High Performance Computing at Texas Tech University

Alan F. Sill, PhD

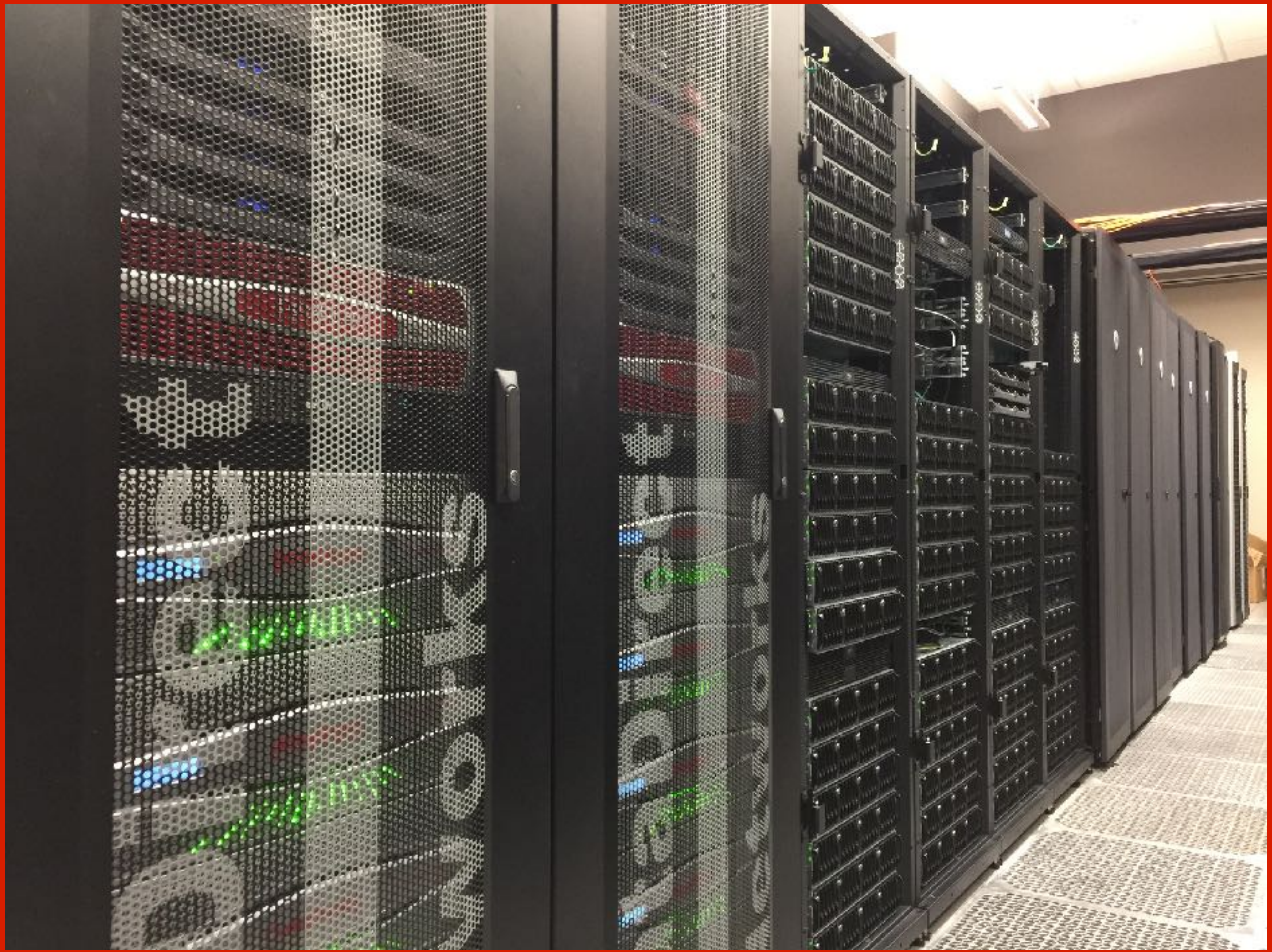
Managing Director, High Performance Computing Center, Texas Tech University

Co-Director, US National Science Foundation Cloud and Autonomic Computing Industry/University Cooperative Research Center

TTU Physics Colloquium | October 9, 2018 | Lubbock, TX

Outline of this talk

- HPCC Data Centers at TTU
  - Personnel
  - Facilities and equipment
  - Software and Tools
- Access and Usage
  - Population that uses HPC
  - Collaborations and resources
- Other Resources You Can Use
- Conclusions



HPCC Staff and Students (Fall 2018)

Staff members:

- Alan Sill, PhD — Managing Director
- Eric Rees, PhD — Research Associate
- Chris Turner, PhD — Research Associate
- Amy Wang, MSc — Programmer Analyst IV
- Travis Cotton, MSc — Programmer Analyst III
- Vidya Eswarappa, MSc — Programmer Analyst II

Also open:

- Assistant Managing Director (vacant)

Graduate students:

- Misha Ahmadian, Graduate Research Assistant
- Jianluo He, Graduate Student Assistant
- Jayacharan Kolla, Graduate Student Assistant

Undergraduate students:

- Josiah Hamilton
- Zachary Hansen
- Joshua Holtz

System Administration

- Three dedicated programmer/analysts monitor operation of the clusters, storage, and other systems daily and provide a first level of response to any problems.
- Backed up by PhD-staffed Facilities and Operations design team for major problems with the resources, and by the ASSURE user support team for detailed intervention needed to support user issues.
- Provides support to and supervision of the student-assisted help desk.
- Answers tickets generated by email from users and forwards or calls for backup from the other teams as necessary.
- This system allows us to operate with immediate response to operational issues while still having further support teams in reserve to take on specialized issues.
- Users experience a high degree of interactive, timely response, with detailed issues referred to our advanced support team.

