RICE

Condos, Resorts and Timeshares:
A Management Model for Scalable Campus Research Cyberinfrastructure

Jan E. Odegard
Ken Kennedy Institute (K2I)
odegard@rice.edu; 713.348.3128
http://k2i.rice.edu

Kim Andrews
Research Computing Support
kimba@rice.edu; 713.348.5726
http://rcsg.rice.edu

TAMU HPC Annual Users Meting; May 1, 2008

RICE

About
Ken Kennedy Institute
for Information Technology

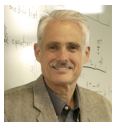
RICE

Vision & History

To build a community of scholars that engages in collaborative research and education covering virtually every aspect of information technology and computing

Institute Directors:

Ken Kennedy (1986-1992)



1945-2007

Sidney Burrus (1992-1998)

Willy Zwaenepoel (1998-2001)

Moshe Vardi (2001-…)

RICE

Research Institutes

Institutes are virtual organizations fostering communities focused on driving a research mission in key areas where Rice have or are developing research strength

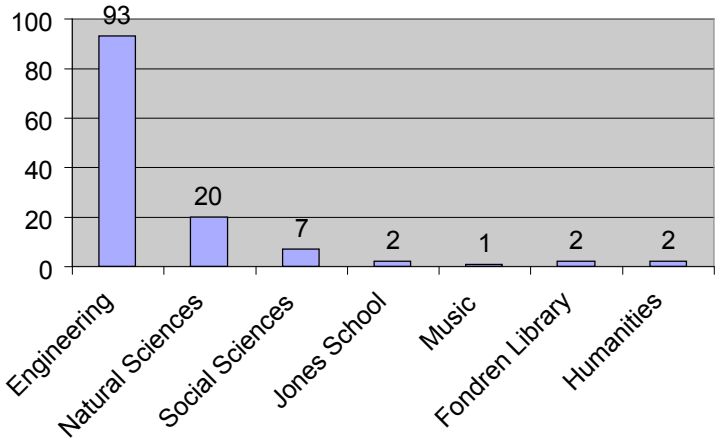
- Ken Kennedy Institute for Information Technology
- Energy and Environmental Systems Institute
- Smalley Institute for Nanoscale Science & Technology
- Institute for Bioscience and Bioengineering
- Rice Quantum Institute
- Rice Space Institute
- Humanities Research Center
- Baker Institute for Public Policy (BIPP)

Institutes are “Permanent” Units

Intellectual Community

7 "divisions" ↔ 18 departments ↔ ~130 members

6 centers ↔ ~15 ad hoc research groups



Research Centers

- Center for Multimedia Communication (CMC)
  - Director: Ashutosh Sabharwal, ECE
- Center for Computational Geophysics (CCG)
  - Co-directors: Bill Symes, CAAM / Alan Levander, ES
- Center for Computational Finance & Economic Systems (CoFES)
  - Director: Kathy Ensor, STAT
- LAboratory for NanoPhotonics (LANP)
  - Director: Naomi Halas, ECE
- Center for Technology in Teaching and Learning (CTTL)
  - Director: Tony Gorry, CS
- Center for Excellence and Equity in Education (CEEE)
  - Director: Richard Tapia, CAAM

- Retired centers
  - Center for High Performance Software
  - Center for Chemical Processing Technology

