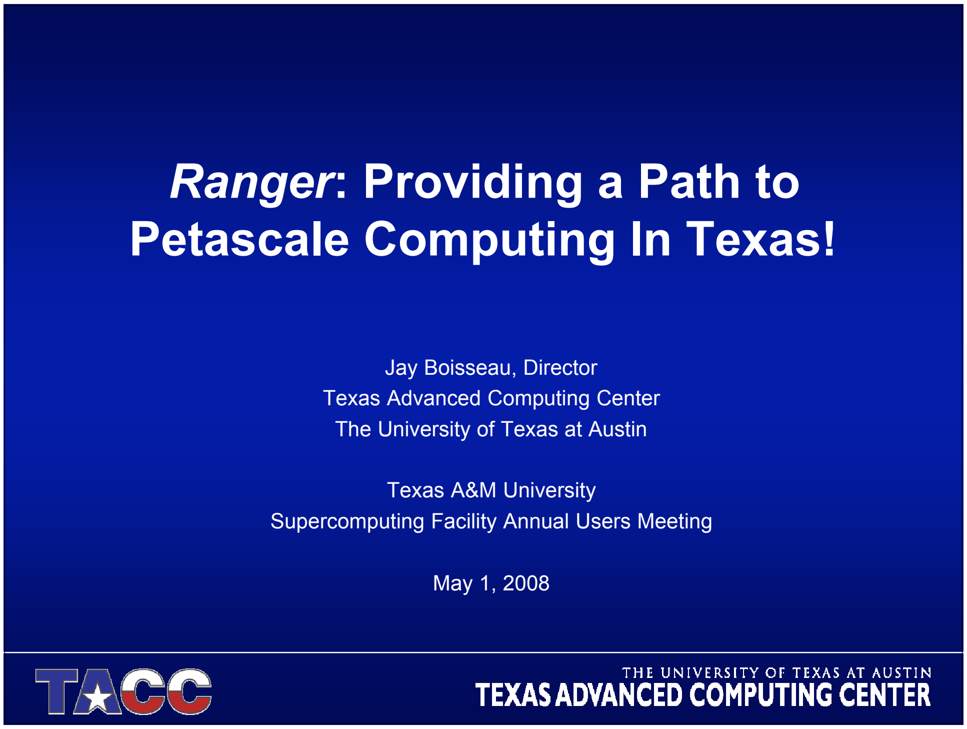
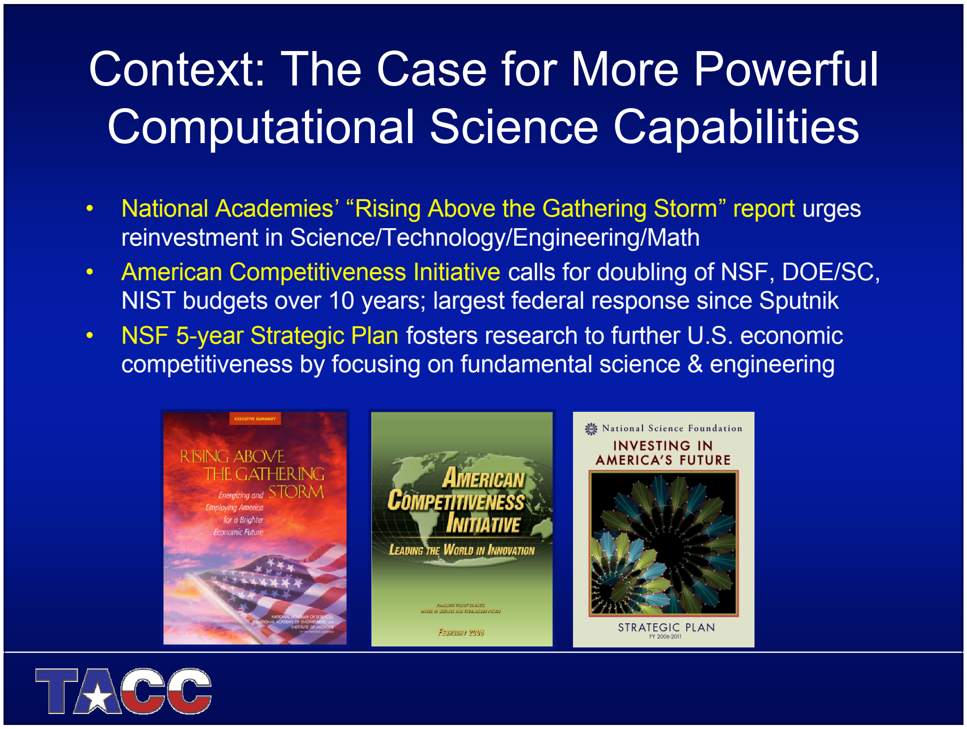


Ranger: Providing a Path to Petascale Computing In Texas!

