Texas Advanced Computing Center (TACC) Overview

John Fonner, PhD
Life Sciences Computing

Texas Advanced Computing Center (TACC)

TACC's mission is to enable discoveries that advance science and society through the application of advanced computing technologies.

UT Austin: What starts here, changes the world

TACC: Powering discoveries that change the world

TACC's Vision

- Provide the most powerful, capable computing technologies and techniques that enable people—researchers, educators, developers, engineers, businessmen, etc.—to advance science and society.

- Provide leadership in the advanced computing community in technology R&D, support, education, and expertise to ensure maximum impact of current and future technologies in diverse applications.

- Enable transformational science and societal achievements that change, influence, and improve our understanding of the world, and the world itself.

About TACC

- Research Center within the University of Texas at Austin
- ~110 staff including ~30 PhD level researchers
- Manages computational resources for the scientific community and provides them <u>free of charge</u> to investigators on a project basis
- Conducts research on a variety of topics
- Funded by NSF, UT, other federal grants, and traditional PI research grants

Ranger

- 62,976 cores, 2.0 GHz AMD Opteron
- Will retire in February 2013
- Still one of the most powerful supercomputers in academia, with over 579 teraflops peak performance.
- To build and deploy Ranger, the National Science Foundation awarded TACC $59 million, the largest single award ever from that agency to UT Austin.

Four year statistics:
- 97% uptime
- Over 2 Billion jobs run
- More than 1.5 Billion hours of compute time used

Lonestar

- Partners include NSF, UT Austin, UT System, Texas A&M and Texas Tech

- 1,888 Dell M610 PowerEdge blade servers, each with two six-core Intel Xeon 5600 "Westmere" processors – **total 22,656 cores**

- Smaller than Ranger, but **more powerful** per node

- 302 teraflops peak performance

Stampede

- Base Cluster (2 Petaflops):
  - Intel Sandy Bridge (8 core) processors
  - Dell dual-socket nodes w/32GB RAM
  - More than 6,000 nodes
  - More than 100,000 cores
- Co-Processors (8 Petaflops):
  - Intel Xeon Phi co-processors
  - Special release of "Knights Corner" (>50 cores)
- Total concurrency approaching 500,000 cores
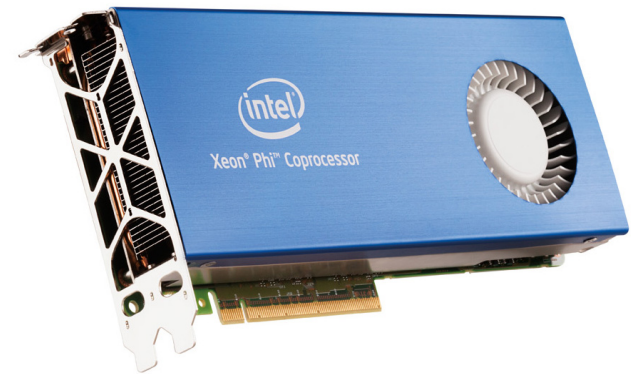
UT Data Repository (Corral)

- 5 Petabytes of geo-replicated storage
  - Data centers in Austin and in Arlington
  - Parallel file system on high-speed network
- Available to all MD Anderson researchers
- Allocations up to 5TB are free of charge
- Larger allocations for $250/TB/year
- Synergistic with Lonestar/Stampede for compute and Ranch for archive storage

Data Storage Costs

	TACC Corral	Google Drive	SkyDrive	Dropbox	iCloud
Free Basic Account	**5TB**	5GB	7GB	2GB	5GB
10GB					$20
20GB			$10		$40
25GB		$30			
50GB			$25	$100	$100
100GB		$60	$50	$200	$200
1TB		$600		$795 (five users)	
2TB		$1,200			
4TB		$2,400			
5TB	**FREE**				
8TB	**$2,000**	$4,800			
16TB	**$4,000**	$9,600			

