Advancing Science through Local, Regional, and National Cyberinfrastructure

Daniel S. Katz
d.katz@ieee.org
Senior Fellow, Computation Institute, University of Chicago & Argonne National Laboratory
Affiliate Faculty, Center for Computation & Technology, Louisiana State University
Adjunct Associate Professor, Electrical and Computer Engineering, LSU
Louisiana

- Area: 134,382 km² (33/51) - VIC: 237,629 km²
- Population: 4,533,000 (2010, 25/51) - VIC: 5,300,000 (2008)
- BS Degree: 21% (2009, 47/51) - AUS: 24% (2006)
- Advanced Degree: 7% (2009, 48/51)

State Goals: talented workforce, great competitiveness, strong educational system, increased economic development
PITAC Report Summary:

• “Computational science -- the use of advanced computing capabilities to understand and solve complex problems -- is critical to scientific leadership, economic competitiveness, and national security. It is one of the most important technical fields of the 21st century because it is essential to advances throughout society.”

• “Universities must significantly change organizational structures: multidisciplinary & collaborative research are needed [for US] to remain competitive in global science”

Complex problems: Innovations will occur at boundaries
Challenges of Complex Problems
“Third Pillar” of Comp. Science

• Applications

• Communities
  – No single group, university, or state can do these problems
  – Must integrate CS, Math, Bio, Sensors, Engineering, more...

• Data everywhere
  – Supercomputers generate petabytes
Cyberinfrastructure

- Term first documented in 1998
- Cyberinfrastructure
  - computing systems
  - data storage systems
  - advanced instruments and data repositories
  - visualization environments
  - people
  - linked together by software and high performance networks
  - to improve research productivity and enable breakthroughs not otherwise possible.
    - Indiana University Cyberinfrastructure Newsletter, Craig Stewart
- Used for e-Science, e-Research
Gravitational Wave Astronomy

- LIGO: Laser Interferometric Gravitational Wave Observatory
- Ties together theory, computation, and experiment
  - Each drives the other two!
Hurricane Katrina
Models
How We Started

• State commitment: $25M/year for Vision 20/20
  – $9M: LSU -> CCT (similarly, ULL -> LITE)
• University commitment to build new programs for 21st century
• State and University willingness to make extraordinary investments
• Opportunity to build new world class program in interdisciplinary research and education, involving all of LSU
• Ed Seidel-led vision to instigate state-wide collaboration
Advancing Research

- Potentially requires advances in three areas, depending on existing strengths.
CCT Organization

CCT

Director Office
Edward Seidel

Corporate Relations

HPC Partnership
McMahon

Cyberinfrastructure Development
Katz

Focus Areas
Allen

Focus Areas
Katz

HPC Partnership
McMahon

Cyberinfrastructure Development
Katz

Focus Areas
Allen

LSU HPC

Performance Team

Core Comp. Sci.

LONI

Systems and Software

Coast to Cosmos

LSU HPC

Performance Team

Core Comp. Sci.

NSF TeraGrid

Systems and Software

Coast to Cosmos

Blue Waters, etc.

Visualization

Cultural Computing

Labs: ACAL, DSL, Viz, LCAT, …

Material World
Cyberinfrastructure Development

• Vision: combine research and infrastructure
  – Research
    o Computer science
    o Applications
    o Tools
  – Infrastructure
    o Hardware
    o Operations
    o Policies

• Both together have squared growth of either alone

• CyD staff – PhDs in CS and apps who understand the whole picture and want to grow the ecosystem
Computing in Louisiana

- LA Tech
- UL-L
- SUBR
- LSU
- UNO
- Tulane
- Shreveport
- Ruston
- Alexandria
- Lake Charles
- Lafayette
- Baton Rouge

**TeraGrid, OSG**

**LONI Institute: People and Collaborations**

**Cybertools: Tools and Services**

**LONI: ~100TF IBM, Dell Supercomputers**

**LONI: 40 Gbps network**

**National Lambda Rail**
LONI - Networking & Computing

Network: partners and customers

- LSU HSC
- La Tech
- ULM
- NSU
- Alex
- SU
- LSU
- ULL
- McNeese
- SLU
- LSU HSC
- UNO
- Tulane

LONI node
- Multiple 10GE
- ~500 core Dell cluster & 112 proc. IBM P5 cluster
- ~4500 core Dell Cluster
LONI Computing Resources (2010)

