# High Performance Computing (platforms) Security and Operations at Pitt

Kim F. Wong
*Center for Research Computing*
*SAC-PA, June 22, 2017*

# Our service

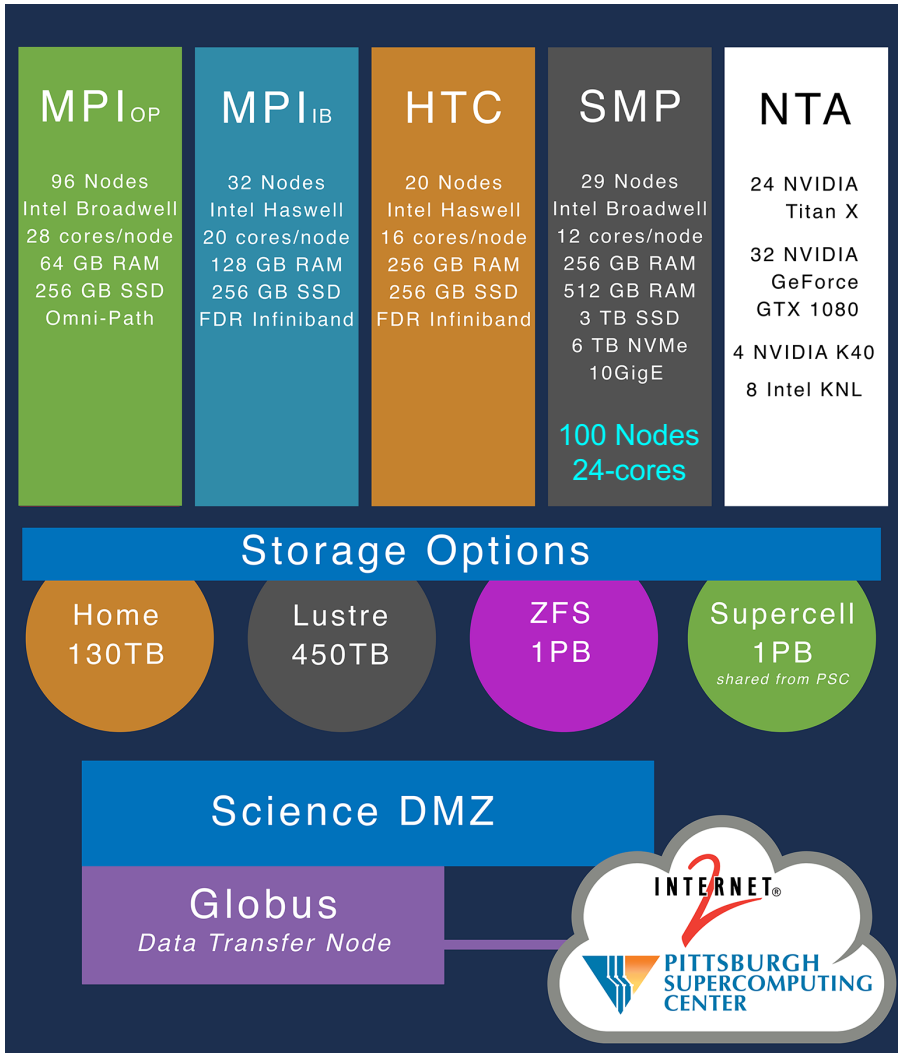
- The mission of the Center for Research Computing is to increase the research productivity of Pitt faculty through the use of advanced computing. We fulfill this mission by

    1. providing our community access to cutting-edge computer hardware and software for enabling transformative research,
    2. providing our community training workshops to educate users how to utilize the computing resources effectively,
    3. providing extended personalized consultation for improving researchers' computational workflow and code performance through selection of better algorithms, parallelization techniques, improved use of input-output strategies, etc.