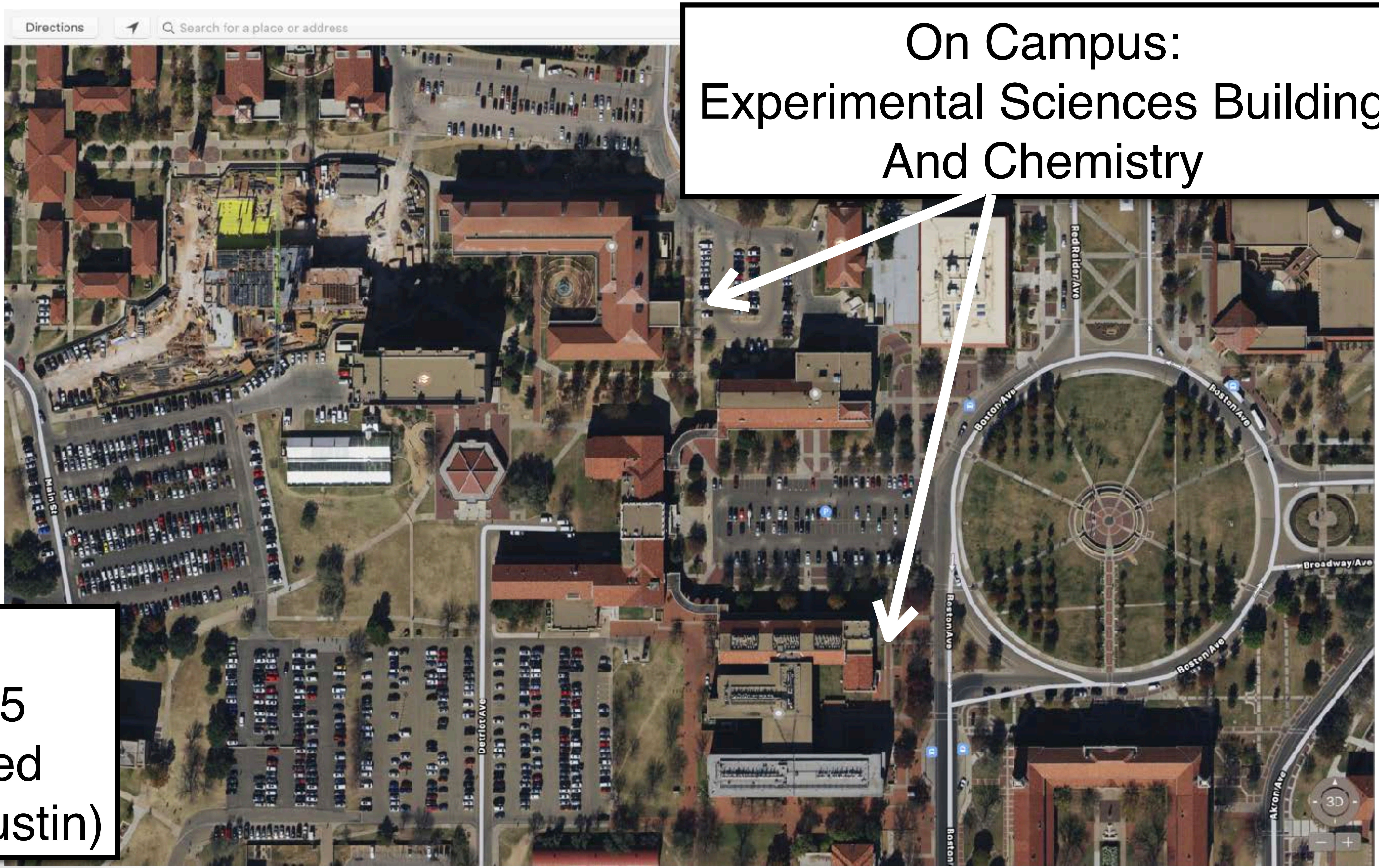
Advanced User Support

Advanced application support

- System-wide applications and special software installation requests
- Technical support for user-specific applications and technologies (containers, custom environments)
- Review of usage pattern issues and performance/system impact topics
- Training and content development for improved user support capabilities

- HPCC user recruitment and training
  - Seminars (deliver and attend)
  - Current and Potential HPCC user engagement
- New funding opportunities
  - Computational Support, and collaborative engagement
- Student-assisted HPCC Help Desk and FootPrints ticketing system
  - Group meets in person once per week; teleconference calls as needed for further coordination
  - Users can get help in person at the HPCC ESB offices during business hours