Jay Boisseau, Director
Texas Advanced Computing Center
The University of Texas at Austin

Texas A&M University
Supercomputing Facility Annual Users Meeting

May 1, 2008



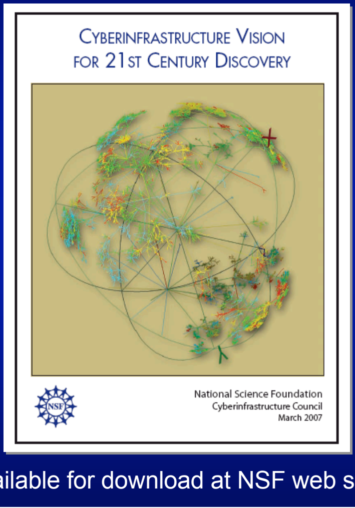
Context: The Case for More Powerful Computational Science Capabilities

- National Academies' "Rising Above the Gathering Storm" report urges reinvestment in Science/Technology/Engineering/Math
- American Competitiveness Initiative calls for doubling of NSF, DOE/SC, NIST budgets over 10 years; largest federal response since Sputnik
- NSF 5-year Strategic Plan fosters research to further U.S. economic competitiveness by focusing on fundamental science & engineering

RISING ABOVE THE GATHERING STORM
Energizing and Employing America for a Brighter Economic Future
AMERICAN COMPETITIVENESS INITIATIVE
LEADING THE WORLD IN INNOVATION
National Science Foundation
INVESTING IN AMERICA'S FUTURE
STRATEGIC PLAN
FY 2006-2011

Context: The NSF Cyberinfrastructure Strategic Plan

- NSF Cyberinfrastructure Strategic Plan released March 2007
  - Articulates importance of CI overall
  - Chapters on computing, data, collaboration, and workforce development
- NSF investing in world-class computing
  - Annual “Track2” HPC systems ($30M)
  - Single “Track1” HPC system in 2011 ($200M)
- Complementary solicitations for software, applications, education
  - Software Development for CI (SDCI)
  - Strategic Technologies for CI (STCI)
  - Petascale Applications (PetaApps)
  - CI-Training, Education, Advancement, Mentoring (CI-TEAM)
  - Cyber-enabled Discovery & Innovation (CDI) starting in 2008: $0.75B!



Cyberinfrastructure Vision for 21st Century Discovery

National Science Foundation
Cyberinfrastructure Council
March 2007

available for download at NSF web site

First NSF Track2 System: 1/2 Petaflop!

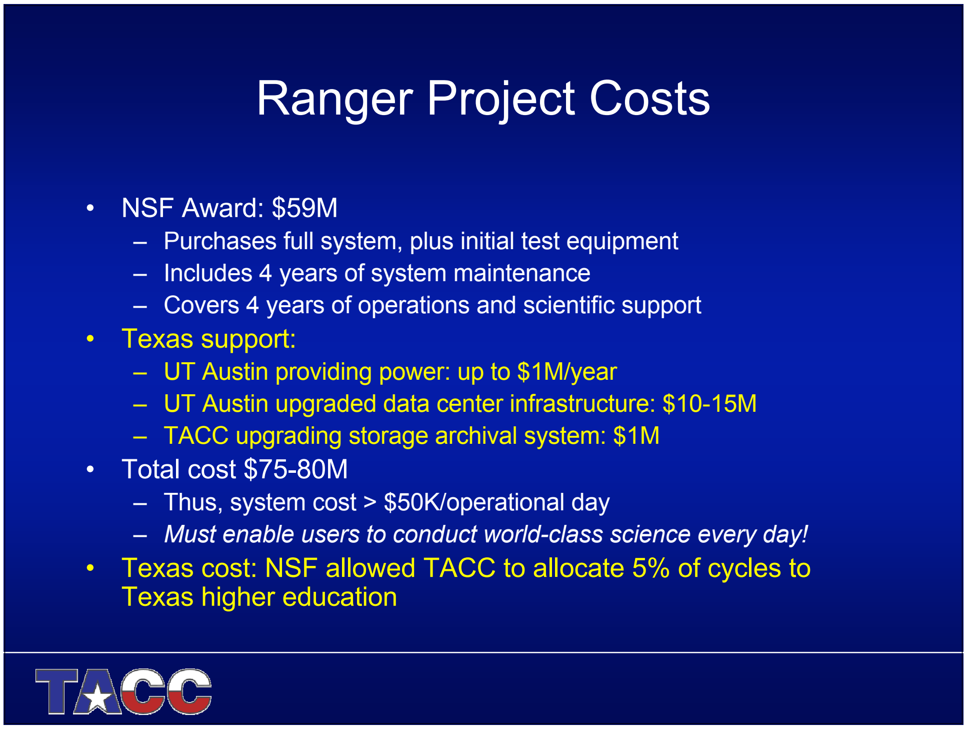
- TACC selected for first NSF ‘Track2’ HPC system
  - $30M system acquisition
  - Sun Constellation Cluster
  - AMD Opteron processors
  - Expandable configuration
- Project includes 4 years operations and support
  - System maintenance
  - User support
  - Technology insertion
  - $29M budget

Team Partners & Roles

- Institutions
  - TACC / UT Austin: project leadership, system hosting & operations, user support, technology evaluation/insertion, applications support
  - ICES / UT Austin: applications collaborations, algorithm/technique transfer and support
  - Cornell Center for Advanced Computing: large-scale data management & analysis, on-site and remote training and workshops
  - Arizona State HPCI: technology evaluation/insertion, user support
- Roles
  - Project Director: Jay Boisseau (TACC)
  - Project Manager: Chief System Engineer (TACC)
  - Co-Chief Applications Scientists: Karl Schulz (TACC), Omar Ghattas (TACC), Giri Chukkapalli (Sun)
  - Chief Technologist: Jim Browne (ICES)