# Our team


- *From left to right.* (a) Ketan Maheshwari: HPC, GPGPUs, Scientific Computing. (b) Karl Johnson: Chem Eng, SaM Co-Director (c) Ken Jordan: Chemistry, SaM Co-Director (d) Fangping Mu: Bioinformatics, Computational Biology, Computational Genomics (e) Kim Wong: Biological Simulation, Agent-based Modeling, Physics-based Modeling (f) Wendy Janocha, Administrative Coordinator (g) Ralph Roskies, Associate Vice Provost for Research Computing (h) Barry Moore II: Quantum Chemistry, HPC (i) Shervin Sammak: Turbulent Combustion, Fluid Dynamics (j) Thomas Troyan: CS Undergraduate, Web Developer.

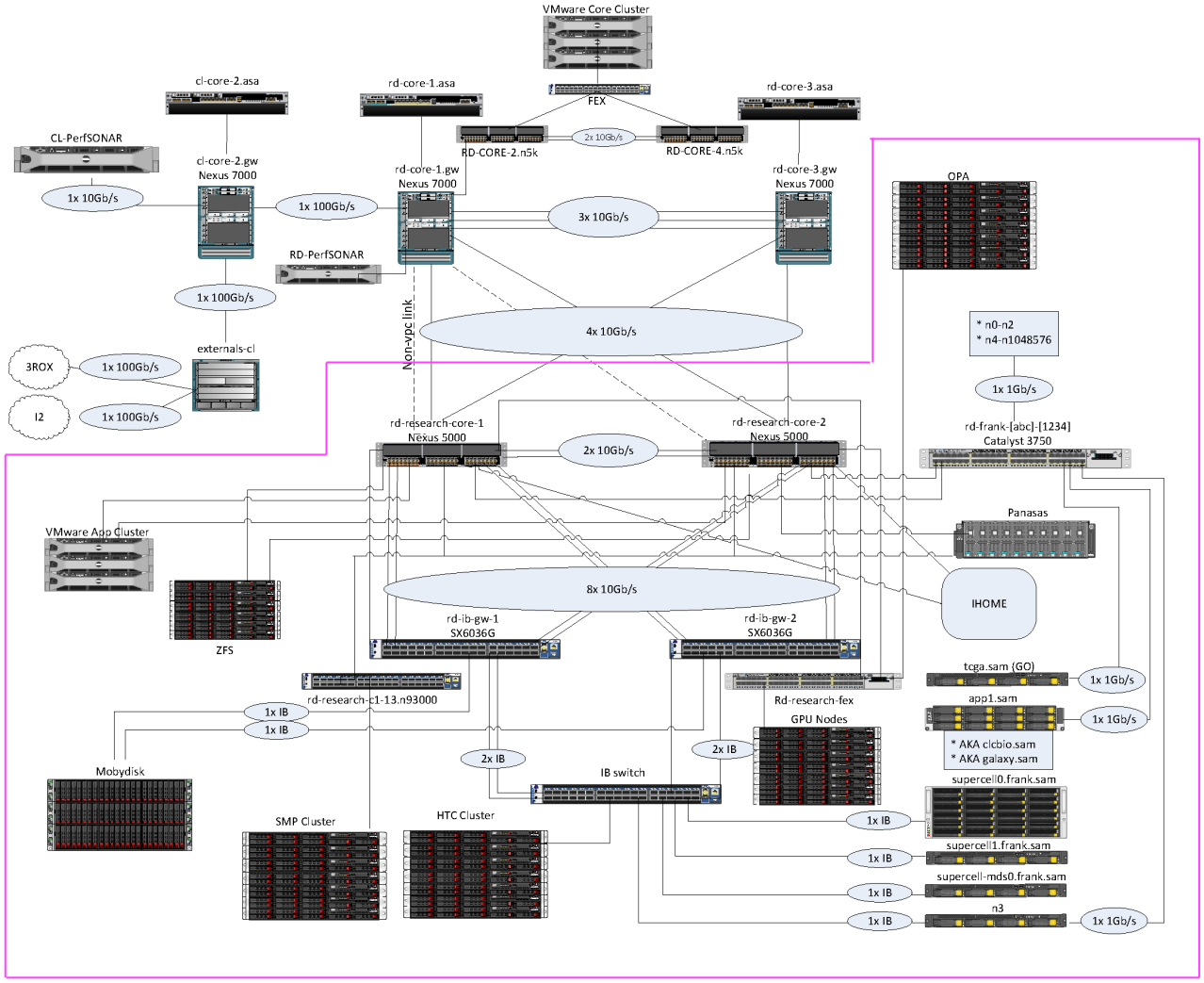