HPCC Data Center Locations

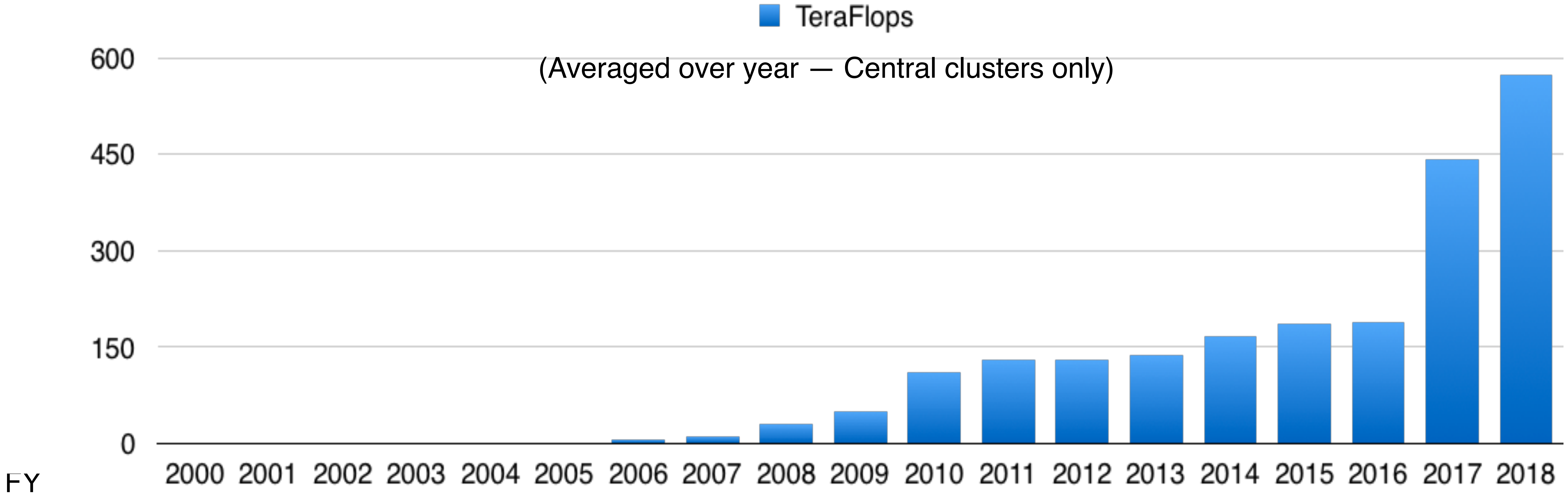


On Campus:
Experimental Sciences Building
And Chemistry

Off Campus:
Reese Building 555
And Texas Advanced
Computing Center (Austin)

Brief History

- Founded in 1999 with significant support by TIEHH and COE; merged into IT Division
- First cluster in FY 2001; TechGrid (campus grid) beginning in 2002
- Statewide grid project (TIGRE) 2005-2007; participation in Open Science Grid since then
- Large cluster with MPI in 2005 with upgrades in 2007, 2010, 2014, and 2017
- First community cluster added in 2006; increments added based on researcher demand
- On Top500 list in 2008 (289), 2009 (176), 2010 (111), 2011 (147), 2012 (414)
- Recent upgrades designed to aim us back towards re-entering the Top500 list



Primary Application Areas

- Molecular Dynamics (Chemistry, Physics, Chemical Engineering)
- Quantum Dynamics
- High Energy Physics
- Numerical Weather Prediction
- Radiation Dosimetry / Health Physics
- Mechanical Engineering (Finite Element Modeling simulations)
- Fluid Dynamics
- Energy exploration / Reservoir modeling
- Bioinformatics
- Image Processing
- Machine Learning, Deep Learning
- Data Science Applications
- Interactive Analysis Methods (Jupyter notebooks, R studio, MatLab)

TTU HPCC: Brief Graphical History

Technology	2000	2005	2010	2015	Scale of Resources
Shared Memory					64 cores
Central Clusters					96 cores (2001) - 25,824 (2018)
Researcher-Owned Clusters					Mix from ~32 to ~1000 cores
Campus Grid					Few to several hundred cores
External Grids and Collaborations					Shared access to 1000k+ cores
External Hosting					Lonestar 4/5 ~1600 cores as overflow

Shading shows approximate relative contribution to total resources

TTU Clusters: Topology, History, & Design

Cluster Name	Initial Year (Status)	Number of Cores	CPU Family and Node Model	Fabric
Grendel	*2008 (Retired)*	*~~1680~~*	*Xeon E5450 Harpertown Dell 1950*	*DDR Infiniband at nodes and switches (non-blocking)*
Hrothgar	**2011 (Operating)**	**~~7680~~ (6562 now)**	**Xeon X5660 Westmere Dell R410**	**DDR Infiniband (nodes) / QDR Infiniband (switches) (non-blocking)**
Hrothgar additions	**2013-2015 (Operating)**	**Additional 2450**	**Xeon E5-2670v2 Ivy Bridge Dell C6320**	**FDR->QDR Infiniband (nodes)/QDR Infiniband (switches) (2:1 oversubscr.)**
Realtime, Nepag, etc.	**2010-2015 (Operating)**	**Assorted up to 1064**	**Assorted from processors and node types above**	**Currently typically FDR Infiniband, single FDR switch (non-blocking)**
Quanah	**2017 (Operating)**	**16812**	**Xeon E5-2695v4 Broadwell Dell C6320**	**Omni-Path 100 Gbps (non-blocking)**