- One central Dell cluster (Queen Bee)
  - 5500 IB-connected cores at ISB in Baton Rouge
  - Archival storage contracted through NCSA
  - 50% of allocations dedicated to TeraGrid from 2008
- Six distributed 512-core Dell clusters
- Five distributed 14-node (112 procs) IBM P5-575 clusters
- Distributed PetaShare storage
  - 32 TB disk @ each small Dell cluster
  - 8 TB disk on LSU & LaTech small Dell clusters – for LBRN
  - 8 TB at SC-S & HSC-NO – for LBRN
  - 250 TB tape
- All run by HPC@LSU, including user support/training
$12M NSF CyberTools Project: Enabler and Driver

- Geno/immuno sensors
- Small molecule sensors
- Bio-transport
- Environmental transport

Biosensor Development
- Soper

Science Drivers

CyberTools
- Seidel

WP1: Data, Scheduling
WP2: Information
WP3: Visualization
WP4: Application

science cybertools people

undergraduate

graduate

postdoctoral

faculty

www.ci.anl.gov
www.ci.uchicago.edu
Cactus

- Component-based HPC framework
  - Freely-available environment for collaborative application development
- Cutting edge CS
  - Grid computing, petascale, accelerators, steering, remote viz
- Active user & developer communities
  - 10 year pedigree, >$10M support
  - Numerical Relativity, CFD, Coastal, Reservoir Engineering, …
- Domain-specific toolkits, e.g. CFD toolkit
  - FD/FV/FE numerical methods
  - Structured, multi-block, unstructured
  - Uses PETSc, Trilinos, MUMPS, HYPRE
  - Used to build Black Oil Toolkit
PetaShare

- Main concept: data is managed (migrated, moved, replicated, cached, etc.) automatically
- Data-aware storage systems, data-aware schedulers, cross-domain metadata scheme
- Provides: 250 TB disk, 400 TB tape storage (and access to national storage facilities)
- Applications: coastal & environmental modeling, geospatial analysis, bioinformatics, medical imaging, fluid dynamics, petroleum engineering, numerical relativity, high energy physics.

Credit: Tevfik Kosar
LONI Institute
“CCT for the Louisiana”

• $15M 5-year project
  – $7M BoR, $8M from LaTech, LSU, SUBR, Tulane, UNO, ULL

• Catalyzes new inter-institutional collaborations, ambitious projects and top level hires:
  – LONI network and computing
  – NSF projects: PetaShare, VizTangibles, TeraGrid, Blue Waters
  – EPSCoR: NSF CyberTools, DOE UCoMS, DoD
  – NIH: $17M LBRN
  – Promote collaborative research at interfaces for innovation
LONI Institute Vision

• LONI investments create world leading infrastructure
• Create bold new inter-university superstructure
  – New faculty, staff, students; train others. Focus on CS, Bio, Materials, but all disciplines impacted
  – Promote research at interfaces for innovation
• Draw on, enhance strengths of all universities
  – Strong groups recently created; collectively world-class
  – Solve complex problems through collaboration & computation
  – Much stronger recruiting opportunities for all institutions
  – Statewide interdisciplinary education & research program
• Create University-Industry Research Centers (UIRCs)
  – Research Triangle, NCSA/UIUC, Bay Area, others
• Transform Louisiana
  – Such committed cooperation between sites extraordinary
LONI Institute Hiring and Projects

- Two new faculty at each institution (12 total)
  - Six in CS, six in Comp. Bio/Materials
- Six Computational Scientists
  - Following Bavarian KONWIHR project
  - Support 70-90 projects over five years; lead to external funding
- Graduate students
  - 36 new students funded, trained; two years each
- One Coordinator/economic development
- All hiring coordinated across state
- Leading faculty across state create multi-institutional seed projects
- Building on seeds, dozens of new projects selected, started
- Exploit common themes, computing environments, tools found in all areas
TeraGrid (XSEDE)

• TeraGrid: world’s largest open scientific discovery infrastructure
• Leadership class resources at eleven partner sites combined to create an integrated, persistent computational resource
  – High-performance networks
  – High-performance computers (>1 Pflops (≈100,000 cores) -> 1.75 Pflops)
    o And a Condor pool (w/ ~13,000 CPUs)
  – Visualization systems
  – Data Collections (>30 PB, >100 discipline-specific databases)
  – Science Gateways
  – User portal
  – User services - Help desk, training, advanced app support
• Allocated to US researchers and their collaborators through national peer-review process
  – Generally, review of computing, not science
• Mid 2011: TeraGrid --> XSEDE
Campus Champions