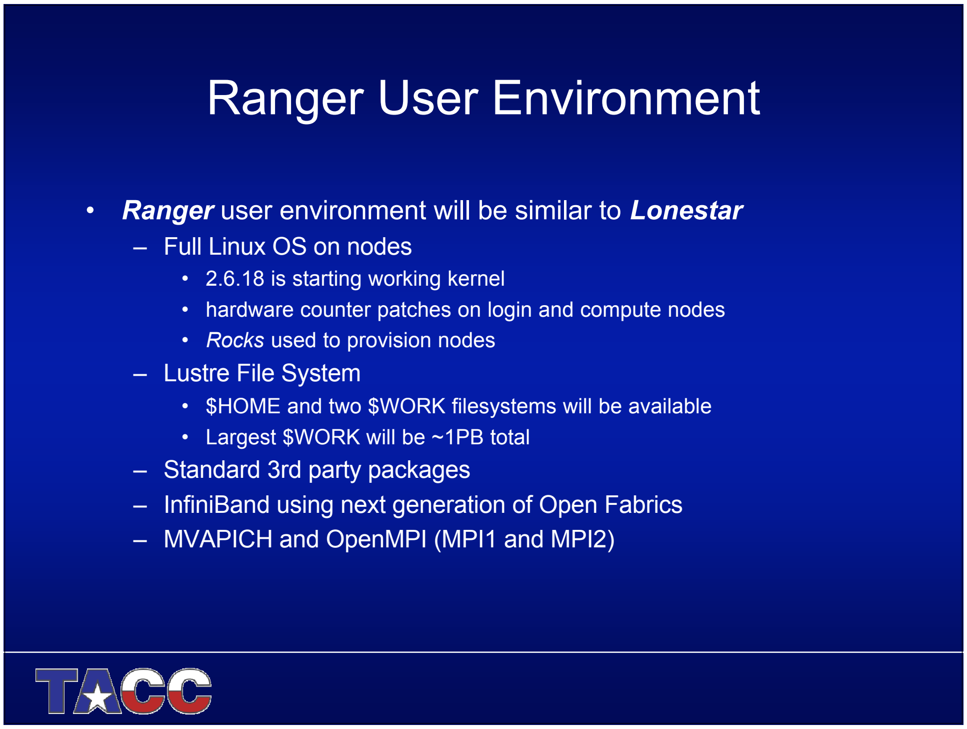
Ranger System Summary

- Compute power - 504 Teraflops
  - 3,936 Sun four-socket blades
  - 15,744 AMD Opteron “Barcelona” processors
    - Quad-core, 2.0 GHz, four flops/cycle (dual pipelines)
- Memory - 123 Terabytes
  - 2 GB/core, 32 GB/node
  - 132 GB/s aggregate bandwidth
- Disk subsystem - 1.7 Petabytes
  - 72 Sun x4500 “Thumper” I/O servers, 24TB each
  - ~72 GB/sec total aggregate bandwidth
  - 1 PB in largest /work filesystem
- Interconnect - 10 Gbps / ~3 μsec latency
  - Sun InfiniBand-based switches (2) with 3456 ports each
  - Full non-blocking 7-stage Clos fabric
  - Mellanox ConnectX IB cards