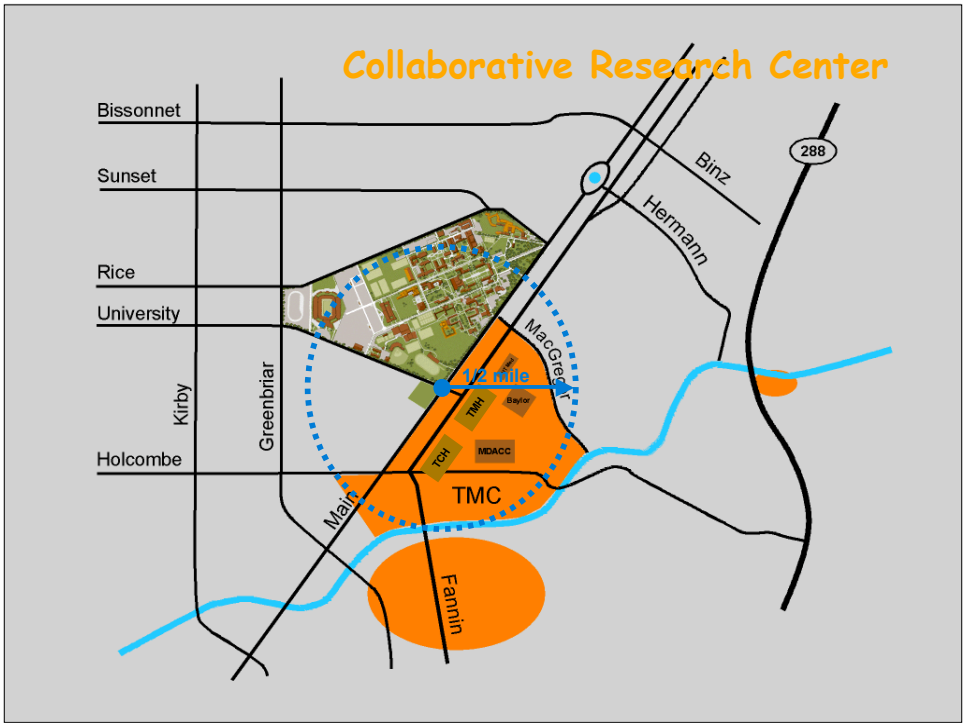
Centers are "Non-Permanent" Units

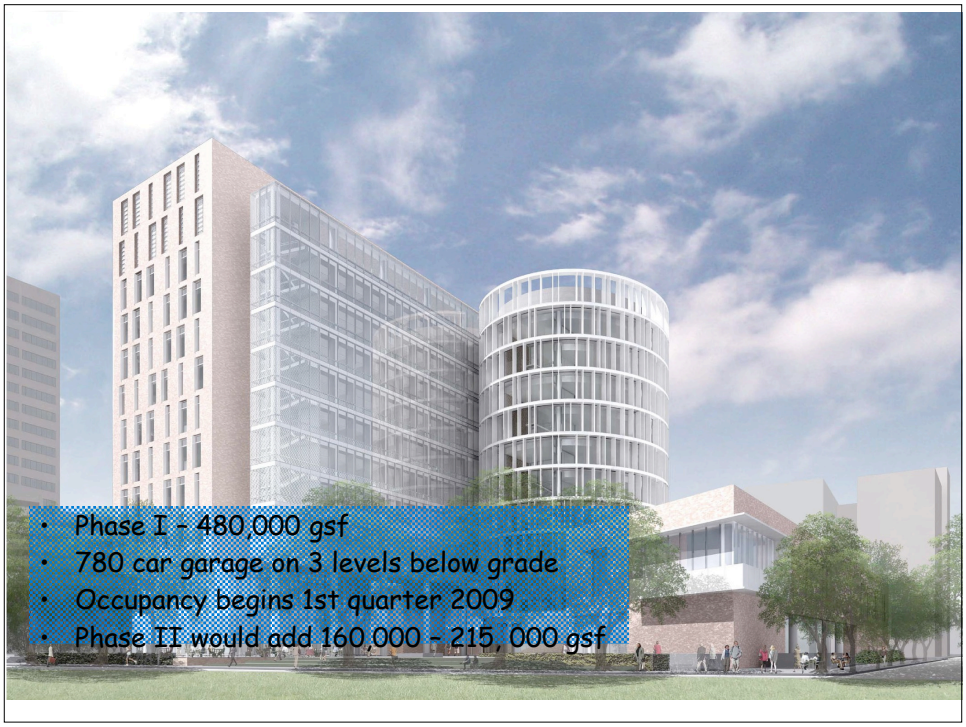
RICE

Research Groups & Labs

- Habanero Multicore Software Project
- Rice Computer Architecture Group
- Gaming Group
- Robotics Group
- Sensor Nets Group
- Bioinformatics Group
- Rice Networking Group
- Digital Signal Processing
- Dynamical Systems Group
- Statistical Consulting Lab
- Complex Flow of Complex Fluids Group
- Theoretical and Computational Neuroscience
- Connexions: Open content education repository
- Advanced Research Initiative on the Emerging Library
- …

Informal & ad hoc Research Collaborations



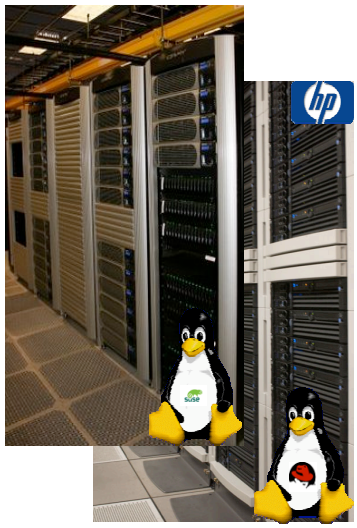


RICE

About
HPC Infrastructure and Usage
at Rice today

RICE

Research Cyberinfrastructure



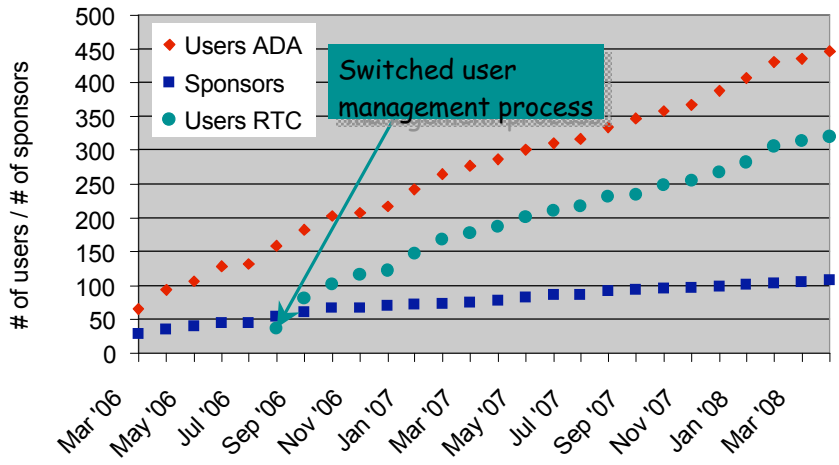
- Rice Terascale Cluster (aka RTC)
  - ✓ HP Integrity Itanium® Linux Cluster
  - ✓ 1 TeraFLOP (peak)
  - ✓ MRI Grant: September 2002
  - ✓ PO: October 2002
  - ✓ Delivered January 2003
  - ✓ Production: June 2003

Rice Computational Research Cluster (aka Ada)

- ✓ Cray XD1 Opteron® Linux Cluster
- ✓ 3 TeraFLOP (peak)
- ✓ MRI Grant: September 2004
- ✓ PO: July 2005
- ✓ Delivered: November 2005
- ✓ Production: March 2006

