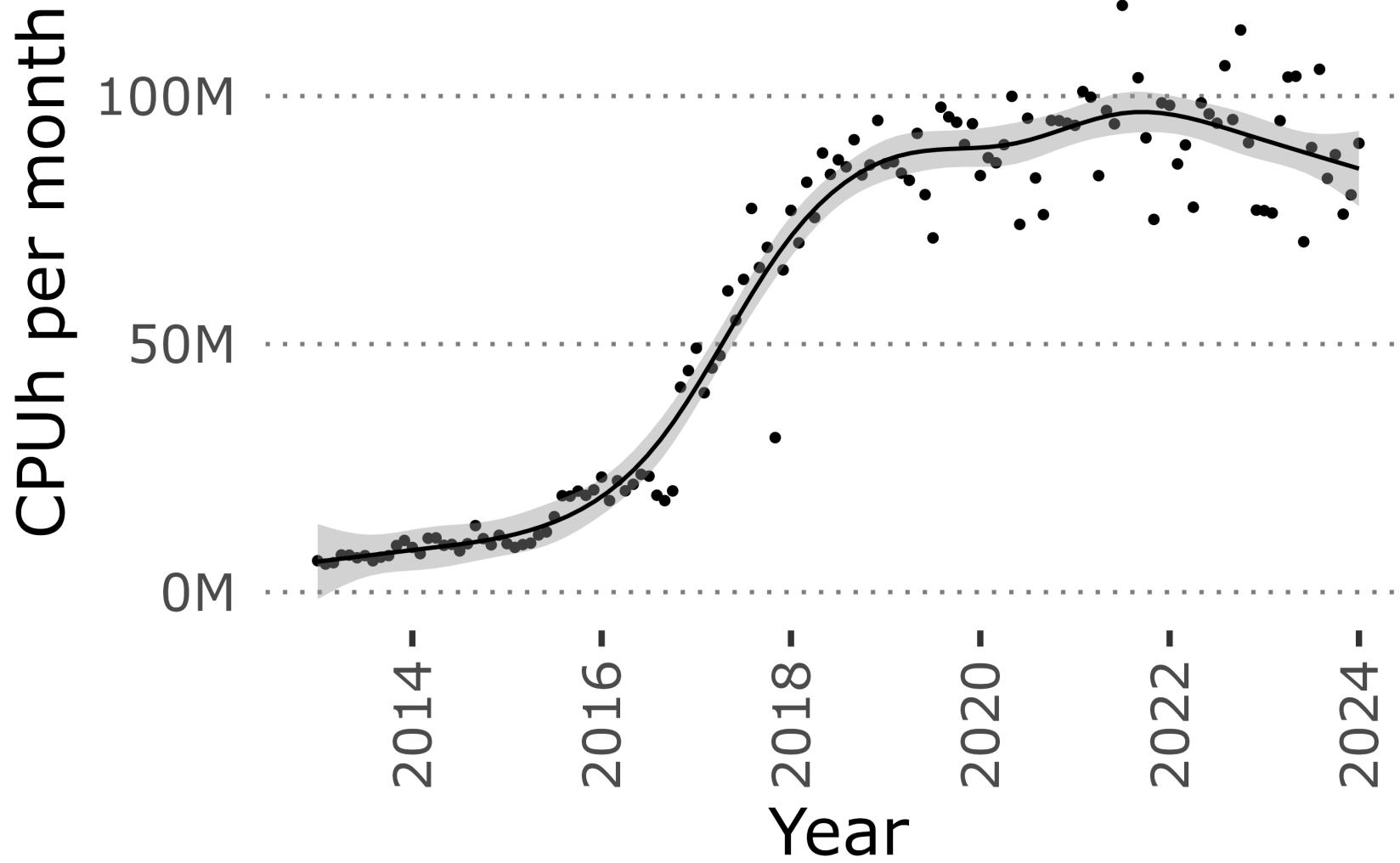


Renbin Yang, Ph.D.