# Our hardware resources




- **MPI-OP & MPI-IB:** for applications that are parallelized using the distributed computing Message Passing Interface framework.
- **HTC:** for high throughput computing workflows such as next-generation sequencing assembly and data-intensive analytics.
- **SMP:** for serial jobs and programs that are parallelized using the shared memory framework.
- **NTA**: for applications written to take advantage of non-traditional architectures such as NVIDIA GPUs and Intel Knights Landing Multi-core CPUs.

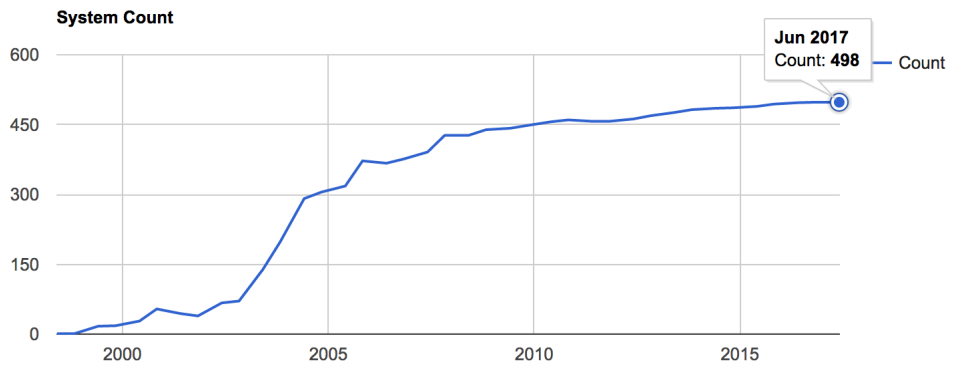
# What does it look like physically?

