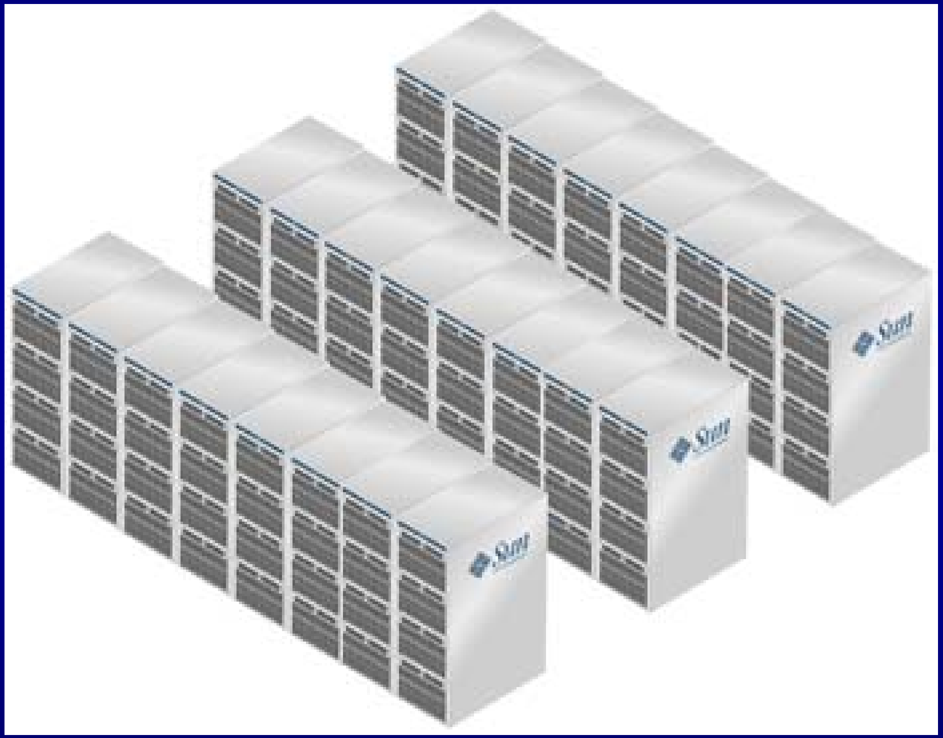
Ranger Update

Jay Boisseau, Director
Texas Advanced Computing Center

June 12, 2007

First NSF Track2 System: 1/2 Petaflop

- TACC selected for first NSF ‘Track2’ HPC system
  - $30M system
  - Sun is integrator
  - 15,744 quad-core AMD Opterons
  - 1/2 Pflop peak performance
  - 125 TB memory
  - 1.7 PB disk
  - ~2 μsec MPI latency
- TACC, ICES, Cornell, ASU supporting system, users four 4 years ($29M)



Ranger Configuration

- Compute power
  - 15744 Opteron “Deerhound” processors
    - Quad-core: 62,976 cores!
    - Four flops/cycle (dual pipelines) per core
  - 504 teraflops aggregate peak performance
- Memory
  - 2GB/core
  - 125 TB total memory
- Expandable
  - May add more compute nodes (may vary memory)
  - May add different compute nodes (GPUs?)

Ranger Configuration

[Most switch data non-disclosure until ISC'07]

- Interconnect
  - Sun proprietary switch based on IB
    - Minimum cabling: robustness and simplicity!
  - MPI latency: 2.3 μsec max latency
  - Peak bi-directional b/w: ~ 1 GB/sec
  - Peak bisection b/w: 7.9 TB/sec

Ranger Configuration

- File system
  - 72 Sun X4500s (“Thumper”)
    - 48 disks per 4U
  - 1.7 PB total disk
    - 3456 drives total
    - 1 PB in largest ***/work*** file system
  - Lustre file system
  - Aggregate b/w: 40 GB/s

Ranger Configuration

System Management

- ROCKS (customized) Cluster Kit
  - *perfctr* patch, etc.
- Sun N1SM for lights-out management
- Sun N1GE for job submission
  - Backfill, fairshare, reservations, etc.

Space & Power

- System power: 2.4 MW
- System space
  - ~80 racks
  - ~2000 sqft for system racks and in-row cooling equipment
  - ~4500 sqft total
- Cooling:
  - In-row units and chillers
  - ~0.6 MW
- Observations:
  - space less an issue than power (almost 3 MW)!
  - power generation distribution less an issue than distribution!