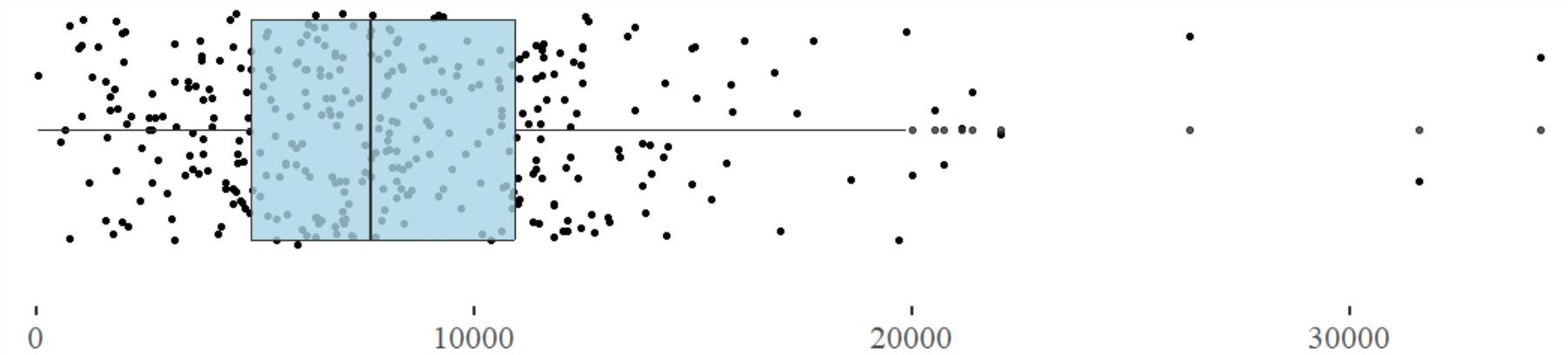
Qi Yu, Ph.D.

HPC@NIH in the future

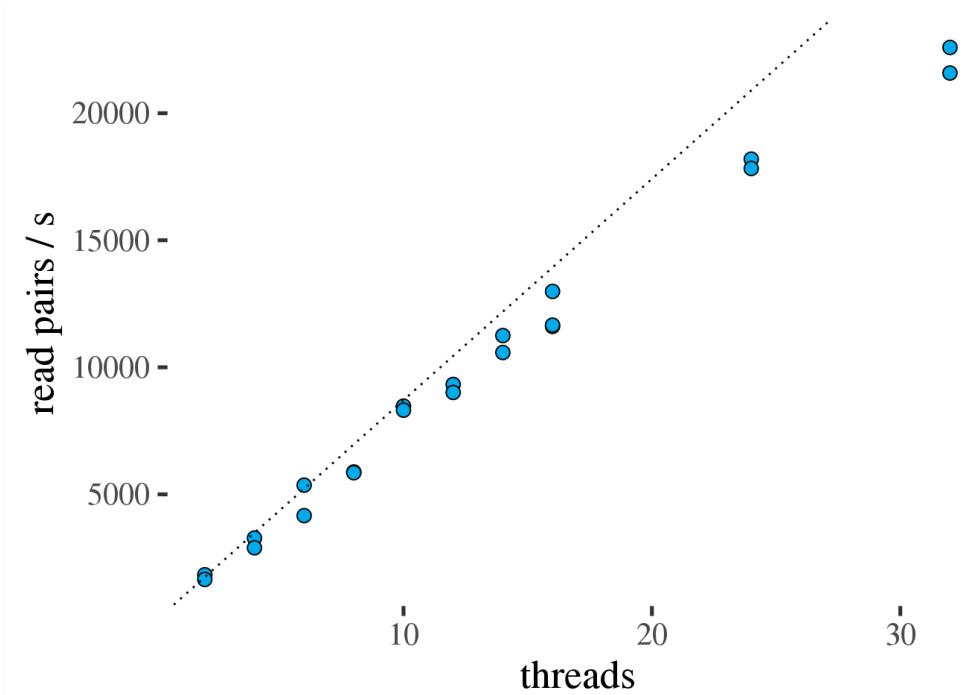
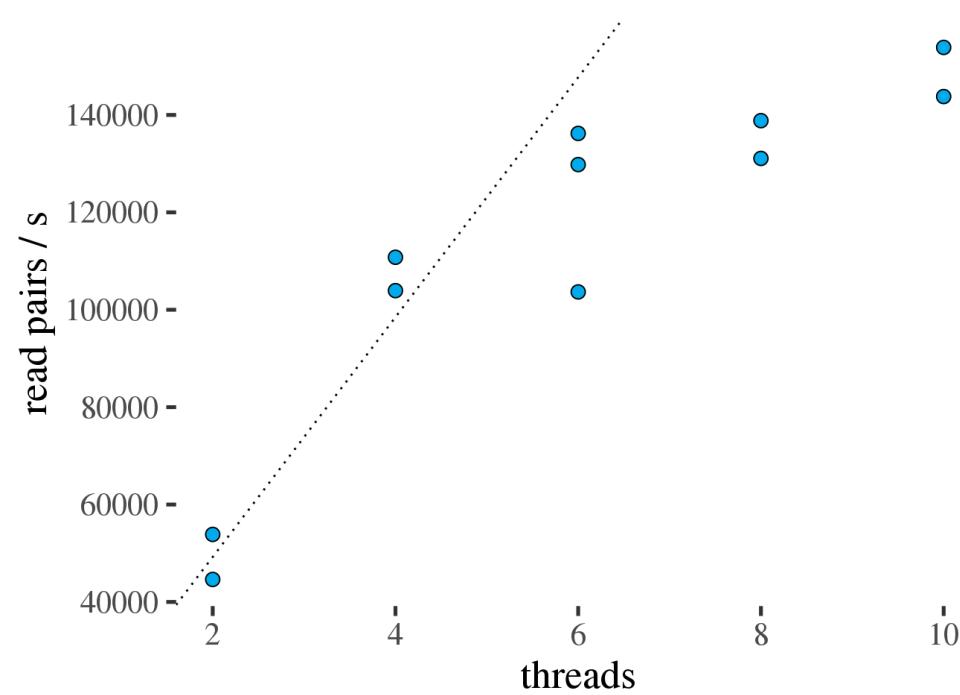