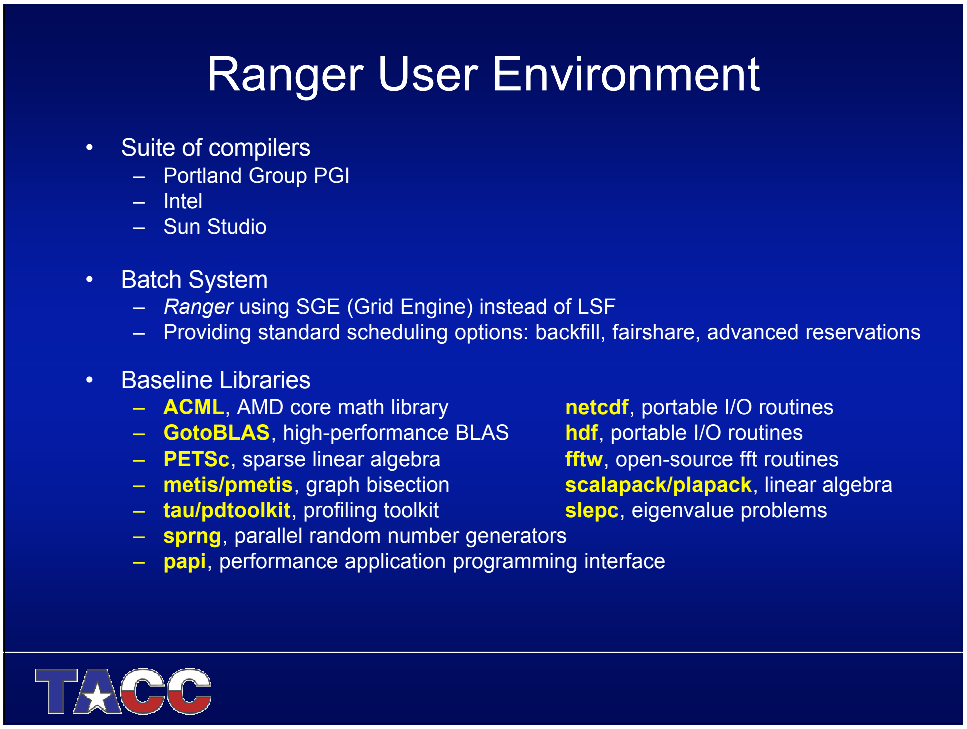
Ranger Project Costs

- NSF Award: $59M
  - Purchases full system, plus initial test equipment
  - Includes 4 years of system maintenance
  - Covers 4 years of operations and scientific support
- Texas support:
  - UT Austin providing power: up to $1M/year
  - UT Austin upgraded data center infrastructure: $10-15M
  - TACC upgrading storage archival system: $1M
- Total cost $75-80M
  - Thus, system cost > $50K/operational day
  - *Must enable users to conduct world-class science every day!*
- Texas cost: NSF allowed TACC to allocate 5% of cycles to Texas higher education



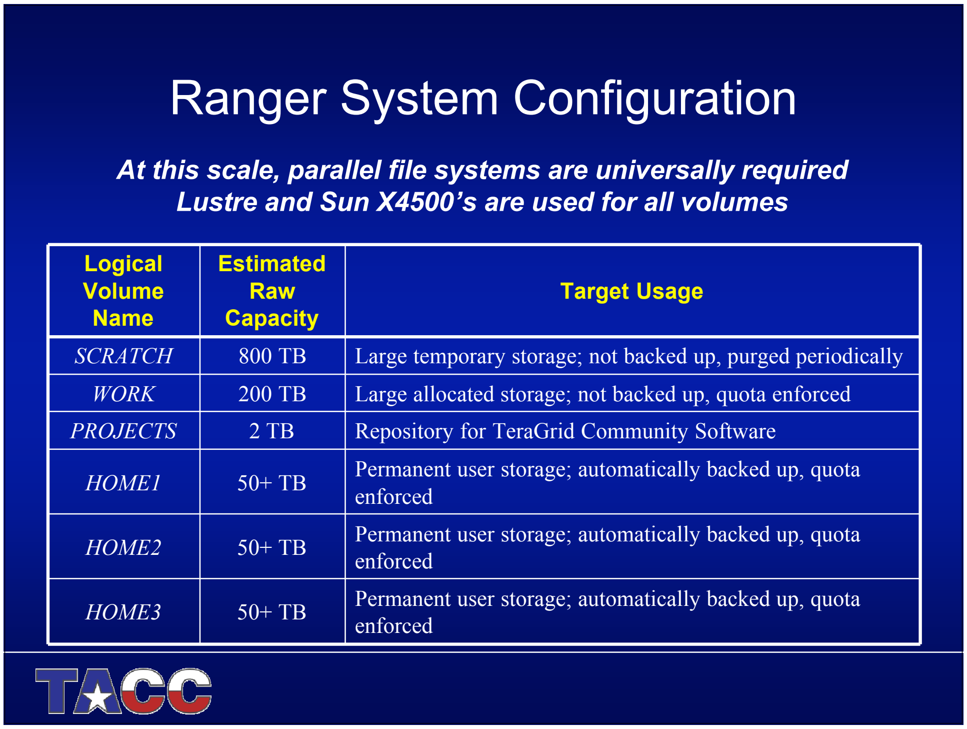
Ranger User Environment

- ***Ranger*** user environment will be similar to ***Lonestar***
  - Full Linux OS on nodes
    - 2.6.18 is starting working kernel
    - hardware counter patches on login and compute nodes
    - *Rocks* used to provision nodes
  - Lustre File System
    - $HOME and two $WORK filesystems will be available
    - Largest $WORK will be ~1PB total
  - Standard 3rd party packages
  - InfiniBand using next generation of Open Fabrics
  - MVAPICH and OpenMPI (MPI1 and MPI2)



Ranger User Environment

- Suite of compilers
  - Portland Group PGI
  - Intel
  - Sun Studio
- Batch System
  - *Ranger* using SGE (Grid Engine) instead of LSF
  - Providing standard scheduling options: backfill, fairshare, advanced reservations
- Baseline Libraries
  - **ACML**, AMD core math library
  - **GotoBLAS**, high-performance BLAS
  - **PETSc**, sparse linear algebra
  - **metis/pmetis**, graph bisection
  - **tau/pdtoolkit**, profiling toolkit
  - **sprng**, parallel random number generators
  - **papi**, performance application programming interface
  - **netcdf**, portable I/O routines
  - **hdf**, portable I/O routines
  - **fftw**, open-source fft routines
  - **scalapack/plapack**, linear algebra
  - **slepc**, eigenvalue problems



Ranger System Configuration

At this scale, parallel file systems are universally required
Lustre and Sun X4500's are used for all volumes

Logical Volume Name	Estimated Raw Capacity	Target Usage
SCRATCH	800 TB	Large temporary storage; not backed up, purged periodically
WORK	200 TB	Large allocated storage; not backed up, quota enforced
PROJECTS	2 TB	Repository for TeraGrid Community Software
HOME1	50+ TB	Permanent user storage; automatically backed up, quota enforced
HOME2	50+ TB	Permanent user storage; automatically backed up, quota enforced
HOME3	50+ TB	Permanent user storage; automatically backed up, quota enforced

Technology Insertion Plans

- Technology Identification, Tracking, Evaluation, and Insertion are crucial
  - Cutting edge system: software won’t be mature
  - Four year lifetime: new R&D will produce better technologies
  - Improve system: maximize impact over lifecycle
- Chief Technologist for project, plus supporting staff
  - Must build communications, partnerships with leading software developers worldwide
  - Grant doesn’t fund R&D, but system provides unique opportunity for determining, conducting R&D!
  - Targets include: fault tolerance, algorithms, next-generation programming tools/languages, etc.