Ranch Tape Archive

- 70 petabyte capacity (and growing)
- Used for long-term storage
- Access provided freely to users of other TACC resources
- Data staged on spinning disk and stored long-term on tape
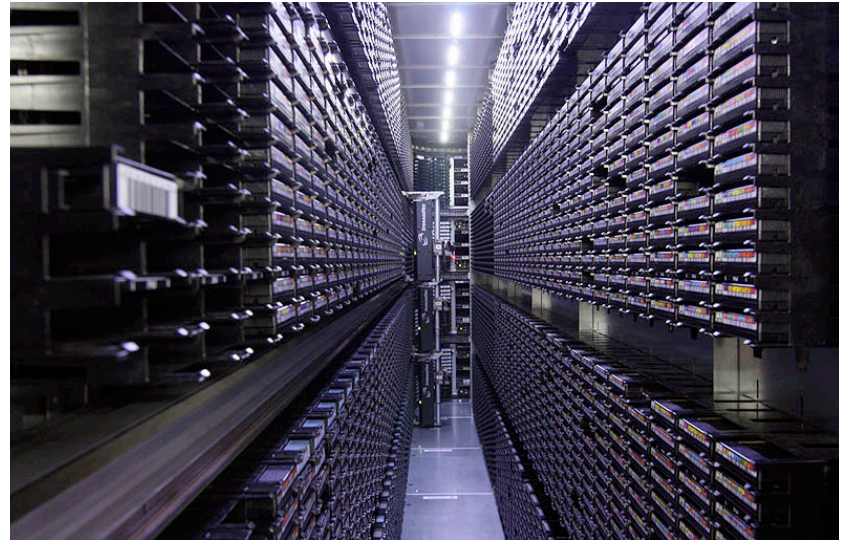
O'Donnell Gift:
Data-Driven Science

- Announced February 21, 2012
  - initial production Q4 2012
- $10 M over three years for:
  - a high-performance petascale data storage system accessible from all of TACC's computing systems
  - a computing system with embedded high-speed storage, optimized for data-intensive computing
- $2 M matching funds from UT for staffing

Visualization on Longhorn

- The largest system in the world dedicated to visualization and data analysis

- System totals: 2048 cores, 13.5 TB distributed RAM, 512 GPUs

- 17M compute hours available annually

- Ideal for remote visualization and GPU computing

Longhorn User Portal

Remote interactive visualization tool, developed at TACC

- Dramatically simplifies the visualization process
- Semi-automated data importation

- Powerful, easy-to-use entry point for researchers
- Quick and easy interactive visualization
- Web-based – works from anywhere!

Life Sciences Computing at TACC

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

- TACC maintains a quarterly release cycle for a core set of biology related applications that:
  - Have high impact and broad appeal to life sciences research
  - Are highly requested by the user community
- Compiler optimized installations using fast math libraries when possible
- 80+ packages installed for genomics, phylogenetics, and computational chemistry

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

- TACC maintains a quarterly release cycle for a core set of biology related applications that:
  - Have high impact and broad appeal to life sciences research
  - Are highly requested by the user community
- Compiler optimized installations using fast math libraries when possible

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

Genomics

Abyss	GSNAP	Picard
BEDTools	HMMER	SAMtools
BioPerl	Libgtextutils	SHRiMP
Bismark	MAFFT	SOAPdenovo
Blat	Maq	SRA toolkit
Bowtie	mpiBLAST	SSAKE
BWA	MUSCLE	Tophat
Cufflinks	NCBI BLAST+	TrinityRNASeq
FASTX-Toolkit	Newbler	Velvet
GATK	Oases	

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

Phylogenetics

BEAST	RAxML-Light
ClustalW	Parsimonator
MrBayes	Phyutility
PHYLIP	QuickTree
RAxML	

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

Computational Chemistry

Amber	GAMESS
APBS	GROMACS
AutoDock	LAMMPS
AutoDock Vina	NAMD
Desmond	NWChem
Espresso	Siesta
Gaussian (coming...)	VASP

Life Sciences Software Stack

www.tacc.utexas.edu/life-sciences-computing/applications

Broadly Used Tools

idev

Java / JDK

launcher

Octave

R (statistics package)

…and many others

User-Specific Software

- What if I need software not already installed?
  - Will it run on Linux?
  - Is the source code available?
  - Are there licensing restrictions?
- Try installing it in your personal directory (root access not required)
- Submit a consulting ticket on the TACC Portal for help
- Email or talk with one of us

Training Opportunities

- Training provided at **no cost** for UT System researchers

- Several training classes per year in Houston
  - BYOCD Workshop – 2/16/2012

- Web participation for training onsite at TACC

- Recordings are now frequently available at: http://www.tacc.utexas.edu/user-services/ training/course-materials

Research Collaboration

Examples:

- The iPlant Collaborative

- Epigenome Dynamics During DNA Replication

- Docking @ UTMB Web Portal

- RepServer

- National Center for Genome Analysis Support

John Fonner

jfonner@tacc.utexas.edu

512.232.5939

For more information:

www.tacc.utexas.edu

THE UNIVERSITY OF TEXAS AT AUSTIN

TEXAS ADVANCED COMPUTING CENTER