Project Timeline

Sep06	award, press, relief, beers
1Q07	equipment begins arriving
2Q07	facilities upgrades complete
3Q07	very friendly users
4Q07	more early users
Dec07	production, many beers
Jan08	allocations begin

User Support Challenges

- NO systems like this exist yet!
  - Will be the first general-purpose system at ½ Pflop
  - Quad-core, massive memory/disk, etc.
- NEW apps challenges, opportunities
  - Multi-core optimization
  - Extreme scalability
  - Fault tolerance in apps
  - Petascale data analysis
- System cost $50K/day--must do science every day!

User Support Plans

- User support: the “usual” +
  - User Committee dedicated to this system
  - 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
  - *Extensive* Training
    - On-site at TACC, partners, and major user sites, and at workshops/conferences
    - Advanced topics in multi-core, scalability, etc
    - Virtual workshops
  - Increased contact with users in TACC User Group

Technology Insertion Plans

- Technology Identification, Tracking, Evaluation, Recommendation are crucial
  - Cutting edge system: software won't be mature
  - Four year lifetime: new R&D will produce better technologies
- Chief Technologist for project, plus other staff
  - Must build communications, partnerships with leading software developers worldwide
  - Grant doesn't fund R&D, but system provides unique opp for determining, conducting R&D!

Technology Insertion Plans

- Aggressively monitor, and pursue:
  - NSF Software Development for Cyberinfrastructure (SDCI) proposals
  - NSF Strategic Technologies for Cyberinfrastructure (STCI) proposals
  - NSF Cyber-enabled for Discovery and Innovation (CDI) proposals (forthcoming)
  - Relevant NSF CISE propsals
  - Corresponding awards in DOE, DOD, NASA, etc.
- Some targets: fault tolerance, algorithms, next-gen programming tools/languages, etc.

Impact in TeraGrid, US

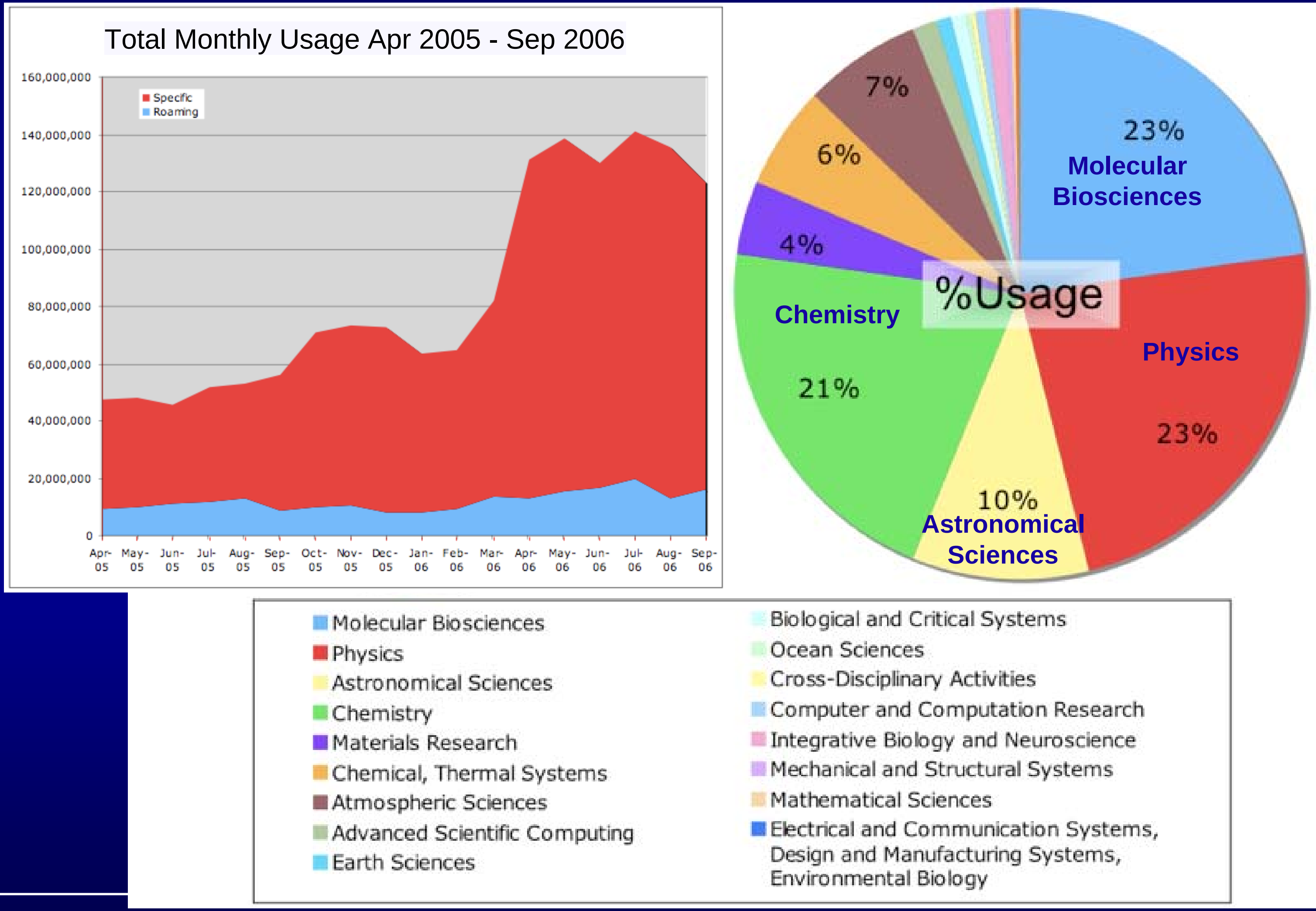
- 500M+ CPU hours to TeraGrid: more than double current total of all TG HPC systems
- 500+ Tflops : almost 10x current top system
- Enable unprecedented research
- Re-establish NSF as a leader in HPC
- *Jumpstarts progress to petascale for entire US academic research community*