# What does it look like topologically?

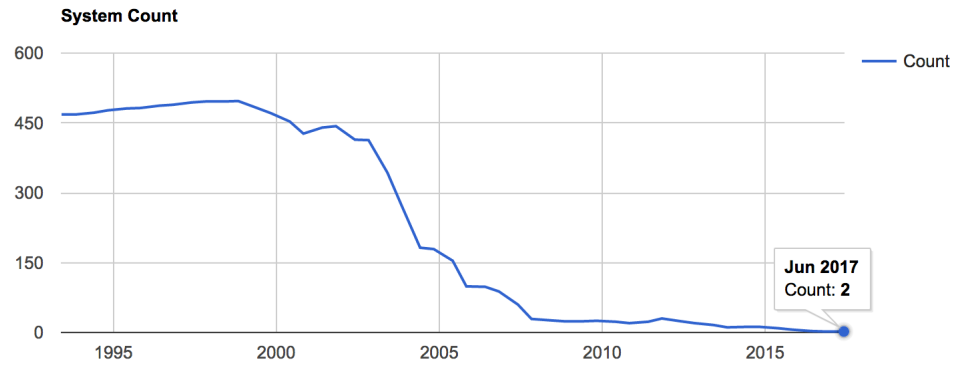


# What is our HPC platform?

OPERATING SYSTEM FAMILY / LINUX



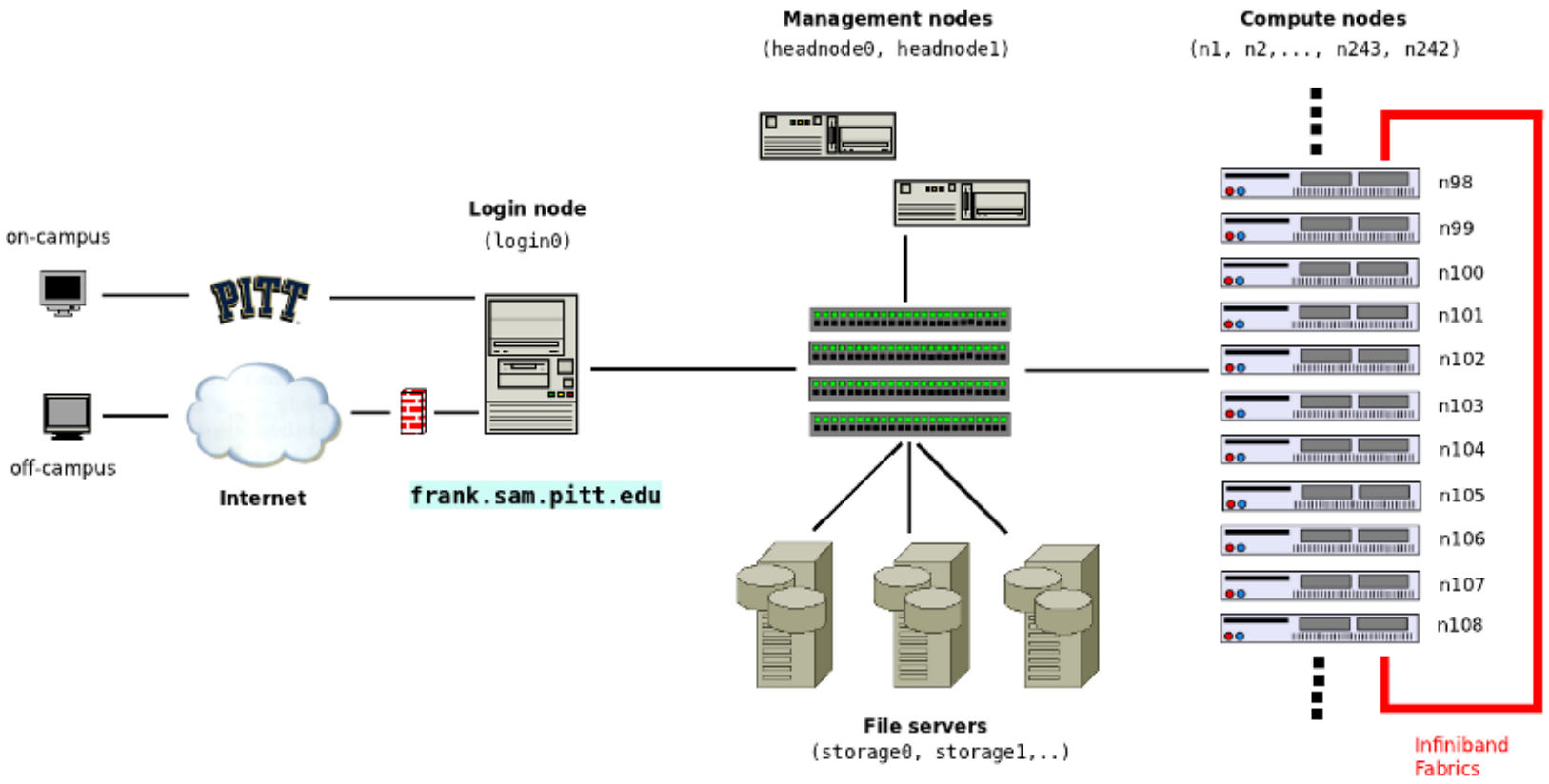

OPERATING SYSTEM FAMILY / UNIX




www.top500.org

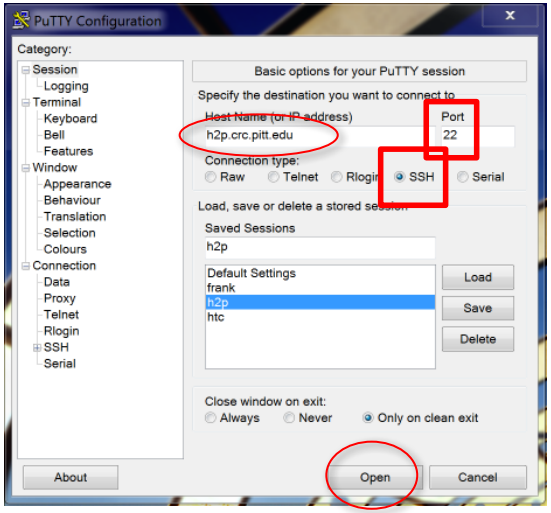
- CRC Clusters run a combination of RHEL 7 and RHEL 6
- Previously we were running CentOS 6

# What is the entry point to the resources?



# Accessing the HPC cluster

Access to H2P is enabled via a secure shell (SSH) connection to the cluster. If off-campus, make sure you have a VPN session open.