Best practices

Test before running large production jobs



Scaling matters

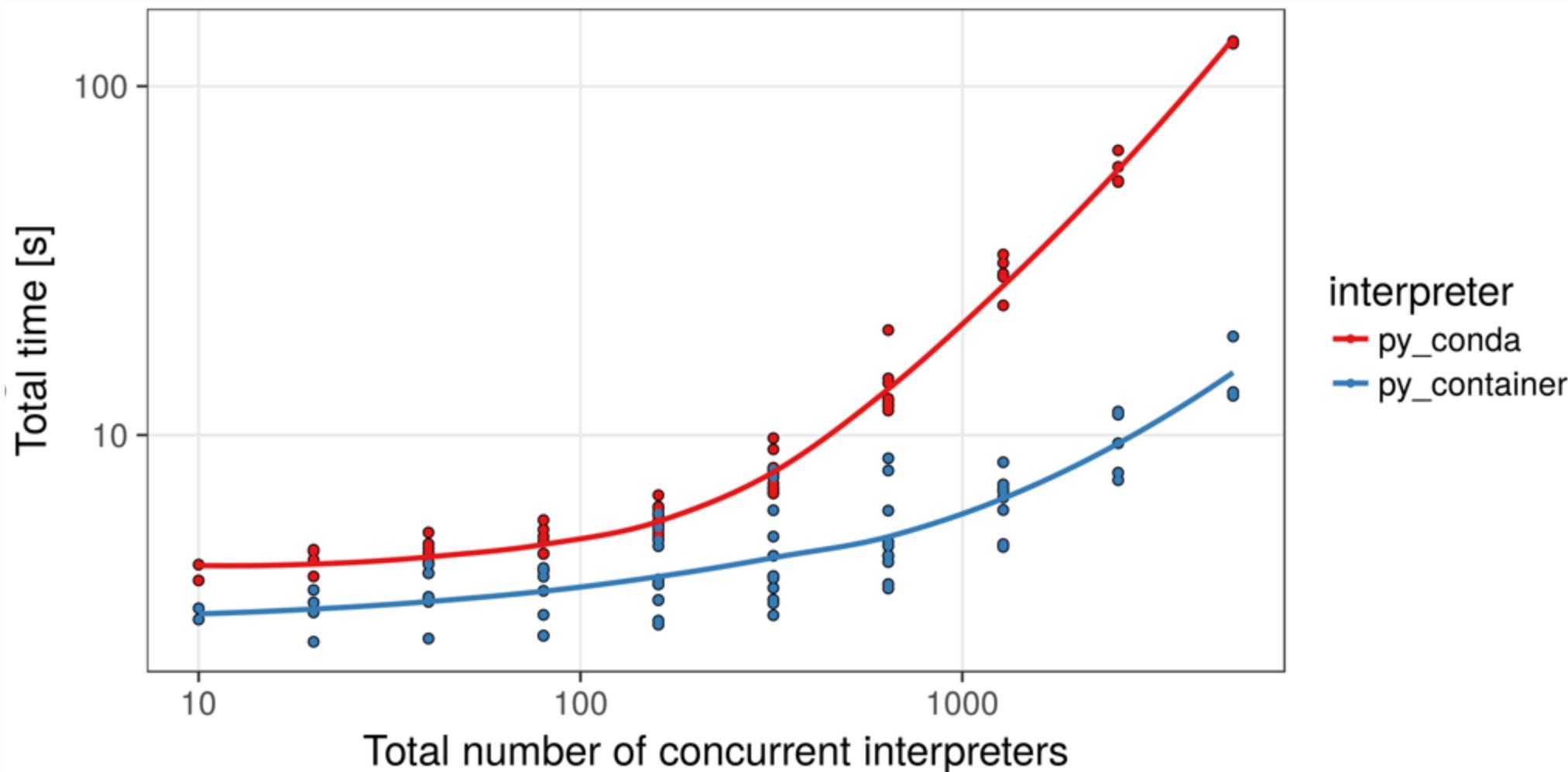


Left: fastp Right: bwa

Be nice to the **file systems**