- “Champion” is a staff or faculty member on a campus that provides information on XSEDE to his/her colleagues
- Currently 114 institutions represented by champions
- Receive training and support from XSEDE staff

Current Campus Champion Institutions (unclassified) – 53
Current Campus Champion Institutions (EPSCoR state) – 30
Current Campus Champion Institutions (Minority Serving Institutions) – 7
Current Campus Champion Institutions (both EPSCoR and MSI) – 5
Total Number of Campus Champion Institutions Overall -- 102

Credit: Scott Lathrop (9/2011)
LONI and National Cyberinfrastructure

• TeraGrid
  – One of the 11 TeraGrid Resource Providers
  – Playing a role in TG-wide governance (TeraGrid Forum, Executive Steering Committee, various working groups, GIG Director of Science)
  – Contributed administrative software AmieGold (glue between TG account info and local info) and CS software (HARC, PetaShare, SAGA)

• OSG
  – Currently providing resources

• XSEDE
  – LONI not a partner in XSEDE, but a service provider

• Nationally
  – Bringing in new users from the southeast US
  – LONI Institute Computational Scientists -> Campus Champions
Recap (to 2010)

• Louisiana decides that science and technology can lead to a better future
• Builds a regional cyberinfrastructure (network, computing, software, ~data, people) that connects to national-scale infrastructure
• Starts to change culture – infuse computation in academic departments, interdisciplinary hiring, large collaborative projects
Lessons

• Three triangle facets (infrastructure, computational, interdisciplinary) have been taken seriously at highest levels, seen as important component of academic research

• Infrastructure need to be integrated at all levels (laboratory, campus, regional, national, international) – users need to be able to easily move work and data to appropriate systems, and collaborate across locations

• Education and training of students and faculty is crucial – vast improvements are needed over the small numbers currently reached through HPC center tutorials; computation and computational thinking need to be part of new curricula across all disciplines

• Emphasis should be made on broadening participation in computation, not just focusing on high end systems where decreasing numbers of researchers can join in, but making tools much more easily usable and intuitive and freeing all researchers from the limitations of their personal workstations, and providing access to simple tools for large scale parameter studies, data archiving, visualization and collaboration

• Vision needs to be consistent – cannot be just one person
• Funding needs to be stable (activities need to be sustainable)
Sources

  – authors from Louisiana State University, Tulane University, University of Louisiana at Lafayette, Louisiana Tech University, Louisiana Community and Technical College System, Southern University, University of New Orleans

• G. Allen and D. S. Katz, “Computational science, infrastructure and interdisciplinary research on university campuses: experiences and lessons from the Center for Computation and Technology,” NSF Workshop on Sustainable Funding and Business Models for Academic Cyberinfrastructure Facilities, Cornell University, 2010
eScience Conference main events  Wednesday – Friday
(keynotes, papers, panels, posters)

Microsoft eScience Workshop  Monday – Tuesday
Additional eScience workshops  Monday – Tuesday
Open Grid Forum OGF36  Monday – Wednesday
GLIF Annual Meeting  Thursday – Friday

Conference Tracks:
eScience Algorithms and Applications
- eScience application areas, including:
  - Physical sciences
  - Biomedical sciences
  - Social sciences and humanities
- Data-oriented approaches and applications
- Compute-oriented approaches and applications
- Extreme scale approaches and applications

Cyberinfrastructure to support eScience
- Novel hardware
- Novel uses of production infrastructure
- Software and services
- Tools

General Chair  Ian Foster
Program Co-Chairs  Daniel S. Katz, Heinz Stockinger
Program Vice Co-Chairs  David Abramson, Gabrielle Allen,
  Rosa M. Badia, Geoffrey Fox
Early Results and Works-in-Progress Posters Chair  Roger Barga
Workshops Chair  Ruth Pordes
Sponsorship Chair  Charlie Catlett
Conference Manager and Finance Chair  Julie Wulf-Knoerzer
Publicity Chairs  Kento Aida, Ioan Raicu, David Wallom