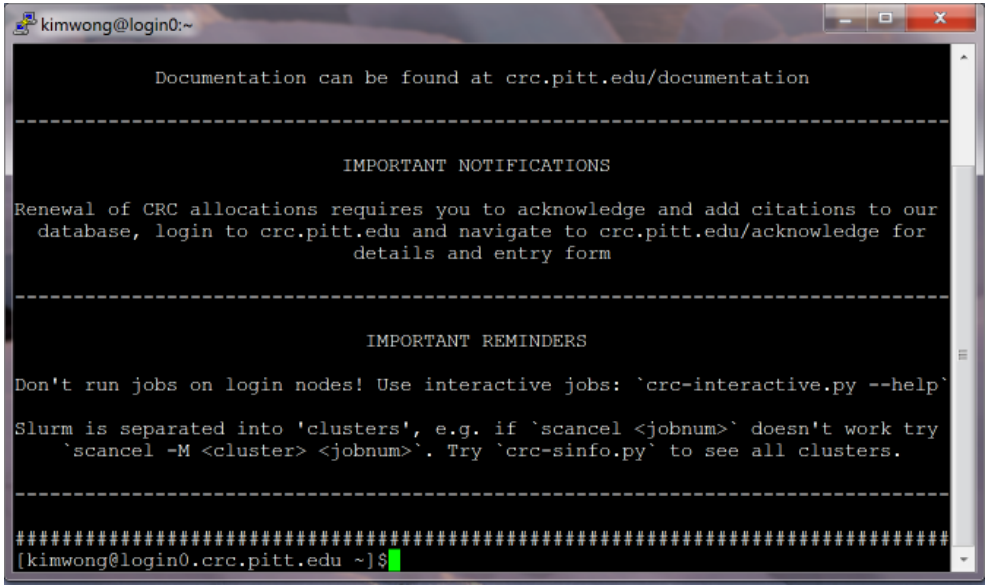

A SSH client called PuTTY is available for Windows

Specify these connection properties:

- Hostname: h2p.crc.pitt.edu
- Port: 22
- Connection type: SSH

Clicking the Open button will *open a SSH terminal*

- login as: <Pitt Username>
- password: <my.pitt password>



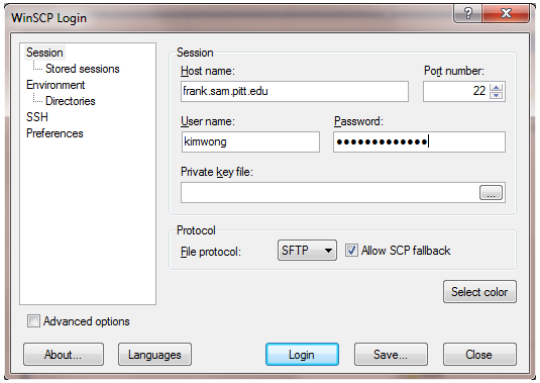

**Linux & Mac Users:** type `ssh -X <username>@h2p.crc.pitt.edu` within a terminal

# Transferring files to Frank (Windows)?

If transferring from off-campus, a VPN session is required.