TACC

User Support Challenges

- NO systems like this exist yet!
  - Will be the first general-purpose system at _ Pflop
  - Quad-core, massive memory/disk, etc.
- NEW user support challenges
  - Code optimization for quad-core, 16-way nodes
  - Extreme scalability to 10K+ cores
  - Petascale data analysis
  - Tolerating faults while ensuring job completion

TACC

User Support Plans

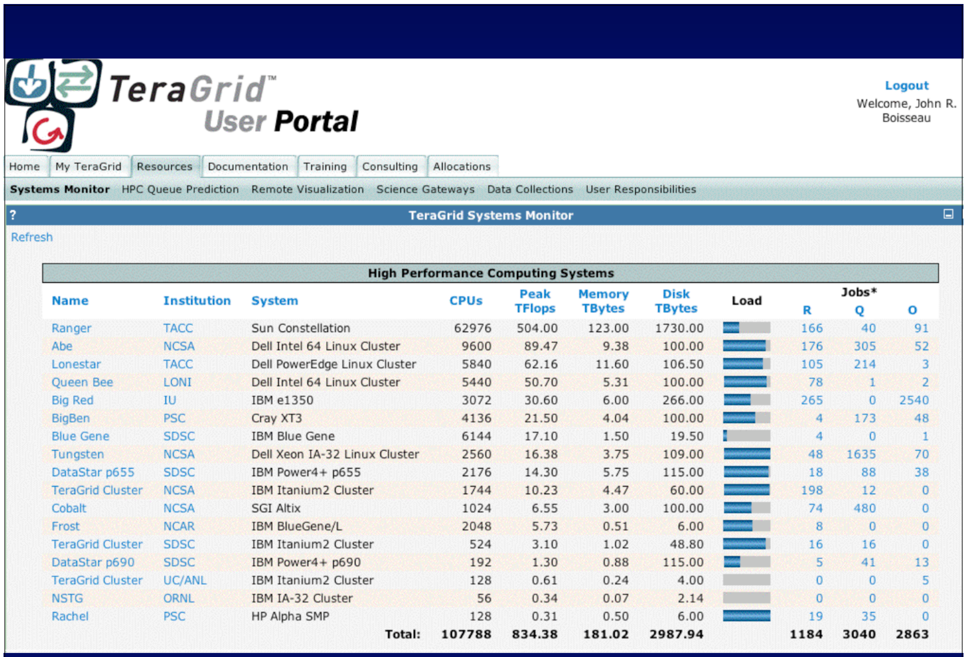
- User support: ‘usual’ (docs, consulting, training) plus
  - User Committee dedicated to this system
    - Active, experienced, high-end users
  - Applications Engineering
    - algorithmic consulting
    - technology selection
    - performance/scalability optimization
    - data analysis
  - Applications Collaborations
    - Partnership with petascale apps developers and software developers

User Support Plans

- Also
  - Strong support of ‘professionally optimized’ software
    - Community apps
    - Frameworks
    - Libraries
  - *Additional* Training
    - On-site at TACC, partners, and major user sites, and at workshops/conferences
    - Advanced topics in multi-core, scalability, etc
    - Virtual workshops for remote learning
  - Increased communications and technical exchange with all users via a TACC User Group