Primary Clusters

Hrothgar Cluster

- 7092 Cores for parallel and serial jobs(12 Cores/Node, 654 Nodes, 24-48 GB/Node), Intel Westmere X5660 @ 2.80GHz , 86.2 Teraflops/sec peak
- 1920 Cores (96 Nodes, 20 Cores/Node, 64-256 GB/Node) Intel Ivy Bridge E5-2670v2 @ 2.50GHz, 101 Tflops/sec
- DDR 20 GB/Second Infiniband fabric for MPI communication & storage
- 1Gb/s Ethernet network for control, cluster management, and communication to serial nodes
- Researcher-owned “community cluster” nodes

Quanah Cluster

- 16812 Cores (467 Nodes, 36 Cores/Node, 192 GB/Node), Intel Xeon E5-2695 v4 @ 2.10GHz
- Dell C6300 4-node enclosures
- Benchmarked 485.7 Teraflops/sec (464/467 nodes)
- Omni-Path 100 Gbits/Second fabric for MPI communication & storage
- 10 Gb/s Ethernet network for cluster management
- 22 nodes owned by researchers, accessible on a class-of-service queue-based project priority basis
- Total power drawn: ~ 200 kW

Both Clusters:

- Dedicated NFS server for system-wide Applications
- OpenHPC cluster management software
- LMod Modules for setting up application environment
- UGE (Univa Grid Engine) for job scheduling

Grendel cluster (2008)

First cluster placed into then-new $1.8M machine room expansion

Hrothgar + additions (2011-present)

Quanah cluster (2017)

High density (72 physical cores per rack unit); High speed 100 Gbps Omni-Path fabric

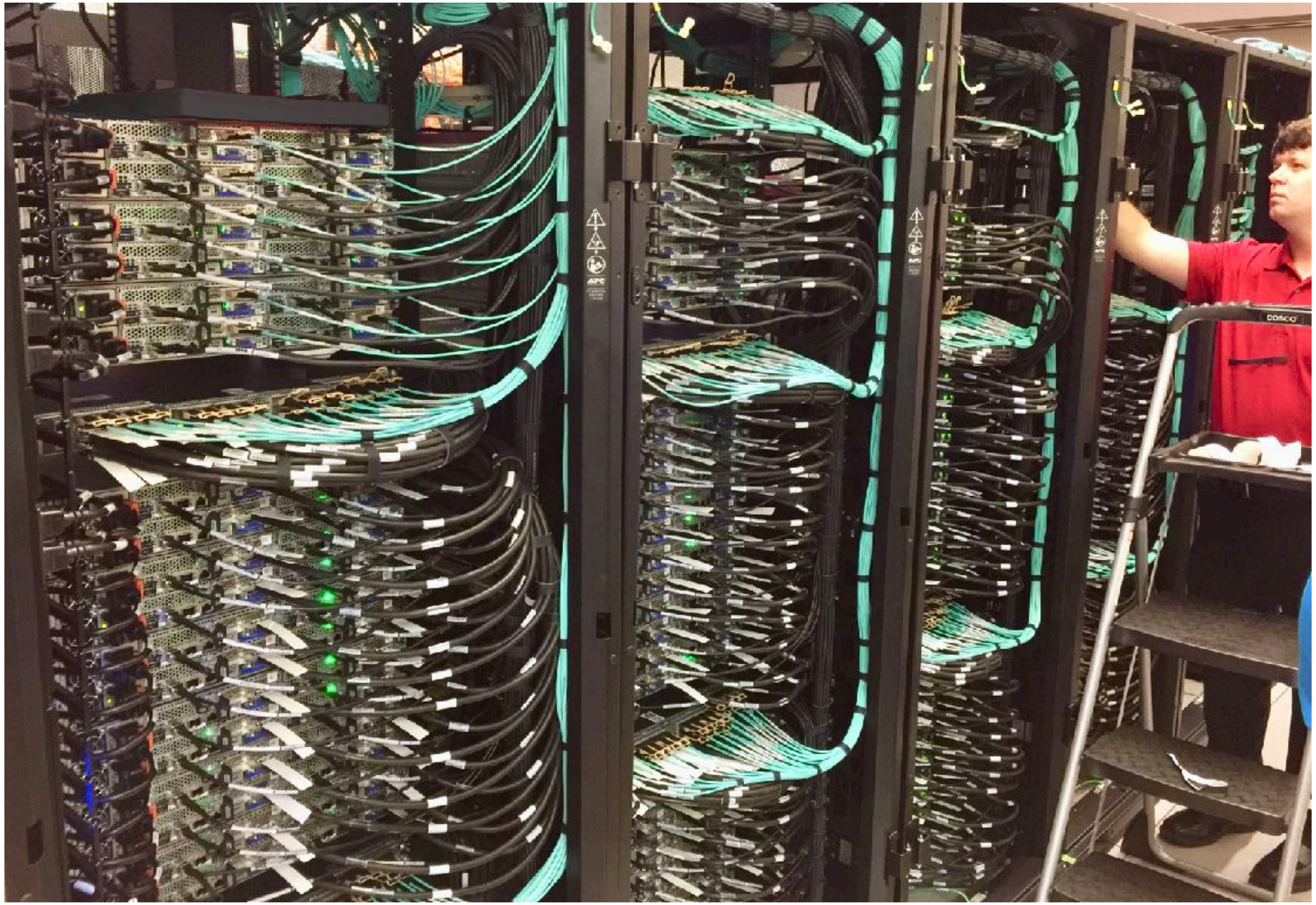
Quanah Cluster (As Upgraded Nov. 2017)

