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,
  - 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?



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



www.top500.org

#### What is the entry point to the resources?



#### Accessing the HPC cluster

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

A SSH client called <u>PuTTY</u> 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

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



🛃 kimwong@login0:~	
Documentation can be found at crc.pitt.edu/documentation	IMPORTANT NOTIFICATIONS ions requires you to acknowledge and add citations to our rc.pitt.edu and navigate to crc.pitt.edu/acknowledge for details and entry form IMPORTANT REMINDERS n nodes! Use interactive jobs: `crc-interactive.pyhelp` o 'clusters', e.g. if `scancel <jobnum>` doesn't work try</jobnum>
	ł
database, login to crc.pitt.edu and navigate to crc.pitt.edu/acknowledge for	l
	I
IMPORTANT REMINDERS	
Don't run jobs on login nodes! Use interactive jobs: `crc-interactive.pyhelp`	ł
<pre>Slurm is separated into 'clusters', e.g. if `scancel <jobnum>` doesn't work try `scancel -M <cluster> <jobnum>`. Try `crc-sinfo.py` to see all clusters.</jobnum></cluster></jobnum></pre>	
[kimwong@login0.crc.pitt.edu ~]\$	
	_

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 <u>http://sourceforge.net/projects/winscp/</u>. Login in to Frank using your Pitt credentials.

Session Stored sessions Environment Directories SSH Preferences	Session <u>H</u> ost name: frank.sam.pitt.edu	Port number:
		word:
	Private key file:	
	Elle protocol: SFTP V	Allow SCP fallback
Advanced options		



	ommands Session Options Ren							
• 🗉 • 🟦	🗳 🕹 🔤 🧬 😤 🐘 🖽		• 🚳 •					
My documents	• 🛅   🔶 + 🔿 + 🚹	ն 🔂 🖄	🖟 kimwong	• 🔤 🛛	(+ + ⇒ +   <b>1</b>	🗖 🚮 🚺	te l	
C:\Users\Kim F. Wong\Do	cuments		/home/sam/kimwong					
Name Êxt	Size Type	Changed Attr	Name Êxt	Size	Changed	Rights	Owner	
<b>i</b>	Parent directory	1/2/2012 1 r	J.qt		4/24/2012 4:39:	rwx	kimwong	
My Music	File folder	1/2/2012 1 sh	SASAppData		7/13/2012 3:19:	rwxr-xr-x	kimwong	
My Pictures	File folder	1/2/2012 1 sh	🏭 .ssh		7/31/2012 12:4	rwx	kimwong	
My Videos	File folder	1/2/2012 1 sh	July .subversion		8/25/2010 11:1	rwxr-xr-x	kimwong	
👔 desktop.ini	402 Configuration	1/2/2012 1 ash	🔐 .thumbnails		7/10/2012 1:27:	rwx	kimwong	
			🔐 .x-formation		9/14/2011 4:21:	rwx	kimwong	
			🍑 Abaqus		2/9/2011 1:30:0	rwxr-xr-x	kimwong	
			AMBER		6/21/2012 6:18:	rwxr-xr-x	kimwong	
			ANSYS		5/7/2012 11:51:	rwx	kimwong	
			III BIN		8/29/2011 11:5	rwxr-xr-x	kimwong	
			🔒 CLC_Data		9/14/2011 11:0	rwxr-xr-x	kimwong	
			LCCdatabases		9/14/2011 11:0	rwxr-xr-x	kimwong	
			🍶 Desktop		12/6/2011 6:38:	rwxr-xr-x	kimwong	
			Engr2403_Lecture4		1/27/2011 1:06:	rwxr-xr-x	kimwong	
			🔒 R		7/9/2012 1:14:4	rwxr-xr-x	kimwong	
			SAM		7/20/2012 1:11:	rwxr-xr-x	kimwong	
			sasuser.v93		8/5/2012 3:42:1	rwxr-xr-x	kimwong	
			software		4/24/2012 4:33:	rwxr-xr-x	kimwong	
			J TEMP		6/25/2012 11:0	rwxr-xr-x	kimwong	
			JEMP2		8/6/2012 10:18:	rwxr-xr-x	kimwong	
			.bash	21	8/25/2010 2:05:	rw-rr	kimwong	
			bash_aliases	1,175	7/26/2012 11:3	rw-rr	kimwong	
			.bash_history	19,902	8/8/2012 6:36:4	rw	kimwong	
			.bash_profile	658	4/24/2012 4:38:	rwxr-xr-x	kimwong	
			.bashrc	380	8/29/2011 11:5	rw-rr	kimwong	
			D bacher alianer	107	1/10/2012 11:1		bimuana	
B of 402 B in 0 of 4			0 B of 4,752 KiB in 0 of 108					

#### Transferring files to Frank (Mac)?

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



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 institutionwide 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)



#### Many facets to securing the PGRR data

- <u>Physical security of site hosting data</u>: access to data center is restricted to allowed personnel and logged upon entry/departure.
- <u>Network security of hosting site</u>: firewall controls and hostbased 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
- <u>Linux filesystem enforcement</u>: group permissions and ACLs for sharing data
- <u>OS Updates</u>: scans and patching of kernel vulnerabilities
- User Database: Rigorous audit of user list
- <u>User Education</u>: 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!