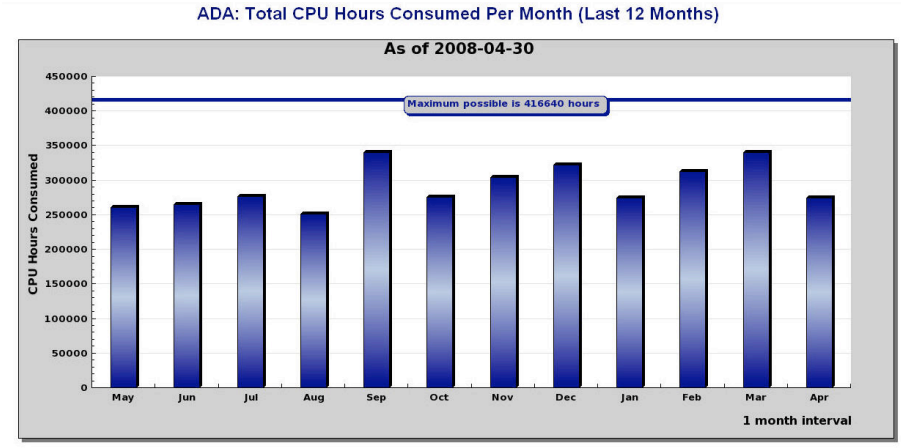
RICE

HPC Users



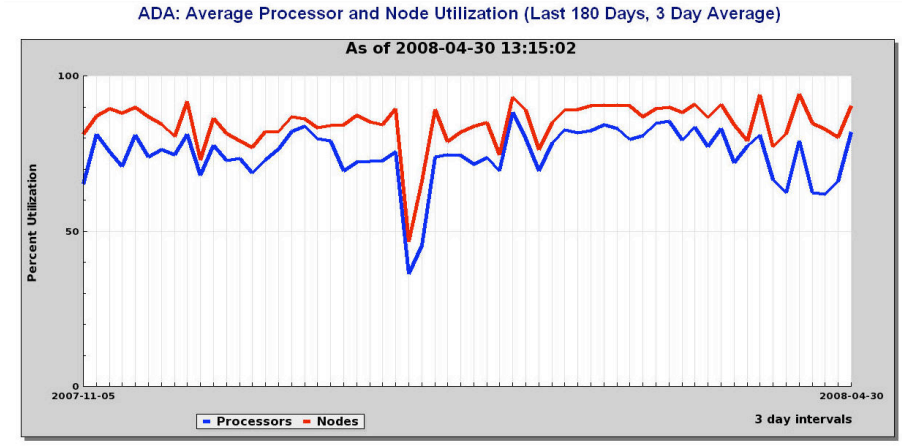
RICE

ADA CPU hours Delivered



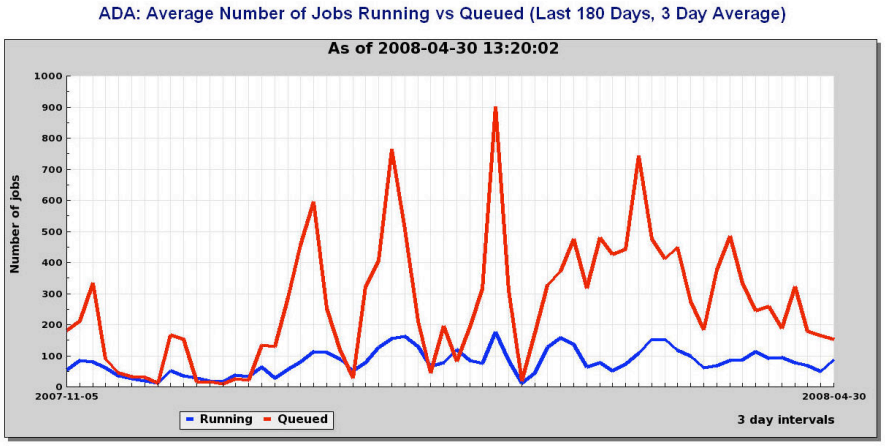
RICE

ADA 180 day Utilization



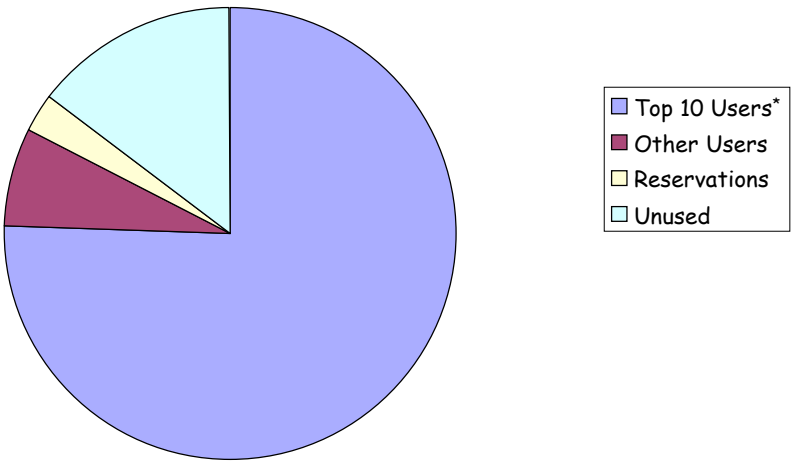
RICE

ADA 180 day Workload



RICE

ADA Users Pattern: April 2008

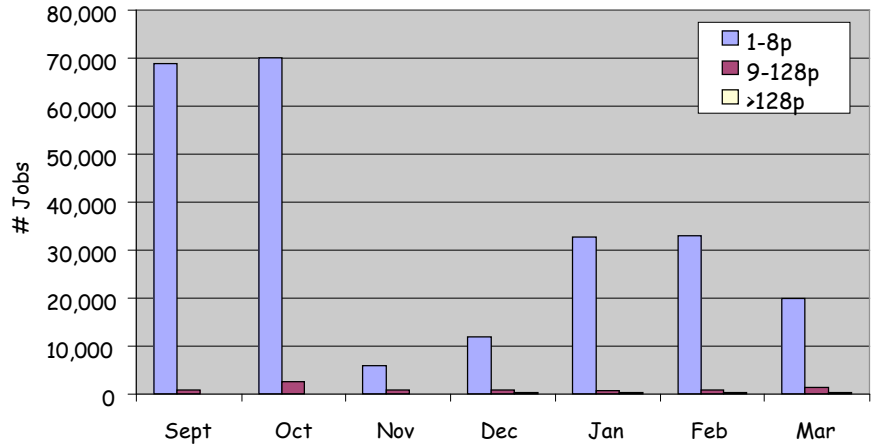


- Total number of users for period: 70 (37 sponsors)
- 332,504/455,040 (73%)