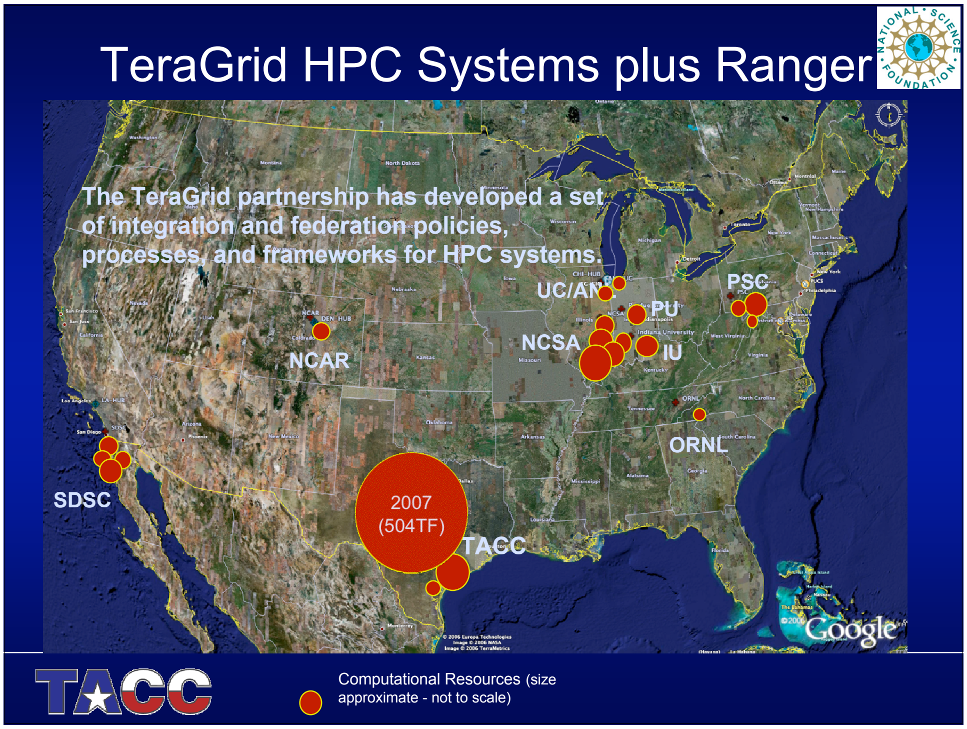
Impact in TeraGrid

- 472M CPU hours to TeraGrid
  - more than sum of *all* current TG HPC systems
- 504+ Tflops
  - 5x current top system
- Enable unprecedented research
  - *Jumpstart progress to petascale for entire US academic research community*
  - Re-establish NSF as a leader in HPC



High Performance Computing Systems

Name	Institution	System	CPUs	Peak TFlops	Memory TBytes	Disk TBytes	Load	Jobs* R	Q	O
Ranger	TACC	Sun Constellation	62976	504.00	123.00	1730.00		166	40	91
Abe	NCSA	Dell Intel 64 Linux Cluster	9600	89.47	9.38	100.00		176	305	52
Lonestar	TACC	Dell PowerEdge Linux Cluster	5840	62.16	11.60	106.50		105	214	3
Queen Bee	LONI	Dell Intel 64 Linux Cluster	5440	50.70	5.31	100.00		78	1	2
Big Red	IU	IBM e1350	3072	30.60	6.00	266.00		265	0	2540
BigBen	PSC	Cray XT3	4136	21.50	4.04	100.00		4	173	48
Blue Gene	SDSC	IBM Blue Gene	6144	17.10	1.50	19.50		4	0	1
Tungsten	NCSA	Dell Xeon IA-32 Linux Cluster	2560	16.38	3.75	109.00		48	1635	70
DataStar p655	SDSC	IBM Power4+ p655	2176	14.30	5.75	115.00		18	88	38
TeraGrid Cluster	NCSA	IBM Itanium2 Cluster	1744	10.23	4.47	60.00		198	12	0
Cobalt	NCSA	SGI Altix	1024	6.55	3.00	100.00		74	480	0
Frost	NCAR	IBM BlueGene/L	2048	5.73	0.51	6.00		8	0	0
TeraGrid Cluster	SDSC	IBM Itanium2 Cluster	524	3.10	1.02	48.80		16	16	0
DataStar p690	SDSC	IBM Power4+ p690	192	1.30	0.88	115.00		5	41	13
TeraGrid Cluster	UC/ANL	IBM Itanium2 Cluster	128	0.61	0.24	4.00		0	0	5
NSTG	ORNL	IBM IA-32 Cluster	56	0.34	0.07	2.14		0	0	0
Rachel	PSC	HP Alpha SMP	128	0.31	0.50	6.00		19	35	0
		Total:	**107788**	**834.38**	**181.02**	**2987.94**		**1184**	**3040**	**2863**



TeraGrid HPC Systems plus Ranger

The TeraGrid partnership has developed a set of integration and federation policies, processes, and frameworks for HPC systems.

● Computational Resources (size approximate - not to scale)

TACC

Ranger in production since February 4

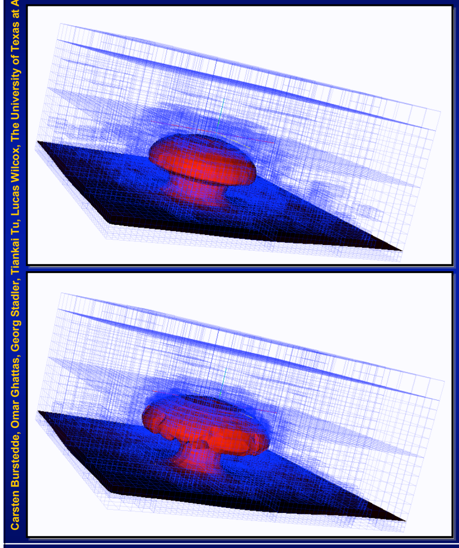
Installation is complete.
It's time to focus on Impact.

TACC

Impact on Science

- TeraGrid resources are available to all researchers at US institutions in all disciplines
- Ranger will enable researchers to attack problems heretofore much too large for TG
- Already seeing applications in astronomy, biophysics, climate/weather, earthquake modeling, CFD/turbulence, and more scale to 1000s of cores
- Just went into production on Monday Feb 4--much more to say very soon!

