Research Computing Support at the University of Utah: An Overview of CHPC

Anita Orendt
Assistant Director
Research Consulting & Faculty Engagement
anita.orendt@utah.edu

CHPC's Mission

In addition to deploying and operating high performance computational resources and providing advanced user support and training, CHPC serves as an expert team to broadly **support** the increasingly **diverse research computing needs** on campus. These needs include support for big data, big data movement, data analytics, security, virtual machines, Windows science application servers, protected environments for data mining and analysis of protected health information, and advanced networking.

- *CHPC's mission:* ***Support Computational Research!***
- CHPC can help if:
  - You need parallel processing
  - You need access to a single high-powered computer
  - You need the ability to run many individual jobs simultaneously
  - You have a large amount of data to store and/or process
  - You need an application you don't have on your computer
  - Your data is IRB-governed PHI
  - You have other computing needs your local resources cannot meet

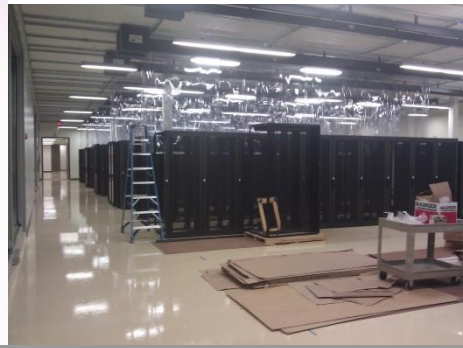
Sources of Useful Information

- Getting Started Guide
  - https://www.chpc.utah.edu/documentation/gettingstarted.php
- CHPC policies
  - https://www.chpc.utah.edu/documentation/policies/index.php
- Cluster Usage Guides
  - https://www.chpc.utah.edu/documentation/guides/index.php
- Application Documentation
  - https://www.chpc.utah.edu/documentation/software/index.php
- Programming Guide
  - https://www.chpc.utah.edu/documentation/ProgrammingGuide.php
- How to Videos
  - https://www.chpc.utah.edu/documentation/videos/index.php

Downtown Data Center

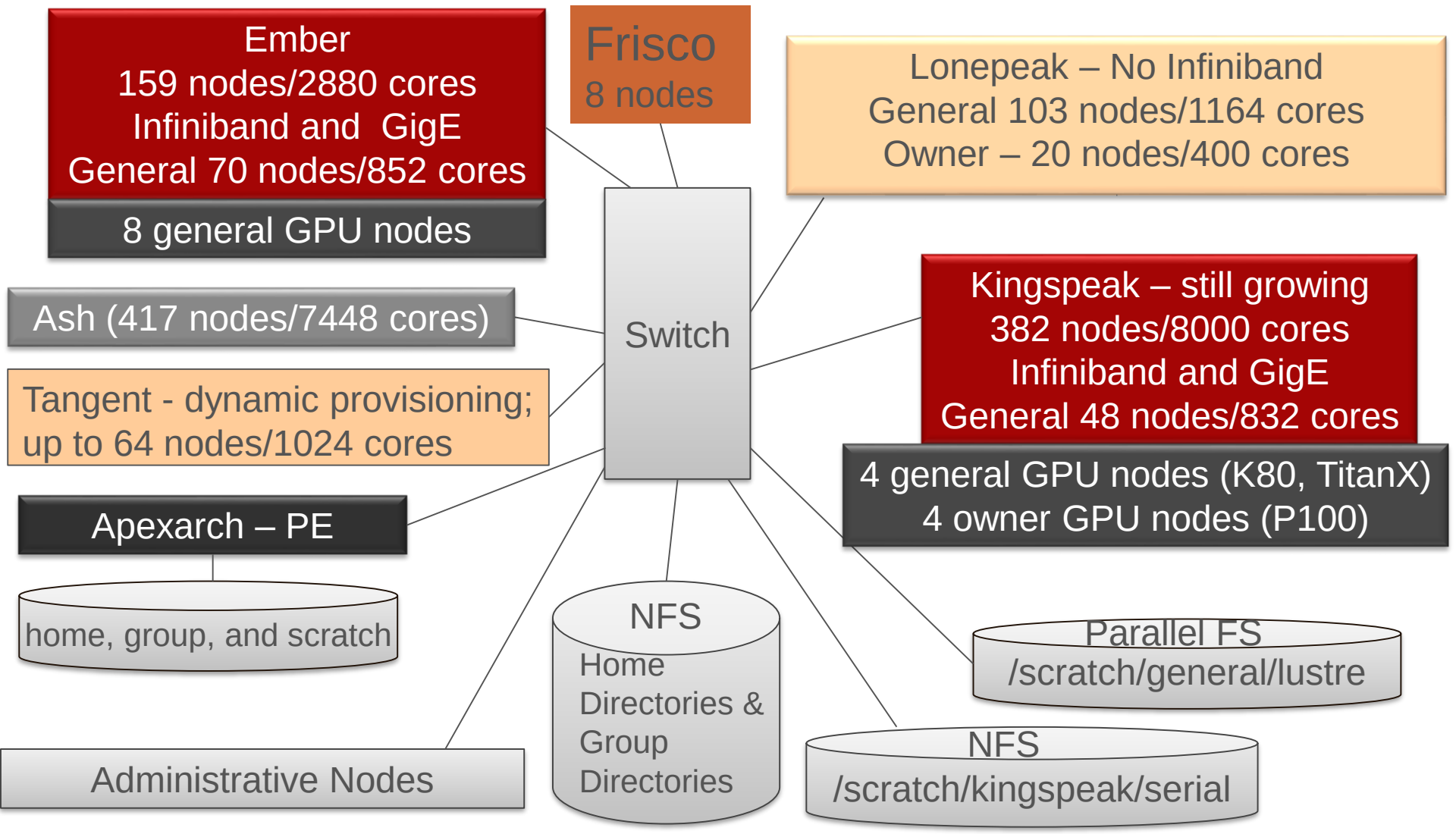
- Came online Spring 2012
- CHPC fully at DDC Spring 2013
- Shared with enterprise (academic/hospital) groups
- 92 racks and 1.2MW of power with upgrade path to add capacity for research computing
- Metro optical ring connecting campus, data center, & internet2
- 24/7/365 facility