* Includes phyz queue

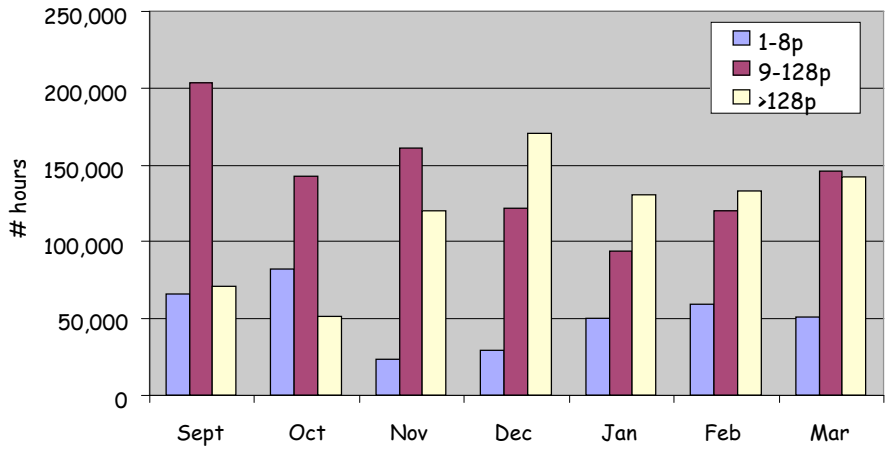
RICE

ADA Workload - Jobs



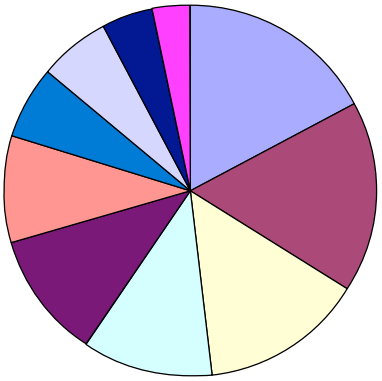
RICE

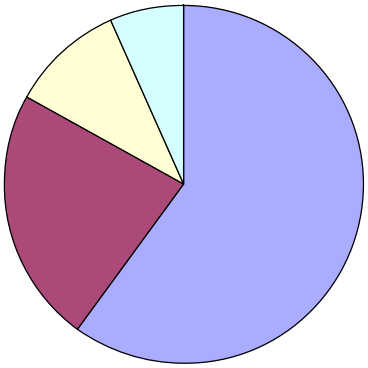
ADA Workload - Hours

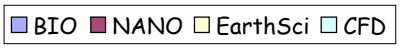


RICE

ADA Top 10 Users

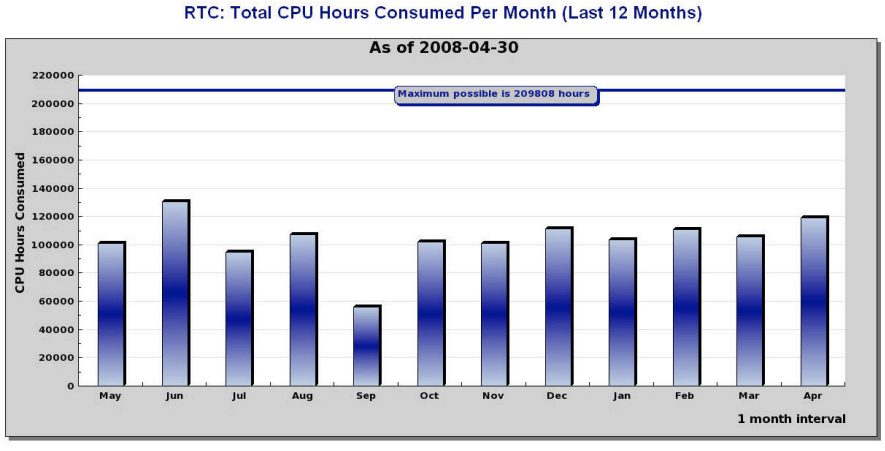






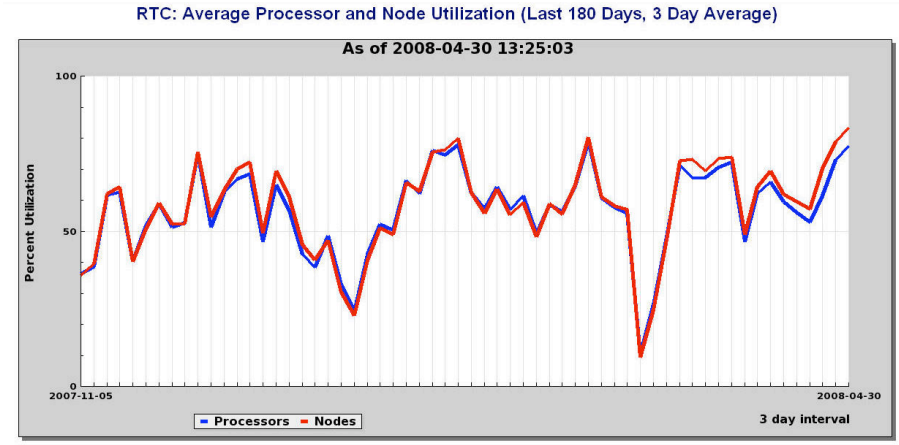
RICE

RTC CPU hours Delivered



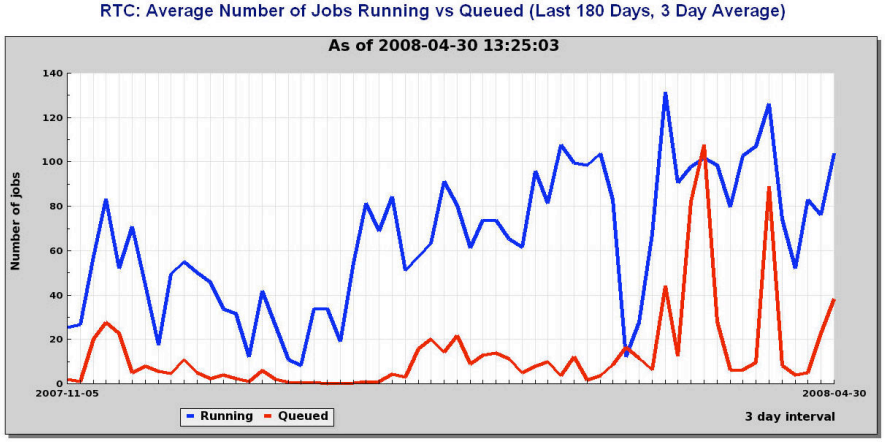
RICE

RTC 180 day Utilization

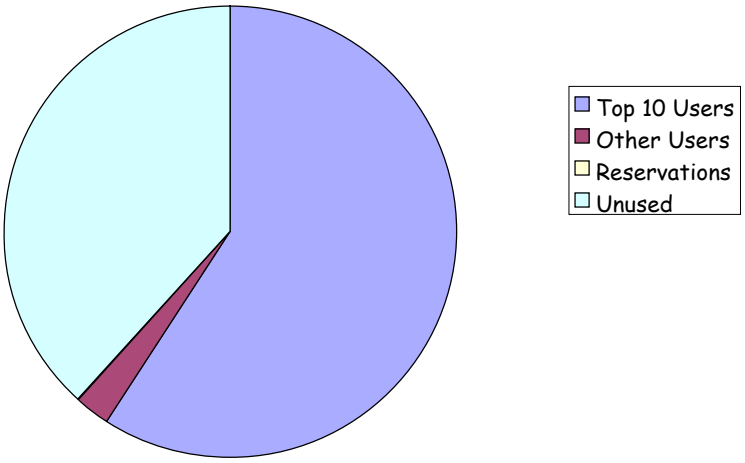


RICE

RTC 180 day Workload

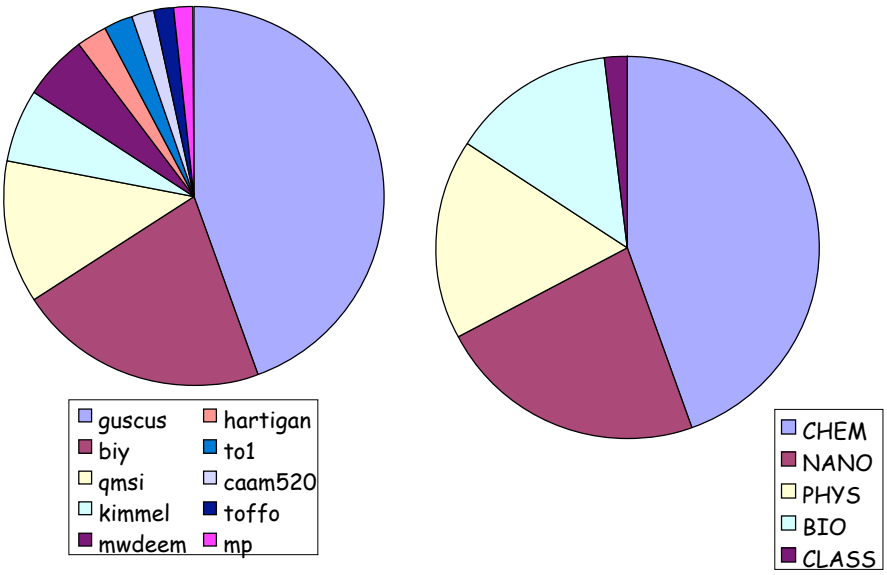


RTC Users Pattern : April 2008



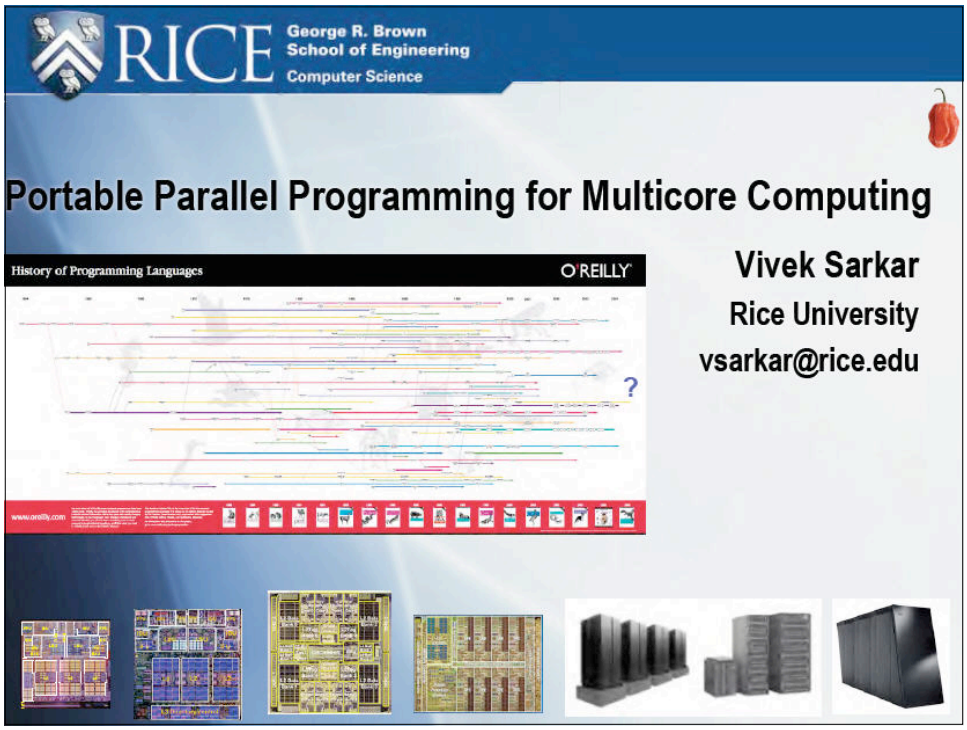
- Total number of users for period: 51 (23 sponsors)
- 123,478/200,160 (62%)

RTC Top 10 Users

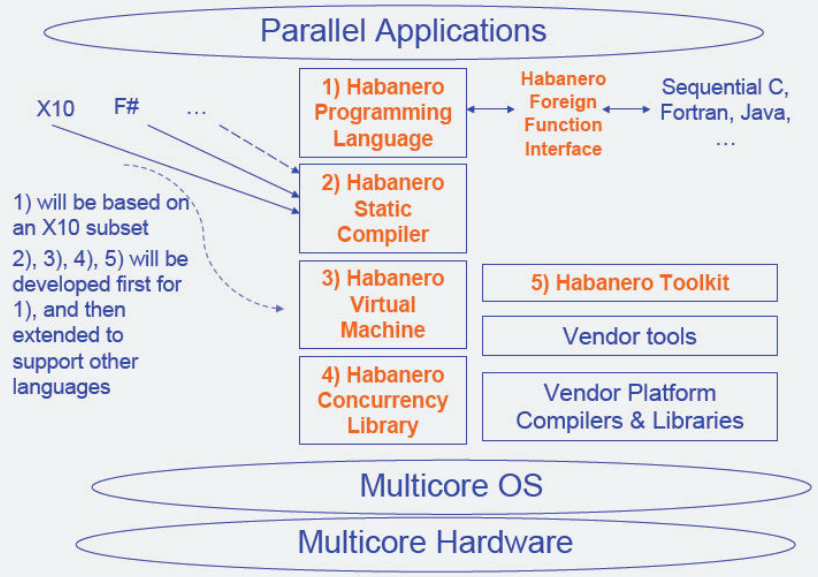


Advanced Computing Technology