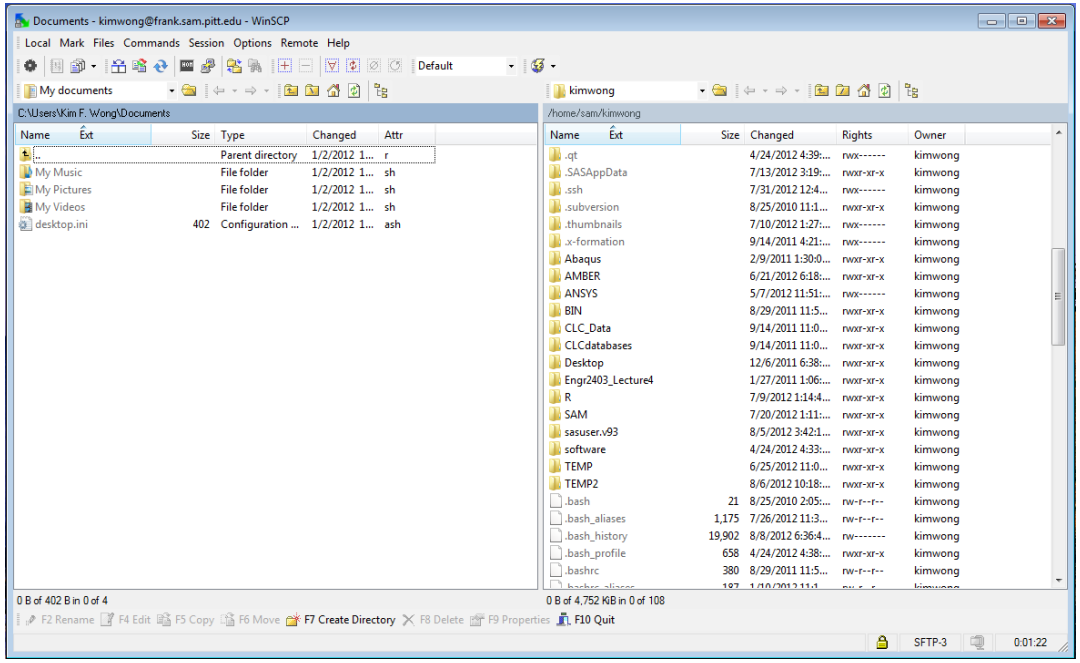
For Windows, use WinSCP http://sourceforge.net/projects/winscp/.
Login in to Frank using your Pitt credentials.

**1**




**2**

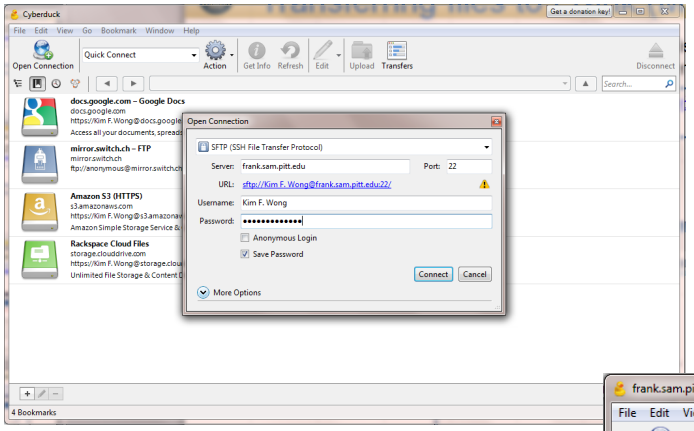
drag and drop
between panels



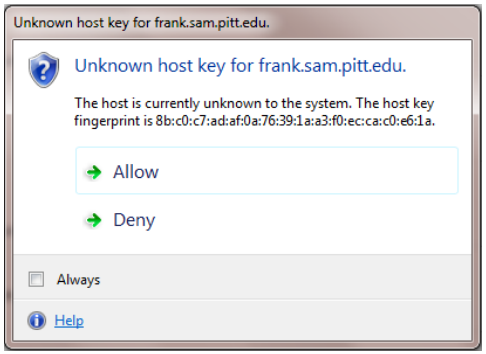
# Transferring files to Frank (Mac)?