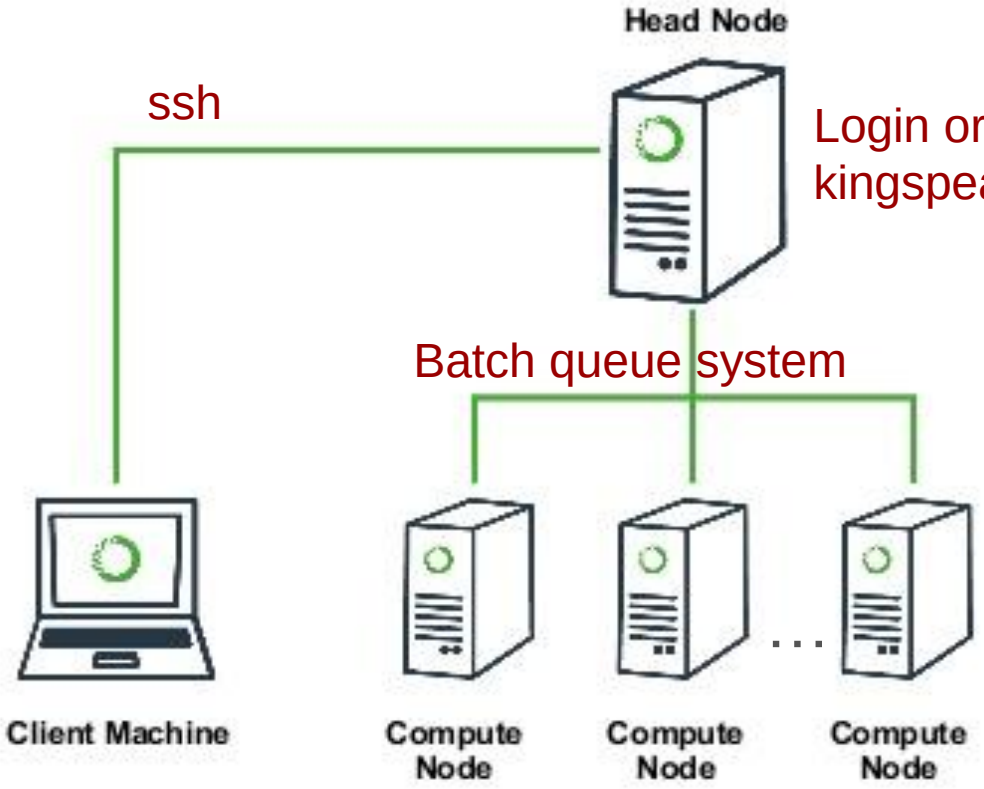
CHPC Resources & Services

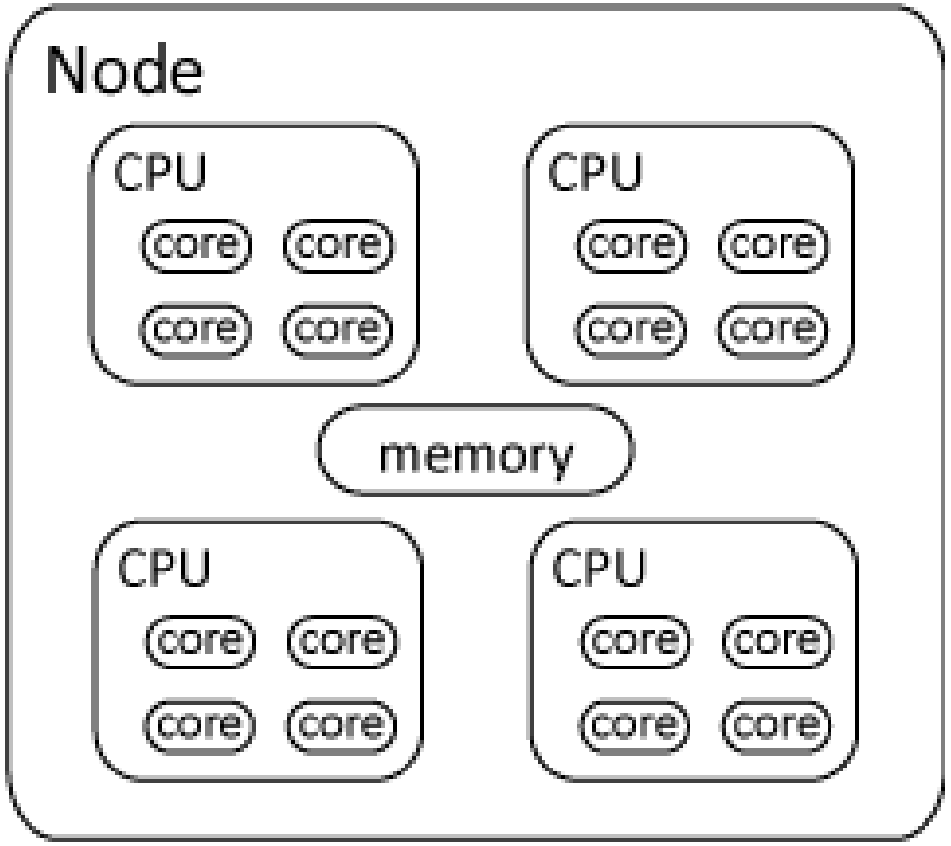
- ***Computational Clusters*** – Kingspeak, Lonepeak, Ember, Ash, Tangent
- ***Storage*** – home, group, and scratch storage along with tape backup and archive storage options
- ***Windows Servers*** – mainly statistics usage and windows only applications
- ***Virtual Machines*** – for needs not met with cluster and windows server
- ***Protected Environment*** –computational cluster Apexarch, storage, VMs, and Windows Server
- ***Networking Support*** – support compute environment; work with researchers on data movement etc
- ***User Support*** – assistance with use of resources; installation of applications; training sessions
- ***Desktop Support*** – for several departments


Ember
159 nodes/2880 cores
Infiniband and GigE
General 70 nodes/852 cores
8 general GPU nodes
Frisco
8 nodes
Lonepeak – No Infiniband
General 103 nodes/1164 cores
Owner – 20 nodes/400 cores
Ash (417 nodes/7448 cores)
Switch
Kingspeak – still growing
382 nodes/8000 cores
Infiniband and GigE
General 48 nodes/832 cores
Tangent - dynamic provisioning;
up to 64 nodes/1024 cores
4 general GPU nodes (K80, TitanX)
4 owner GPU nodes (P100)
Apexarch – PE
home, group, and scratch
NFS
Home Directories & Group Directories
Parallel FS
/scratch/general/lustre
Administrative Nodes
NFS
/scratch/kingspeak/serial

CONTINUUM ANALYTICS

Cluster Architecture Diagram




Node
CPU
core
core
core
core
CPU
core
core
core
core
memory
CPU
core
core
core
core
CPU
core
core
core
core