- **10 racks, 467 Dell™ C6320 nodes**
  - 36 CPU cores/node Intel Xeon E5-2695 v4 (two 18-core sockets per node)
  - 192 GB of RAM per node
  - Compute power: ~1 Tflop/s per node (485 Tflop/s benchmarked)
  - 16,812 worker node cores total

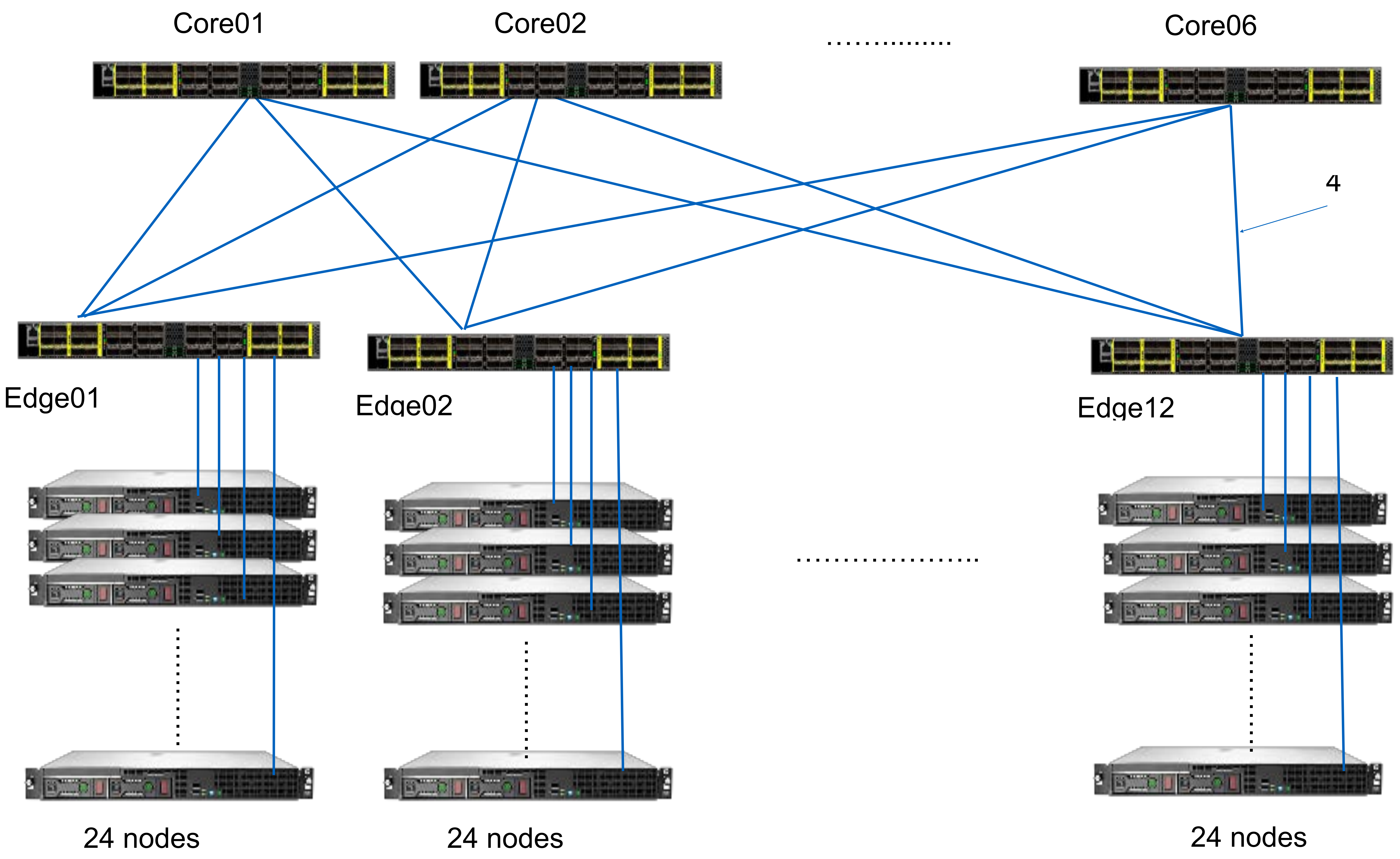
- **Operating System:**
  - CentOS 7.4.1708, 64-bit, Kernel 3.10

- **High Speed Fabric Network:**
  - Intel ™ OmniPath, 100Gbps/connection
  - Non-blocking fat tree topology
  - 12 core switches, 48 ports/switch
  - 57.6 Tbit/s core throughput capacity

- **Management/Control Network:**
  - Ethernet, 10 Gbps, sequential chained switches, 36 ports per switch



Single –rail, 2 tier, non-blocking fat-tree (folded Clos network)



Current design:
12 Core Switches

Each core switch:
48 ports @ 100 Gbits/sed
with director module

24 Edge Switches

Each with 24 ISL uplink cables in 12 bundles of 2 uplink cables per edge switch, and 24 host cables per edge switch