Early Research: Computing the Earth’s Mantle



Carsten Burstedde, Omar Ghattas, Georg Stadler, Tiankai Tu, Lucas Wilcox, The University of Texas at Austin

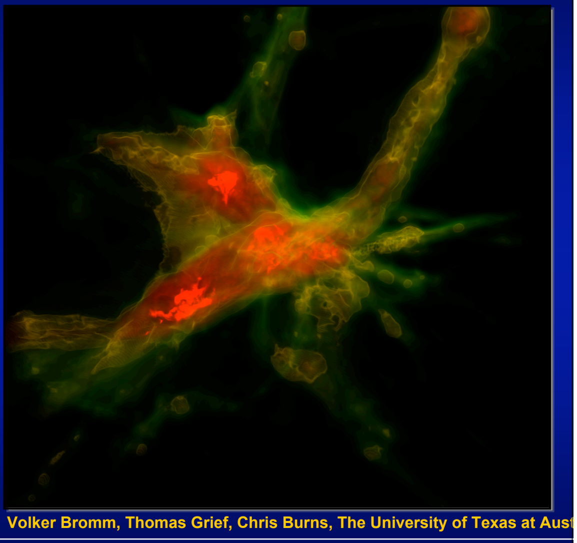
Omar Ghattas is studying convection in the Earth’s interior. He is simulating a model mantle convection problem. Images depict rising temperature plume within the Earth's mantle, indicating the dynamically-evolving mesh required to resolve steep thermal gradients.

Ranger's speed and memory permit higher resolution simulations of mantle convection, which will lead to a better understanding of the dynamic evolution of the solid Earth

Early Research: Researching the Origins of the Universe

Volker Bromm is investigating the conditions during the formation of the first galaxies in the universe after the big bang.

This image shows two separate quantities, temperature and hydrogen density, as the first galaxy is forming and evolving.



Volker Bromm, Thomas Grief, Chris Burns, The University of Texas at Aust

Ranger will operate for four years

Over 700 (and counting!) are <u>already</u> doing world-class science!

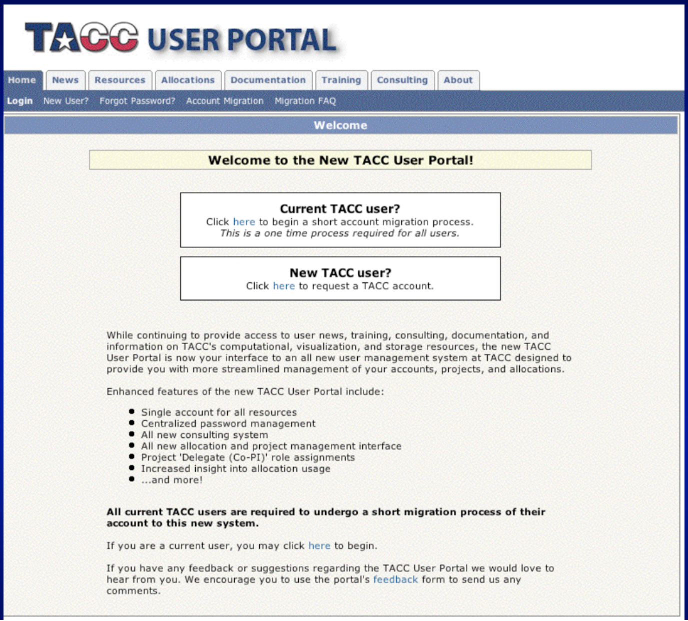
More to come

How Does This Help Texas?

- TACC may allocate *up to 5%* of the cycles (26M CPU hours!) to Texas higher ed institutions
  - User can still use as much of system at once as TG users
- Allocations requests must be submitted to TACC
- Review/decisions will be based on four criteria:
  - Research/education merit
  - Team capability/expertise for using system
  - Opportunity for impact in Texas
  - Level of support needed

How Do Texans Apply?

- Apply through the TACC User Portal:
  - http://portal.tacc.utexas.edu
- Future deadlines will be one month before beginning of quarter (March 1, June 1, September 1, December 1)
  - Next deadline in 31 days
- Instructions are on the TACC User Portal—click on ‘New User’



Welcome to the New TACC User Portal!

Current TACC user?
Click here to begin a short account migration process.
This is a one time process required for all users.

New TACC user?
Click here to request a TACC account.

While continuing to provide access to user news, training, consulting, documentation, and information on TACC's computational, visualization, and storage resources, the new TACC User Portal is now your interface to an all new user management system at TACC designed to provide you with more streamlined management of your accounts, projects, and allocations.

Enhanced features of the new TACC User Portal include:

- Single account for all resources
- Centralized password management
- All new consulting system
- All new allocation and project management interface
- Project 'Delegate (Co-PI)' role assignments
- Increased insight into allocation usage
- ...and more!