and use `/lscratch`
and pipes

Interpreter startup is
expensive

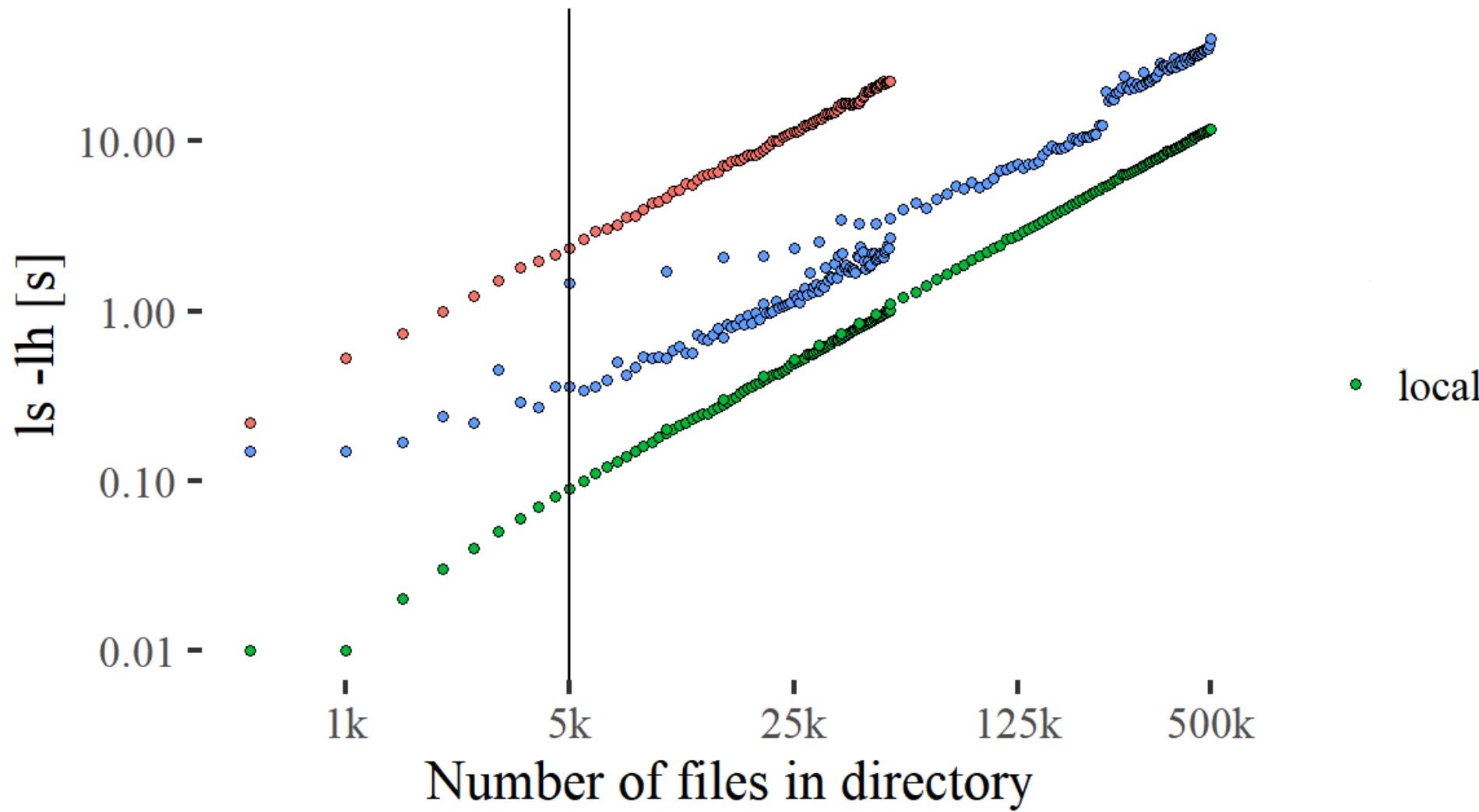
For example: concurrent python processes