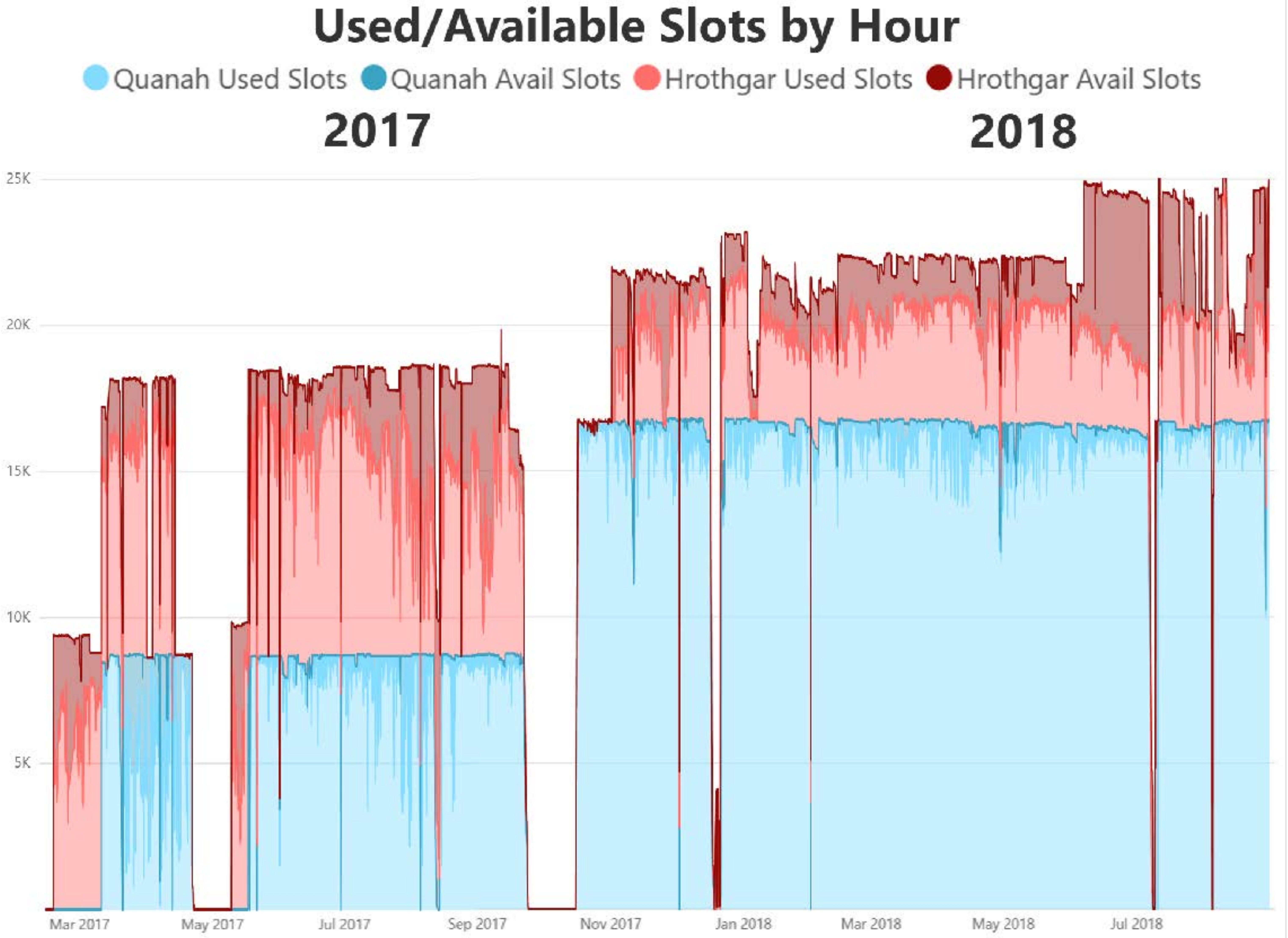
Primary Cluster Utilization

Hrothgar cluster averaged about 80% utilization before addition of Quanah phase I, commissioned in April 2017.

Addition of Quanah I roughly doubled the total number of utilized cores to ~16,000, with a total number of available cores of ~18,000.

Addition of Quanah II in November 2017 required decreasing size of Hrothgar, but still led to over 20,000 cores in regular use. Power limits prevent running all of former Hrothgar, but will be solved with new UPS and generator upgrade soon.

Expect ~25,000 cores total to be online beginning mid-November 2018 after power upgrade is complete.



Power and Cooling

Reliable power and cooling are critical for high performance computing infrastructure operation

Facility		Power		Cooling

Facility: ESB

Power

- 250 kVA
- 10 minute battery backup
- Located on the first floor of the ESB

~~150 KvA~~ → ***Upgrade to 300 kVA!***

60 Battery Storage Cabinet

10 minute battery backup

+ Addition of a 750 kVA generator!

Cooling

- 2 in-room CRAC Units with 60 tons of cooling each
- Additional building air handler (up to 30 tons)

Facility: Reese

150 kVA

48 Battery Storage Cabinet

1.5 hour battery backup

- One in-room CRAC Unit
- 15 tons of cooling with present chiller; up to 30 tons with auxiliary chiller turned on
- One portable 2.5-ton auxiliary cooling unit

Auxiliary Resources and Facilities

- Small portion (about 700 cores) of the Hrothgar cluster, plus a dedicated Atmospheric Sciences cluster and specialty nodes not part of the main clusters, are housed at the off-campus Reese Center facility.
- Research Data Management System is located in two different physical facilities (main campus University Data Center and at the Reese Center) for redundancy.
- Approximately 1600 cores of Lonestar 5 processing time at TACC in Austin are owned by TTU and allocated by the HPCC to provide extra processing to particular users. Application for access to these allocations are open to all TTU research groups.
- The HPCC also advises and helps to support other special resources at several different locations for specific research groups not operated by the HPCC (e.g., Chemistry, Center for Biology and Genomics, Atmospheric Sciences).
- The TTU IT Division also supports a new Data Science effort designed to help users getting started in this area who do not yet need HPCC resources.