TeraGrid HPC Systems plus Ranger

Computational Resources (size approximate - not to scale)

Who Might Use It?

Current TeraGrid HPC Usage by Domain



TACC

1000 projects, 3200 users

Some of the Big Challenges

- Scalable algorithms
- Scalable programming tools (debuggers, optimization tools, libraries, etc.)
- Achieving performance on many-core
  - Cray days of 2 reads & 1 write per cycle long gone
- Fault tolerance
  - Increased dependence on commodity (MTBF/node not changing) and increased number of nodes -> uh oh!

Some of the Big Challenges

- Data analysis ‘in the box’
  - Data will be too big to move (network, file system bandwidths not keeping pace)
  - Analyze in simulation if able, or at least while data still in parallel file system
- Power constraints (generation, distribution) limit number, location of petascale centers
  - but expertise becomes even more important than hosting expertise

TACC Strategic Focus Areas 2008+

- Petascale Computing
  - Integration, management, operation of very large systems for reliability and security
  - Performance optimization for multi-core processors
  - Fault tolerance for applications on large systems
  - Achieving extreme performance & scalability: algorithms, libraries, community codes, frameworks, programming tools, etc.
- Petascale Visualization & Data Analysis
  - ‘In-simulation’ visualization, HPC visualization applications
  - Remote & collaborative visualization
  - Feature detection and other tera/peta-sale analysis techniques
- Remote & collaborative usage of petascale resources
  - Tools for desktop & local cluster usage/integration
  - Portals for community apps to increase user base

Summary

- NSF again a leader in petascale computing as component of world-class CI, with solicitations for hardware, software, support, applications
- Ranger is a national instrument, a world-class scientific resource
- Ranger and other forthcoming NSF petascale systems (and software, and apps) will enable unprecedented high-resolution, high-fidelity, multi-scale, multi-physics applications

Advertisement: The University of Texas at Austin Distinguished Lecture Series in Petascale Computation

- Web accessible in real-time and archived: http://www.tacc.utexas.edu/petascale/
- Past Lectures include
  - “Petaflops, Seriously,” Dr. David Keyes, Columbia University
  - “Discovery through Simulation: The Expectations of Frontier Computational Science,” Dr. Dimitri Kusnezov, National Nuclear Security Administration
  - “Modeling Coastal Hydrodynamics and Hurricanes Katrina and Rita,” Dr. Clint Dawson, The University of Texas at Austin
  - “Towards Forward and Inverse Earthquake Modeling on Petascale Computers,” Dr. Omar Ghattas, The University of Texas at Austin
  - "Computational Drug Diagnostics and Discovery: The Need for Petascale Computing in the Bio-Sciences,” Dr. Chandrajit Bajaj, The University of Texas at Austin
  - "High Performance Computing and Modeling in Climate Change Science,” Dr. John Drake, Oak Ridge National Laboratory
  - "Petascale Computing in the Biosciences - Simulating Entire Life Forms," Dr. Klaus Schulten, University of Illinois at Urbana-Champaign
- Suggestions for future speakers/topics welcome