CHPC Clusters - Condominium Model

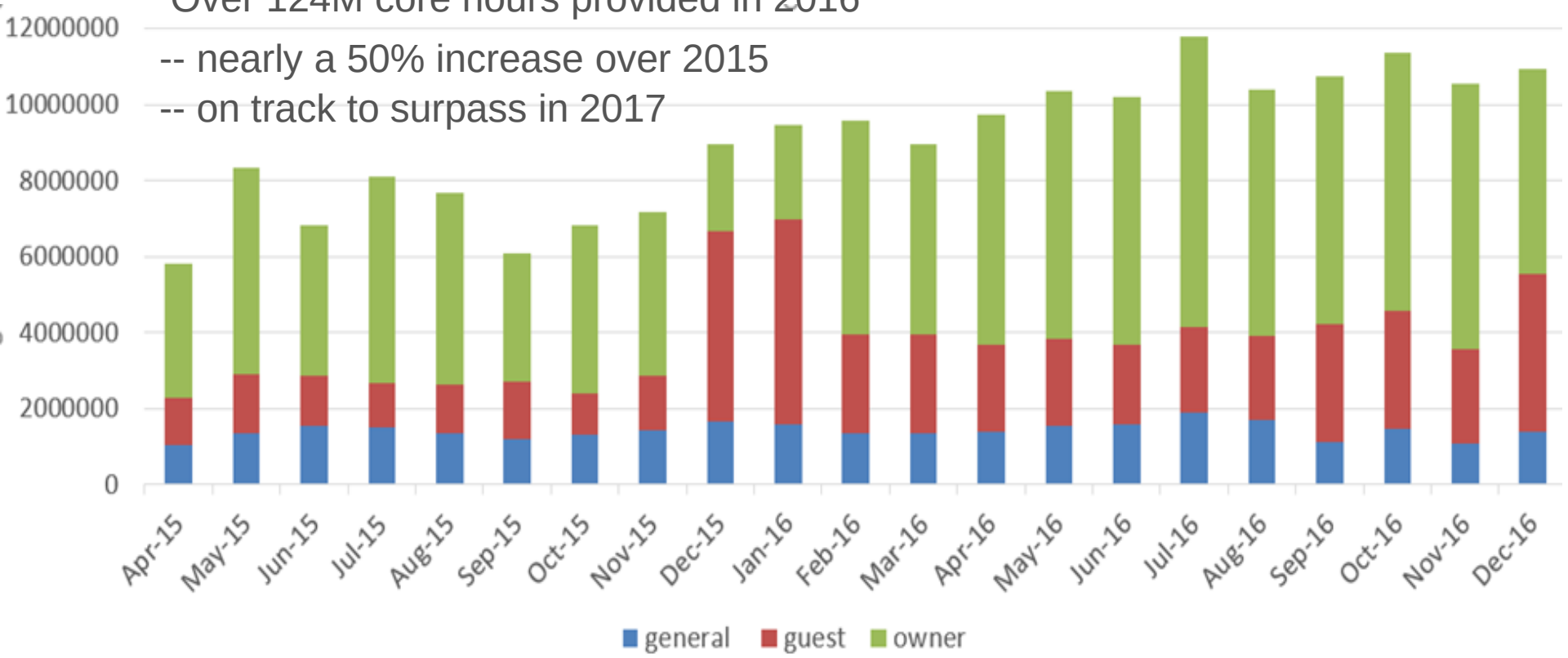
- General resources – No charge for usage
  - Allocation process for cycles on kingspeak and ember
  - Out of allocation – freecycle allowed (with preemption)
- Owner nodes – latest configuration
  - 28 core Intel Broadwell @ 2.4GHz cpu speed, 128GB RAM, 1TB local hard drive, 5 year warranty, IB connectivity @ ~$6800/node (New Intel Skylake – XeonSP coming next month)
  - Others allowed on as owner-guest when not in use (preemption)
- Lonepeak – no allocation and no preemption
- Tangent – dynamically provisioned cluster resource; no allocation and no preemption
- Ash – Owner cluster – has guest access as smithp-guest

Core Hour Usage

Over 124M core hours provided in 2016

-- nearly a 50% increase over 2015

-- on track to surpass in 2017



CHPC Provides Core Infrastructure

- Physical needs (racks, power, cooling)
- Core ethernet and IB fabric; IB to ethernet bridging
- Login/head/management nodes
- Scheduling, batch and allocation systems
- HPC scratch space
- Some application and licensing costs
- High speed data transfer nodes (DTNs)
- 'Science DMZ' network
- CHPC Staff