Storage and LNet Routers

- Main storage is currently 2.5 PB (formatted) DDN SFA-12K Exascaler array with 880 2-TB spinning disks. *Plan to add additional 3.4 PB in November 2018.*
- Users also can purchase additional dedicated, backed-up storage space. *Cost: $80/TB/year with full separate backup, or $40/TB/year with backup waiver.*
- Lnet routers are used to connect clusters to storage fabric:
  - Each router is a small server with a fabric adapter of each type equipped with a small core count single high-clock-rate CPU
  - Number needed = roughly same as the maximum number of object storage servers active at any given time to avoid being bandwidth bottleneck.
  - In our case, we provisioned 8 OmniPath to Infiniband routers for Quanah and six IB-to-IB routers for each of the ivy and west queues to separate Hrothgar from storage fabric the different Hrothgar queues from each other for increased reliability and roadmap positioning

HPCC Lustre File System

You have access to three storage areas

- *Home - /home/<eraider>*
  - Quota: 150GB
  - Backed up nightly
  - Never purged
- *Work - /lustre/work/<eraider>*
  - Quota: 700GB
  - Never backed up nor purged
- *Scratch - /lustre/scratch/<eraider>*
  - Quota: None
  - Never backed up
  - Purged monthly or as needed to maintain < 80% scratch usage

Area	Quota	Backup	Purged
/home/<eraider>	150 GB	Yes	No
/lustre/work/<eraider>	700 GB	No	No
/lustre/scatch/<eraider>	None	No	As needed to maintain < 80% usage

Quota status shown on each login:

```
Current Storage Usage for user:
           /home - Currently using  71 of 150 GB (47%).
     /lustre/work - Currently using 166 of 700 GB (23%).
```

Environment Settings

- Using LMod Modules
  - *User Guide: http://www.depts.ttu.edu/hpcc/userguides/general_guides/software_environment.php*

- Modules commands:
  - *module avail*
  - *module list*
  - *module load <module_name>*
  - *module unload <module_name>*
  - *module spider <keyword>*
  - *module purge*

```
quanah:$ module avail

----------------------------------------- /opt/apps/nfs/module/modulefiles ------------------------------------------
   gnu/5.4.0    gurobi/v751    intel/17.3.191    java/1.8.121    matlab/R2017b    perl/5.16.3

Use "module spider" to find all possible modules.
Use "module keyword key1 key2 ..." to search for all possible modules matching any of the "keys".

quanah:$ module spider python

----------------------------------------------------------------------------------------------------------------
  python:
----------------------------------------------------------------------------------------------------------------
    Description:
      Python-2.7.9 compiled with Intel

     Versions:
        python/2.7.9-gnu
        python/2.7.9-intel

----------------------------------------------------------------------------------------------------------------
  For detailed information about a specific "python" module (including how to load the modules) use the module's full
 name.
  For example:

     $ module spider python/2.7.9-intel
----------------------------------------------------------------------------------------------------------------

quanah:$
```

Submitting Jobs

- How to submit jobs
  - *User Guide: http://www.hpcc.ttu.edu/userguides/general_guides/job_submission.php*
- Example Script
  - *Set the name of our job to "MPI_TEST_JOB"*
  - *Requests to run on the Quanah cluster using the omni queue.*
  - *Sets the parallel environment to "mpi" and requests 36 cores.*
- Submit job
  - *qsub <job submission script>*

```
#!/bin/sh
#$ -V
#$ -cwd
#$ -S /bin/bash
#$ -N MPI_Test_Job
#$ -o $JOB_NAME.o$JOB_ID
#$ -e $JOB_NAME.e$JOB_ID
#$ -q omni
#$ -pe mpi 36
#$ -l h_vmem=5.3G
#$ -l h_rt=48:00:00
#$ -P quanah

module load intel impi

mpirun --machinefile machinefile.$JOB_ID -np $NSLOTS ./mpi_hello_world
```

```
#!/bin/sh
#$ -V
#$ -cwd
#$ -S /bin/bash
#$ -N MPI_Test_Job
#$ -o $JOB_NAME.o$JOB_ID
#$ -e $JOB_NAME.e$JOB_ID
#$ -q west
#$ -pe west 12
#$ -P hrothgar

module load intel impi

mpirun --machinefile machinefile.$JOB_ID -np $NSLOTS ./mpi_hello_world
```

Check Cluster Usage/Status

- Viewing the current queue status and running/pending jobs
  - *For a queue overview, run the command: "qstat -g c"*
  - *Visit the queue status webpage: http://charlie.hpcc.ttu.edu/qstat/qstat.html (Updates every 2 minutes)*