Make sure you have **space**
in /data for your jobs' output

Keep < 5000 files in a
directory

listing files in large directories



and listing files is the least of the problems

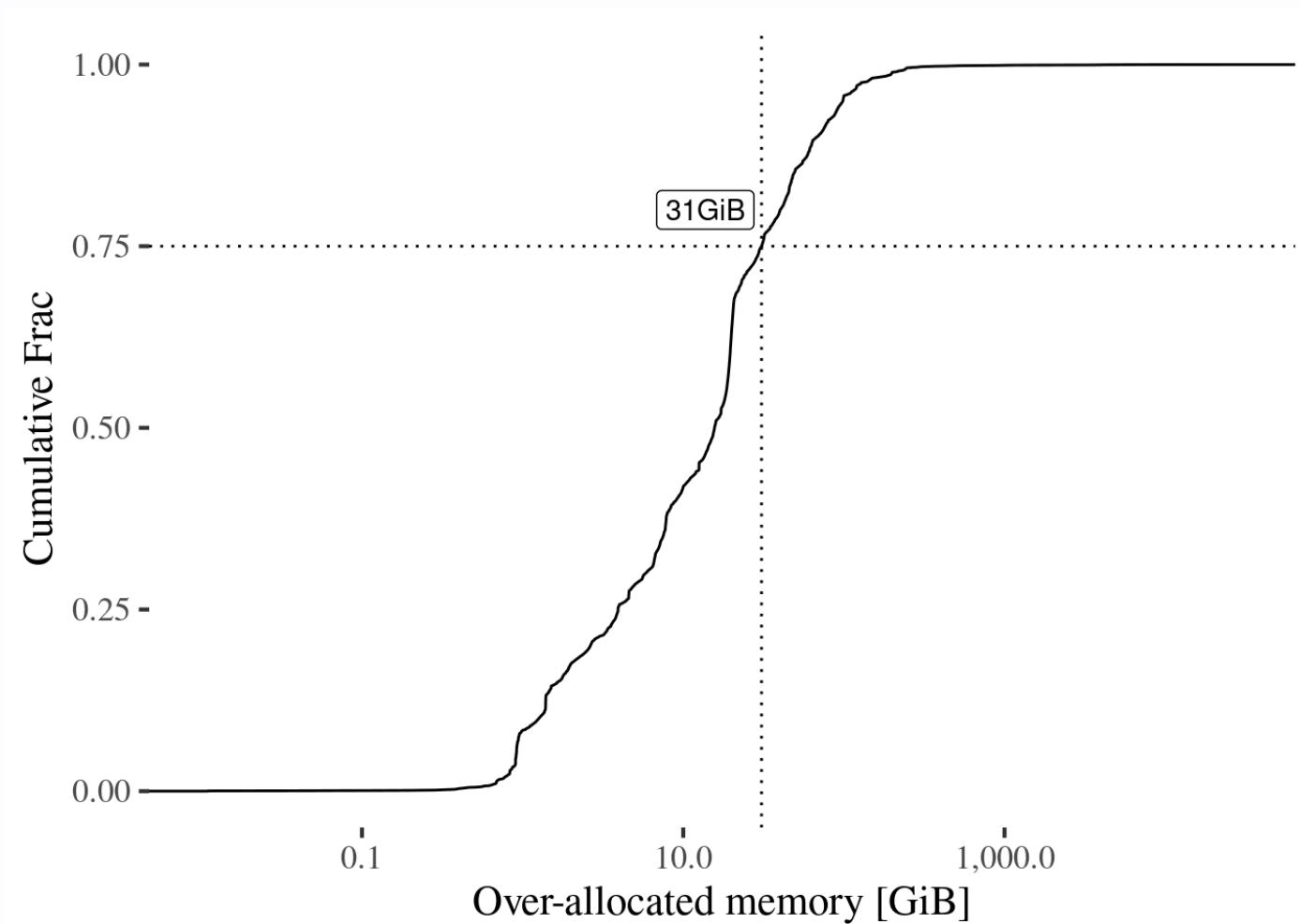
Biowulf storage is not for
archiving(*)

(*) except the object store,
kind of

Backup your data

Have **good resource**
estimates for your jobs

Completed jobs in 2023



Don't **abuse** newwall

Use **job arrays / swarms**
when you can

Try to run jobs with
walltime > 15 min

Keep startup files **simple**
(no conda please)

Please **read your email**

Be **detailed** and **specific**
when you file an issue (and
use text
instead of screenshots)

Please acknowledge
biowulf in your
publications

Talk to us

staff@hpc.nih.gov