For Macs, I have heard that Cyberduck works well: https://cyberduck.io/. Select SFTP (SSH File Transfer Protocol).

**1**

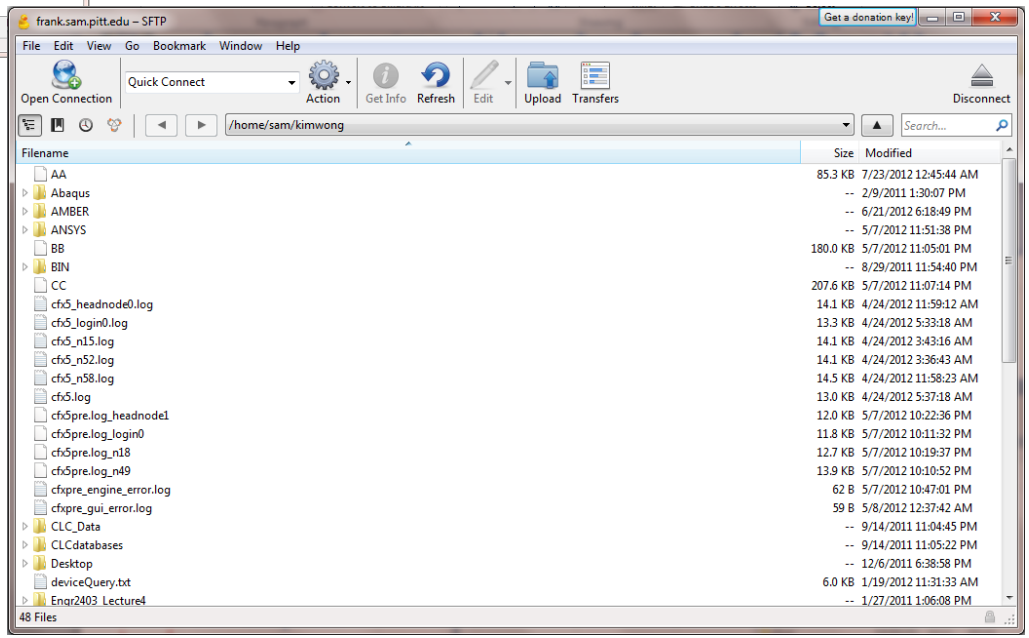



**2**




**3**

After authentication, a new window shows up. Drag and drop between that window and your local desktop/folders.