- Job Runtime Limits

Cluster	Queue	Project	Runtime Limit	# of Cores per node	Memory per node
Quanah	omni	quanah	48 hours	36 cores	192 GB
Quanah	omni	xlquanah	120 hours	36 cores	192 GB
Quanah	omni	hep	(*)	36 cores	192 GB
Hrothgar	ivy	hrothgar	48 hours	20 cores	64 GB
Hrothgar	west	hrothgar	48 hours	12 cores	24 GB
Hrothgar	serial	hrothgar	120 hours	12 cores	24 GB
Hrothgar	community cluster queues (Chewie, R2D2, Yoda, ancellcc, blawzcc, caocc, dahlcc, phillipscc, tangcc, tang256cc)	communitycluster	(*) none for researcher-owned resources	vary	vary

X Windows Remote Connections

- Interactive GUI using Linux/Mac
  - *Install X Server (Xquartz)*
  - *Add the following options to your normal ssh command:*
    - -nNT
    - -L 2222:<cluster>.hpcc.ttu.edu:22
  - *Example:*
    - ssh -nNT -L 2222:quanah.hpcc.ttu.edu:22 eraider@quanah.hpcc.ttu.edu
  - *Run a test command like xclock.*

- Interactive GUI using Windows
  - *Install X Server (Xming)*
  - *Start the X Server application.*
  - *Turn on X11 forwarding in your client of choice.*
  - *Log into the cluster.*
  - *Run a test command like xclock.*

- Matlab Interactive Job Example:
  - *From Quanah:*
    - qlogin –q omni –P quanah
    - module load matlab
    - matlab

Transferring Data

- Transfer files using Globus Connect
  - *User Guide:* *http://tinyurl.com/hpcc-data-transfer*
- Whenever possible, refrain from using:
  - *scp,*
  - *sftp,*
  - *rsync,*
  - *Or any other data transfer tool.*
  - *Such tools can slow down use of the login nodes and interfere with the work of others.*

- **Why use Globus?**
  - *Globus Connect service is well connected to the campus network with 10 Gbit/second networks.*
  - *Globus Connect works with Linux, Mac and Windows and is controlled through a web GUI.*
  - *Globus connect service eliminates the data transfer load from the cluster login node.*
  - *The data transfer nodes are better located within HPCC networks for transferring user data.*
  - *Numerous other sites (including TACC) support Globus Connect data transfers.*
  - *Interfaces well with planned Globus Publish methods to be used in upcoming TTU Research Data Management System (RDMS).*
  - *Data transfers can be automated and scripted.*

Training & Documentation

User Training opportunities are announced and kept up to date on the HPCC web site:
http://www.hpcc.ttu.edu

Upcoming User Training Workshops

Course	Date	Time	Location	Register
GPU Programming Using OpenACC (XSEDE)	Nov. 6	10am - 4pm	TLPDC #153	--
New User Training	Jan. 29	1pm - 3pm	ESB #120	N/A
New User Training	Feb. 13	3pm - 5pm	ESB #120	N/A
New User Training	June 6	1pm - 3pm	ESB #120	N/A
New User Training	July 11	3pm - 5pm	ESB #120	N/A

XSEDE HPC Workshop: GPU Programming Using OpenACC

XSEDE, along with the Pittsburgh Supercomputing Center will be presenting an OpenACC GPU programming workshop on November 6, 2018.

OpenACC is the accepted standard using compiler directives to allow quick development of GPU capable codes using standard languages and compilers. It has been used with great success to accelerate real applications within very short development periods. This workshop assumes knowledge of either C or Fortran programming. It will have a hands-on component using the Bridges computing platform at the Pittsburgh Supercomputing Center.

Registration is required. A link will be provided here when registration opens in mid-October.

Sessions:

- 10am - 4pm, November 6

Location: Texas Tech University Library, TLPDC Room #153

Slides from the most recent User Training Workshop are also made available there, as well as online user guides for common tasks such as logging on remotely, using applications, submitting jobs and doing qlogin interactive analysis, transferring data, etc.

Beyond Texas Tech

In addition to the campus primary clusters, researcher-owned community cluster and class-of-service shared access cpu resources, specialty clusters, dedicate and shared network and storage resources, TTU participates in several collaborative projects in support of our research community. These include:

- Open Science Grid - community of discipline-specific virtual organizations that operate with a common set of middleware software and negotiate access to common resources
- Campus Compute Cooperative - a new pilot program being developed by several research universities to exchange access to specific types of hardware and resources
- XSEDE - through support for monthly HPC training classroom-format training seminars covering topics such as parallel and MPI programming, supercomputing techniques, big data toolkits and access to national-scale resources

If you need access to any of the above, or have a specialty need for collaboration, please contact us at hpccsupport@ttu.edu

Conclusions

The HPCC operates a wide variety of resources designed to provide access to high speed computing, networking for data transfer, and storage to support the needs of the Texas Tech research community.

Increasingly, we are seeing that the number and diversity of groups making use of these resources is increasing as nearly every field of study develops the need to use advanced computing resources (and certainly every field of study is beginning to use major amounts of data).

The HPCC hopes to bridge the gap between desktop, laptop, and mobile devices and the very high capacity resources of the national and international supercomputing facilities for our users. We do this by providing access to state-of-the-art, modern resources that are provisioned with high quality software and tools, and by supporting the TTU research community in its mission through training seminars, consulting, on-boarding activities and online documentation aimed at getting faculty, staff, and students able to use these resources. Please contact us at hpccsupport@ttu.edu if we can help you further!