- Collection (sanbox) of bleeding edge technology
- Purpose: benchmarking, testing & planning
- **Not** production computing
- Current in-house resources
  - 4x Nvidia Tesla S870
  - Sun UltraSparc T1, T2
  - ClearSpeed Advance X620 accelerator
  - DRC coprocessor module w/ Xilinx Virtex-4 FPGA
  - Cray XD1 w/6 Xilinx Virtex-4/LX160 FPGA
  - 1x4S Quad core Barcelona system
  - 2x2S Quad core Barcelona systems
- Remote accessible resources
  - STI Cell
  - IBM Cyclops (C-64)
  - IBM Power5+, Power6
  - Blue Gene
  - SiCortex



Habanero Project (habanero.rice.edu)



RICE

5

Habanero Target Applications and Platforms

Applications:

Parallel Benchmarks
- SSCA's #1, #2, #3 from DARPA HPCS program
- NAS Parallel Benchmarks
- JGF, JUC, SciMark benchmarks

Medical Imaging
- Back-end processing for Compressive Sensing (www.dsp.ece.rice.edu/cs)
- Contacts: Rich Baraniuk (Rice), Jason Cong (UCLA)

Seismic Data Processing
- Rice Inversion project (www.trip.caam.rice.edu)
- Contact: Bill Symes (Rice)

Computer Graphics and Visualization
- Mathematical modeling and smoothing of meshes
- Contact: Joe Warren (Rice)

Computational Chemistry
- Fock Matrix Construction
- Contacts: David Bernholdt, Wael Elwasif, Robert Harrison, Annirudha Shet (ORNL)

Habanero Compiler
- Implement Habanero compiler in Habanero so as to exploit multicore parallelism within the compiler

Platforms:
- AMD Opteron Quad-Core
- Clearspeed Advance X620
- DRC Coprocessor Module w/ Xilinx Virtex FPGA
- IBM Cyclops-64 (C-64)
- IBM Power5+, Power6
- Intel Xeon Quad-Core
- NVIDIA Tesla S870
- STI Cell
- Sun UltraSparc T1, T2
- . . .

Additional suggestions welcome!