Storage Options

- *Home Directories* -- /uufs/chpc.utah.edu/common/home/<uNID>
  - Home directories 50GB, not backed up (some exceptions)
  - Groups can purchase 1TB max home directory/group
  - New compellent solution coming on line now
- *Group Level File Systems*
  - Group space @$150/TB; can get quarterly archives with purchase of tapes
- *Scratch File Systems*
  - For use by all users; scrubbed of files older than 60 days
  - 700TB Lustre Parallel file system (/scratch/general/lustre)
  - 175TB NFS mounted file system (/scratch/kingspeak/serial)
- *Disk Based Archive Storage – Now Available!*
- *Tape Backups – as discussed above*

File Systems

- Access speed based on connectivity
- Local disk fastest – local to each node; varies in size
  - /scratch/local
- Network mounted scratch file systems
  - /scratch/general/lustre
  - /scratch/kingspeak/serial
  - home directories and group spaces (***don't use for large i/o!***)

Remember NFS mounted spaces – including file systems of group spaces – are a shared resource!

To check the current status of the file systems (and clusters)

– www.chpc.utah.edu → Usage → Cluster Utilization Graphs

Protected Environment

- Dedicated protected resources for handling of data/projects with protected information
- Currently HIPAA, looking at FISMA & FERPA
- Also for projects with other types of sensitive data/restrictions
- Significant area of growth for CHPC
- Described in recent paper (Bradford *et al*.) *www.ncbi.nlm.nih.gov/pubmed/23911553*

Coming Soon

- New Cluster – Notchpeak
  - Early fall – Kingspeak successor. Will stop growing kingspeak
  - New Intel Skylake processors – higher core count per node and AVX512 support
- Refresh of Protected Environment - Redwood
  - NIH Instrumentation Grant
  - New HPC, storage, VM, windows
  - On line early 2018
- Additional General Lonepeak nodes
- New Home Directory Space
  - Watch for news in next couple of months

Getting a CHPC Account

- CHPC uses campus uNID and password
- PIs must have account and will need to approve accounts for any members of their research group (can delegate)
- Account Application Procedure – Online process
  - Complete CHPC account form at https://www.chpc.utah.edu/role/user/account_request.php
  - For collaborators outside of University of Utah must complete affiliate form with HR to get uNID https://www.hr.utah.edu/forms/affiliate.php_ and then use account application

Security Policies

- No clear text passwords, use ssh and scp
- You may not share your account under any circumstances
- Don't leave your terminal unattended while logged into your account
- Do not introduce classified or sensitive work onto CHPC systems unless on Protected Environment
- Do not try to break passwords, tamper with files etc.
- Do not distribute or copy privileged data or software
- Report suspicions to CHPC (security@chpc.utah.edu)
- See http://www.chpc.utah.edu/docs/policies/security.html for more details

Accessing Clusters

- Login or interactive nodes with each cluster
  - ssh –Y ***cluster***.chpc.utah.edu where ***cluster*** is kingspeak, ember, tangent, ash-guest, lonepeak or apex
- Interactive nodes only used for short compiles, editing and very short test runs
- ***No more than 15 minutes and no jobs of any length that make heavy use of cpu or memory!***
- Have script which watches running processes and notifies users when in violation of the acceptable usage policy

Accessing Login nodes

- Use FastX from Mac, Windows, or Linux desktops -- preferred
  - https://www.chpc.utah.edu/documentation/software/fastx2.php
- Alternatively:
  - From windows need ssh client
    - PuTTY http://www.chiark.greenend.org.uk/~sgtatham/putty/
    - Xshell http://www.netsarang.com/products/xsh_overview.html
  - For X forwarding applications also need
    - Xming http://www.straightrunning.com/XmingNotes/
      - Look for “mesa” version
  - From mac/linux – use terminal ssh (with –Y for X forwarding)
- Access to protected environment needs Duo two factor authentication (and VPN if off campus)

FastX – Tool for Remote X

- https://www.starnet.com/fastx
- Used to interact with remote linux systems graphically in much more efficient and effective way then simple X forwarding
- Graphical sessions can be detached from without being closed, allowing users to reattach to the session from the same or other systems
- Server on all interactive nodes as well as the frisco nodes; some servers have graphics cards and support OpenGL
- Clients for windows, mac and linux; can be installed on both university and personal desktops.