# I am not only going to tell you about the security operations within our HPC platform, I'm going to demonstrate it.

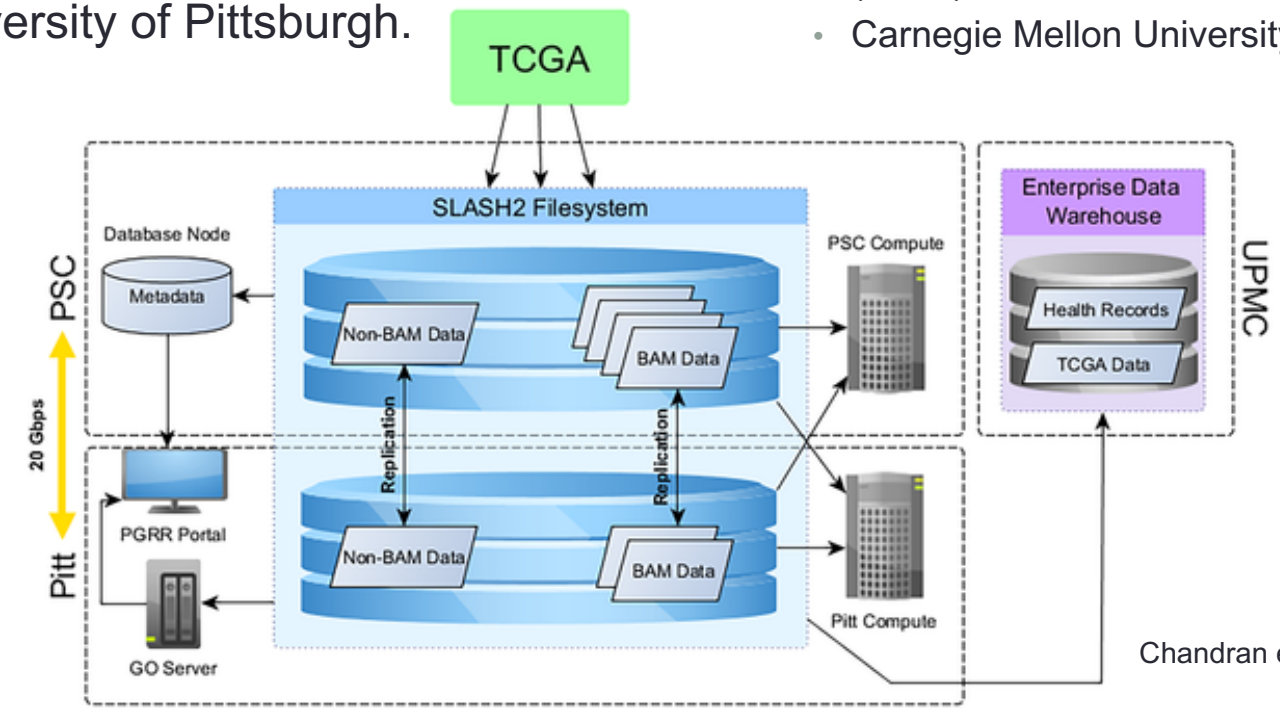
- Cluster: h2p, htc, mpi, frank
- Filesystem security
- Access control list (ACLs)
- Exploiting Linux groups to manage access to software and data repository
- setfacl -Rm u:gnowmik:rX,d:u:gnowmik:rX /ihome/sam/kimwong
  - R is recursive
  - d is default
  - m is needed to add/modify rules
  - rx are read and execute permissions

Hands-on demonstration via SSH to clusters.

# PGRR: A case study in complexity

- The Pittsburgh Genome Resource Repository (PGRR) is an institution-wide HPC infrastructure enabling controlled access to The Cancer Genome Atlas (TCGA) data for investigators named on a common Data Use Agreement through the University of Pittsburgh.

- Collaboration members:
  - Pittsburgh Supercomputing Center (PSC)
  - Center for Simulation and Modeling (SaM)
  - Institute for Personalized Medicine (IPM)
  - Department of Biomedical Informatics (DBMI)
  - Cancer Bioinformatics Service (CBS) of the University of Pittsburgh Cancer Institute (UPCI)
  - Carnegie Mellon University (CMU)




Chandran et al. *PLoS ONE*. 2016.

# Many facets to securing the PGRR data

- Physical security of site hosting data: access to data center is restricted to allowed personnel and logged upon entry/departure.
- Network security of hosting site: firewall controls and host-based ACLs. SSH access to login nodes. Centralized password authentication via the University Active Directory. Passwords age out every 180 days and complexity of passwords are checked
- Linux filesystem enforcement: group permissions and ACLs for sharing data
- OS Updates: scans and patching of kernel vulnerabilities
- User Database: Rigorous audit of user list
- User Education: Rigorous user training on best practices

# Security challenges on the horizon

- Computing-enabled research is becoming more collaborative
  - Need to share data and resources among local groups
  - Need to accommodate external collaborators
- Science gateways and web portals as tools for lowering the barrier to access HPC resources
  - Need to delineate who has access and the scope
  - Is this delineation consistent with funding agency requirements
- Distributed data repositories

Thank you!