RICE

6

RICE

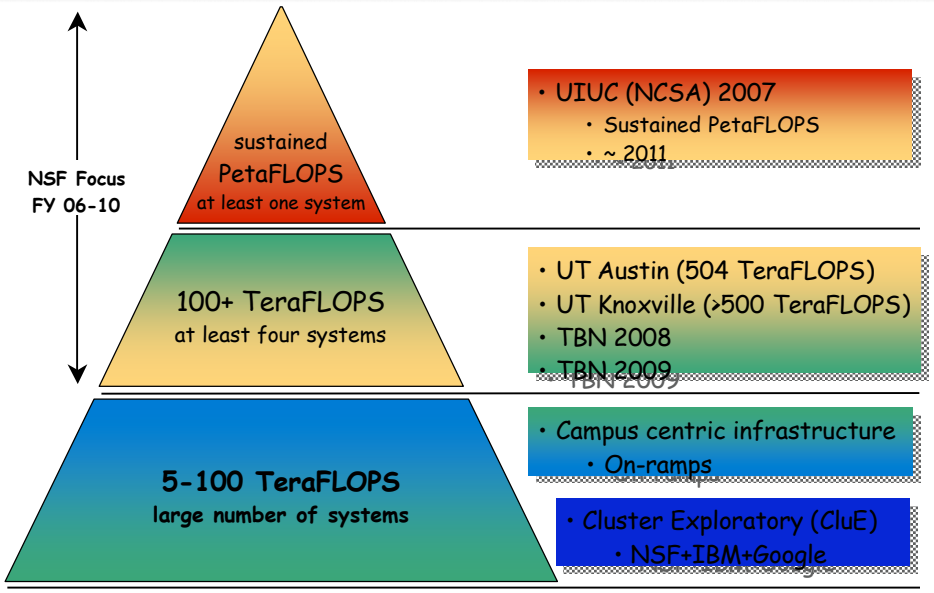
About
HPC Infrastructure and Usage
at Rice tomorrow

RICE

Understanding Challenges

- Funding
  - Have invested ~$2,000,000 per 3 years in HW
  - NSF is not sustainable funding model for campus
  - NSF MRI program awarded <10% for computing
  - Small group and individual awards still fund some hw
  - Charging for cycles …
    - Can we do it?
      - No, if the system is funded by federal grant
      - Yes, if system funded by private funds
    - Does it work?
      - No, no & no
  - OMB A-21: Cost Principles for Educational Institutions
- Barriers to entry
- Total cost of ownership
- Sustainability & life cycle management

National Computing Landscape



Campus HPC Condominium

- Focus on capacity not "capability"
  - Capacity systems are easier to grow incrementally
  - Capability systems are difficult to grow incrementally
- Develop the **resort**
  - Hire staff (management and operation)
  - Build out the core (space, power & cooling)
  - Core facilities (utility-nodes, storage & system software)
  - Provide funding for minimal shared compute core
  - Provide funding for end-user software
- Market & sell **condos** to interested partners
- Offer visitors and partners "**timeshares**"

RICE

Primary Data Center

RICE

Leverage & TCO

- Typical sources of partner funding
  - Faculty startup funds
  - Investigator awards w/small equipment budget
- Invest scarce university resources to
  - Minimize **total cost of ownership**
  - Minimize compute resource **idle time**
  - Maximize **# beneficiaries**

Partner terms

- Partners:
  - Can set and manage (within specified extremes set by resort) the queuing policy for own condo
  - Benefits from access to more capacity than they have funding to support private (shared-pool policy apply)
  - Benefits from access to a managed system (resort is fully staffed)
  - Must be willing to give up (permanently share) part ($x or % of nodes in condo) of their total investment for access to the **resort**
  - Must be willing to share compute cycles (to other partners and/or visitor) in vacant condo
    - Partner usage pattern may make this a good or bad deal for the resort

Visitor terms

- Visitors:
  - Benefits from access to shared resources
  - Benefits from access to timeshares (vacant condos)
  - Will be able to finish any job(s) running on vacant condos before resources are handed back to partners
  - Benefits from access to a managed system (resort is fully staffed)
  - Must be willing to live with universal queuing policy and a smaller compute resource during high partner occupancy

The Hard Condo Problems

- How much of a share will be allocated to the condo versus the common pool?
  - 20/80, 25/75, 30/70
- Do partners pay less to common pool if they are willing to wait longer to get access to their condo?
  - 4, 8, 12, 24 hours
- What is the appropriate run time for jobs in common pool?
  - 8, 12, 24, 48 hours
  - Impact: wait time, fair share, …
- How long can a partner be permitted to run a single job inside a condo?
  - Impact: resort manageability, risk & "good occupant"
- Should there be a minimal condo unit size?
  - 4, 8, 16 (nodes)?
- Do we limit MPI on the resort/condo?

"HPC" Resources (12 months ago)

Manufacturer	Architecture	OS	CPU	Memory (GB)	Local Disk (TB)	% tot
Cray	AMD Opteron	SuSE	672	1400	26	45%
HP	Intel Itanium	RedHat	278	608	8	19%
Linux Networx	Intel Xeon	RedHat	84	84	1.2	6%
Appro	Intel Xeon	RedHat	150	300	1	10%
Appro	AMD Athlon	RedHat	84	256	1	6%
Appro	AMD Athlon	RedHat	44	44	0.5	3%
Appro	AMD Opteron	RedHat	32	64	4	2%
PSSC	Intel Xeon	RedHat	24	48	1.5	2%
Appro	Intel Xeon	RedHat	24	48	1.5	2%
Appro	AMD Athlon	RedHat	20	40	1	1%
PSSC	AMD Opteron	Fedora	20	40	0.5	1%
IBM	Regatta	AIX	16	32	1	1%
IBM	Regatta	AIX	16	64	1	1%
Chaogic	AMD Athlon	RedHat	8	8	0.2	1%
IBM	Power5	AIX	4	8	1	0%
Dell	PowerEdge 6850	Windows	4	16	0.5	0%
Total for all Compute Clusters			**1480**	**3060**	**56.1**	**100%**