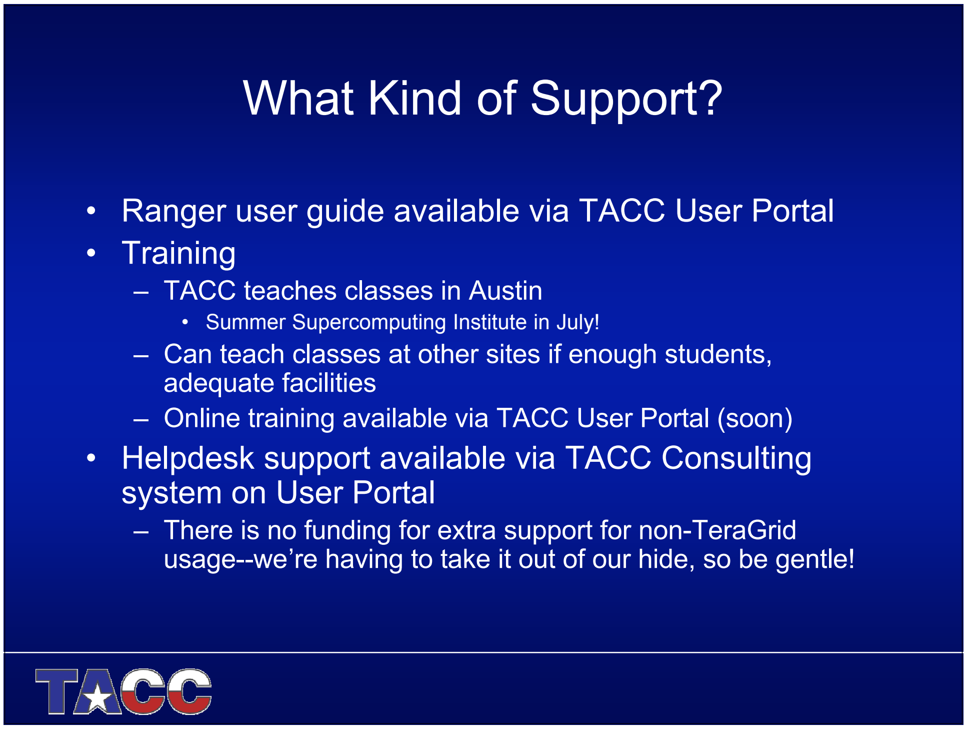
All current TACC users are required to undergo a short migration process of their account to this new system.

If you are a current user, you may click here to begin.

If you have any feedback or suggestions regarding the TACC User Portal we would love to hear from you. We encourage you to use the portal's feedback form to send us any comments.

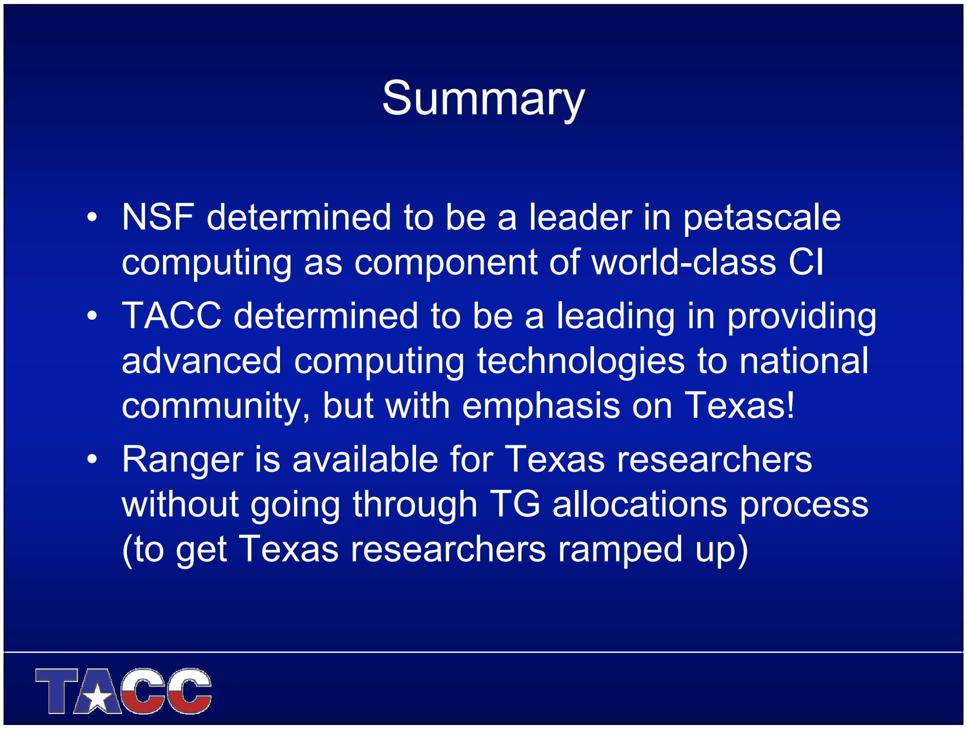
What Kinds of Allocations for Texans?

- Research
  - Default: Up to 500K CPU hours
  - Last for one year
  - Can request up to 1M by special arrangement
- Education
  - Up to 100K hours
  - Last for 2 quarters
- Startup
  - Up to 50K hours
  - Last for 1 quarter
  - Used for gaining expertise, preparing larger requests
  - May be repeated once



What Kind of Support?

- Ranger user guide available via TACC User Portal
- Training
  - TACC teaches classes in Austin
    - Summer Supercomputing Institute in July!
  - Can teach classes at other sites if enough students, adequate facilities
  - Online training available via TACC User Portal (soon)
- Helpdesk support available via TACC Consulting system on User Portal
  - There is no funding for extra support for non-TeraGrid usage--we’re having to take it out of our hide, so be gentle!



Summary

- NSF determined to be a leader in petascale computing as component of world-class CI
- TACC determined to be a leading in providing advanced computing technologies to national community, but with emphasis on Texas!
- Ranger is available for Texas researchers without going through TG allocations process (to get Texas researchers ramped up)