FastX

- For FastX – see “To Use” section of documentation at https://www.chpc.utah.edu/documentation/software/fastx2.php
- Download client following directions on page
- Do install
- Start program
- Set host to kingspeak1.chpc.utah.edu OR kingspeak2.chpc.utah.edu OR other interactive node OR one of the frisco nodes (frisco1-frisco8.chpc.utah.edu)

Login scripts

- CHPC provides login scripts ("dot" files) when creating account for both tcsh and bash shells
- These files set the environment so that applications are found, batch commands work – ***Do not remove***
- Choose shell at account creation – can change at www.chpc.utah.edu (sign in, select edit profile)
- Four files: .bashrc, .tcshrc, .custom.sh, .custom.csh
  - **The first two should not be edited**
  - **The second two is where to add custom module loads**
- Will automatically execute an .aliases file if it exists
- Modules for environment control – see https://www.chpc.utah.edu/documentation/software/modules.php

CHPC Uses Modules for Setting Environment

- CHPC provides login scripts ("dot" files) when creating account for both tcsh and bash shells
- These files set the environment so that applications are found, batch commands work – ***Do not remove or edit!***
- **https://www.chpc.utah.edu/documentation/software/modules.php** for information
- Presentation on Modules – Thursday, September 14th

Batch System Information

- Used to access compute nodes which must be used for any extensive use
- Use SLURM – Simple Linux Utility for Resource Management
- https://www.chpc.utah.edu/documentation/software/slurm.php for information
- Presentation on Slurm – Tuesday Sept 19th

Software on Clusters

- Have a variety of compliers, mpi packages, math libraries and applications installed
- Some licensing restrictions may apply
- If you need a package we do not currently have installed – ask us!
- Currently we place most installations at:
  - /uufs/chpc.utah.edu/sys/pkg ***OR*** /uufs/chpc.utah.edu/sys/installdir
- Have a searchable application database
  - https://www.chpc.utah.edu/software/chpc/

Allocation

- General Allocation Process Information
  - https://www.chpc.utah.edu/documentation/policies/1.4AllocationPolicies.php
- Regular allocation form
  - https://www.chpc.utah.edu/apps/profile/allocation_form.php
  - Requests due Sept 1, Dec 1, Mar 1, and Jun 1
  - Allocation in core hours
- Quick allocation
  - https://www.chpc.utah.edu/apps/profile/allocation_quick_form.php
- Check usage -- https://www.chpc.utah.edu/usage/cluster/current-project-general.php

Windows Statistics Server

- Kachina/Swasey – each 48 core, 512TB memory
- Presently has the following software installed
  - SAS 9.4 with text miner
  - SPSS
  - R
  - STATA
  - Mathematica
  - Matlab
- ***If you need other software, please contact us to discuss***

Virtual Machine Farm

- For needs and applications that do not fit in compute cluster or Windows server
- Multiple VM servers with fail over
- VM storage
- Have community mysql/mssql VMs, git repositories, web servers, etc

CHPC Fall Presentation Series

In INSCC Auditorium at 1-2pm unless otherwise noted – can join remotely via skype for business – *1-3pm; **9am-3pm

Thursday August 31st	Hands on Introduction to Linux, part 1*	Wim Cardoen & Anita Orendt
Tuesday September 5th	Hands on Introduction to Linux, part 2*	Wim Cardoen & Anita Orendt
Thursday September 7th	Hands on Introduction to Linux, part 3*	Wim Cardoen & Anita Orendt
Tuesday & Wednesday September 12th &13th	XSEDE HPC Monthly Workshop: Big Data** NOTE – must pre-register with XSEDE	Martin Cuma
Thursday September 14th	Module Basics	Anita Orendt
Tuesday September 19th	Slurm and Slurm Batch Scripts	Anita Orendt
Thursday September 21st	Introduction to the Use of Open Science Grid Resources	Wim Cardoen

https://www.chpc.utah.edu/presentations/Fall2017CHPCPresentationSchedule.php

If you would like training for yourself or your group, CHPC staff would be happy to accommodate your request. Please contact anita.orendt@utah.edu

Getting Help

- CHPC website
  - www.chpc.utah.edu
    - Getting started guide, cluster usage guides, software manual pages, CHPC policies
- Jira Ticketing System
  - Email: issues@chpc.utah.edu
- Help Desk: 405 INSCC, 581-6440 (9-5 M-F)
- We use chpc-hpc-users@lists.utah.edu for sending messages to users; also have Twitter accounts for announcements -- @CHPCOutages & @CHPCUpdates