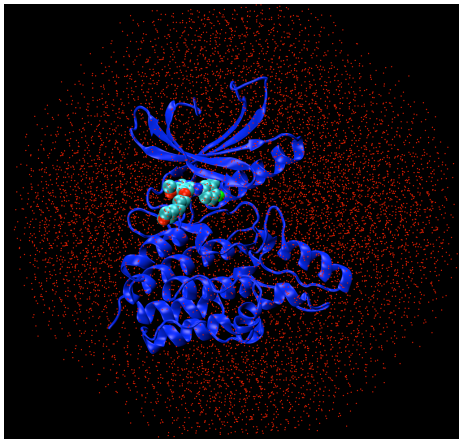
Shared University Grid @ Rice

SUG@R

- 123 (107+16) nodes
- SunFire x4150 (dual soc)
- Intel Xeon E5440
  - 2.83 GHz (quad core)
  - 2GB/core (16GB/node)
- GigE interconnect
- ~ 10 TeraFLOPS
- 19 TB Panasas
- 12 TB NFS (iSCSI)
- 60 TB Coraid (Disaster Rec.)
- RedHat linux
- Friendly users NOW
- Production: ~May 20
- (Sigma Solutions / Sun)

Molecular Dynamics of Proteins

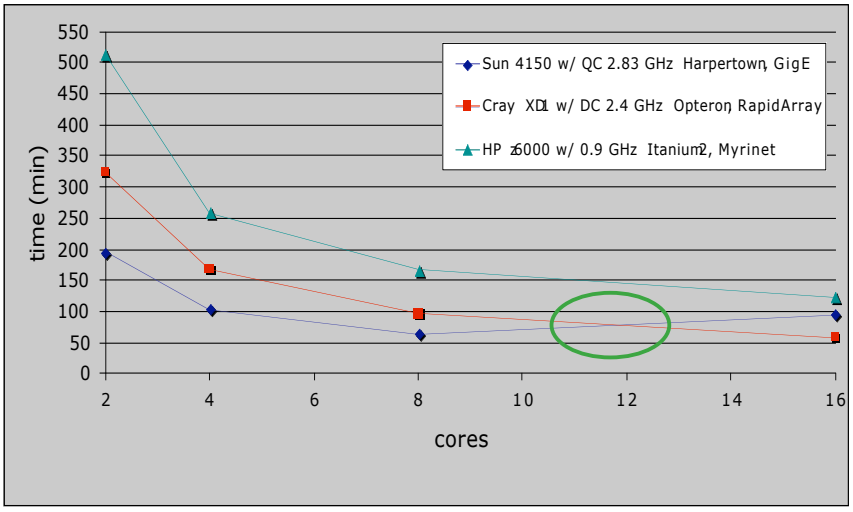
- Compute type and strength of interactions between kinase and drug through time (20-40ns)
- Calculate binding free energy of the drug from MD simulation.
- Amber 9
  - Pmemd: parallel version to do MD for systems with explicit water
- System (31908 atoms)
  - GAK kinase (4628 atoms)
  - inhibitor (55 atoms)



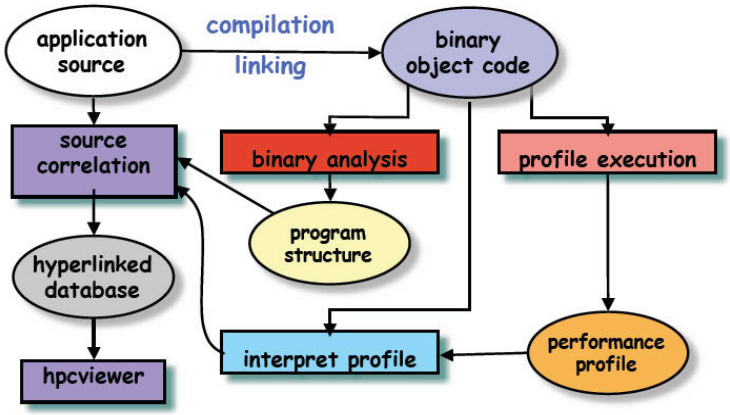
MD of G-associated **kinase** (GAK) in complex with inhibitor (Iressa) for treatment of lung cancer

RICE

Code Performance



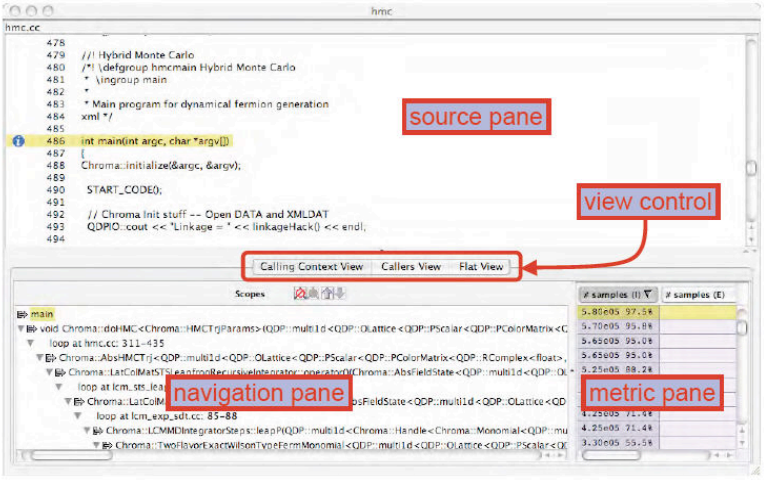
HPCToolkit Workflow



http://www.hipersoft.rice.edu/hpctoolkit/

6

hpcviewer User Interface



21

RICE

Summary

- Condominium offers campus solution for
  - Leveraging institutional HPC investments
  - Life cycle management
  - TCO management
  - Resource cycle (3 staggered resorts):
    - Old, slated for decommissioning (still large utilization)
    - Primary production system
    - New system entering production
- Track 1 & 2 (and others), national solution for
  - Recourse needs beyond local capacity/capability
  - Tight integration critical
  - Local staff for user support critical
- Visualization - not addressed at Rice yet