Scott Lathrop
NCSA Blue Waters Technical Program Manager for Education
TeraGrid Area Director for Education, Outreach and Training
scott@ncsa.uiuc.edu
www.hpcuniv.org

Oklahoma Supercomputing Symposium 2008
October 7, 2008
What is HPC University?

A **Virtual Organization** of people and institutions committed to preparing knowledgeable & skilled HPC (Digital Services) professionals, researchers, educators, and students.

| Assess learning and workforce development needs | Provide a catalog of high quality, peer-reviewed training and education resources |
| Assess requirements of a large and diverse computing community of practice | Provide a calendar of live and synchronous events |
| Define a roadmap for acquiring HPC competencies according to needs and requirements. | Identify gaps to drive development of new EOT resources |
| | Open a forum for community sharing |
| | Incorporate HPC across all disciplines |
Who is Involved?

• TeraGrid Resource Providers
• NCSA Blue Waters project
• Department of Energy HPC Centers
• State and regional HPC Centers
• Current and potential HPC community of practice
• Shodor
• Computational Science Education Reference Desk (CSERD) Pathway

• We welcome contributions and participation by all interested organizations
A Brief History

**HPC University RAT**

**Charter:** The HPC Training RAT will identify successful paths to creating qualified, effective HPC professionals, capable of exploiting current terascale and upcoming petascale technologies for the advancement of scientific research. Gaps in training materials and delivery methods will be identified, and recommendations for filling these gaps will be presented.

**Report Topics:**
- Catalog Map
- Topics
- Mastery Levels
- Quality Assurance
- Delivery Methodologies
- Target Populations
  - Demographics
  - Disciplines
- Scaling & Dissemination
- Getting to Petascale

**Key Issues Identified:**
- **Gaps** – Identifying and plugging gaps is a critical and ongoing requirement for viability
- **Persistence** - Are the materials available when users need them?
- **Quality assurance** – Do the materials provide a validated, verifiable experience for the users?
- **Delivery methods** – Are the materials available to the users independent of geography or temporality?
- **Scaling the training** – Are good trainers and training practices being identified and made available to new trainers?
- **Petascale Preparation** – Effective use of petascale resources require proactive efforts now and into the future

The RAT Report will be on hpcuniv.org shortly.
Report Recommendations

• Create persistent inventory of resources and events
• Ensure Quality Assurance via VV&A process
• Provide Multiple Delivery Methodologies
  – Support a variety of teaching methods and styles
  – Redesign materials for effective interactive online learning
• Broaden participation across under-served demographics and fields of science
  – Eliminate geographic and temporal boundaries
• Scale best practices of content and delivery
  – Provide education accounts and HPC University Portal
• Advance Petascale Learning
  – Engage HPC and petascale experts for guidance
  – Include petascale apps and scaling experiences via case studies
Build on strong CSERD Foundation

Expand the HPC University offerings for the community

Catalog

Infrastructure

Verification
Validation & Accreditation

Roadmap
Address Community Needs & Services

Catalog

Competencies
Computational Science

New & Improved Materials & Content

Evaluation & Assessment

Data

Needs
Community & Users

V V & A

Roadmap

NEW

Science Gateway

FORTIFIED

MORE

Increased participation

IMPROVED
Short-Term Activities

• Establish resource archive – persistent training materials
  – Train the trainer sessions to improve delivery
• Establish mentoring program
  – Identify mentors who can make the commitment
• Identify reviewers who can verify/validate existing offerings
  – Begin VVA process
• Formative evaluation and assessment
• Petascale survey and computational science survey
  – we would appreciate your help to disseminate these!
Long-Term Activities

- Iterative process to identify and fill gaps
- Use evaluation to study longitudinal impact
- Coordinate with professional societies on program accreditation
- Share meta-data with ACM and IEEE Digital Libraries
SC07-SC10 Education Program

- Multi-year, year-long, Education Programs to provide continuity and sustained impact
- Integrate computational science into high school and undergraduate STEM classrooms
- Use digital libraries for teaching and learning - CSERD/NSDL
- Sponsors: ACM, IEEE, TeraGrid, Blue Waters, NCSI, CSERD, Krell, and NSF
- Recruiting faculty to create innovative curriculum
Virtual School of Computational Science and Engineering

- A multi-state, multi-institutional organization
- Led by University of Michigan and Great Lakes Consortium for Petascale Computation
- Driven by faculty and graduate students needs
- Goal: prepare the current & next generation of scientists and engineers to utilize leading edge computer systems.
  - Initial focus: graduate & postgraduate education, & beyond
  - Develop core competencies for HPC and petascale
  - Create new curriculum
  - Offering Summer Schools and on-line learning materials
TeraGrid Campus Champions Program

• Training program for campus representatives
• Campus advocate for HPC and CI
• TeraGrid ombudsman for local users
• Quick start-up accounts for campus
• TeraGrid contacts for problem resolution
• We’re looking for interested campuses!
• Over thirty institutions are joining to date!
Science Gateways - Broadening Participation

- Special PRiority and Urgent Computing Environment (SPRUCE)
- National Virtual Observatory (NVO)
- Linked Environments for Atmospheric Discovery (LEAD)
- Computational Chemistry Grid (GridChem)
- Computational Science and Engineering Online (CSE-Online)
- GEON(GEOsciences Network)
- Network for Earthquake Engineering Simulation (NEES)
- SCEC Earthworks Project
- Network for Computational Nanotechnology and nanoHUB
- GI Science Gateway (GiSolve)
- Biology and Biomedicine Science Gateway
- Open Life Sciences Gateway
- The Telescience Project
- Grid Analysis Environment (GAE)
- Neutron Science Instrument Gateway
- TeraGrid Visualization Gateway, ANL
- BIRN
- Gridblast Bioinformatics Gateway
- Earth Systems Grid
- Astrophysical Data Repository (Cornell)
For Additional Information

http://www.hpcuniv.org (Beta Version)
http://www.teragrid.org
http://cserd.nsdl.org
http://www.nsdl.org

Scott Lathrop - lathrop@mcs.anl.gov
Laura F McGinnis - lfm@psc.edu
Brad Armosky - barmosky@tacc.utexas.edu
Shawn Brown - stbrown@psc.edu
Petascale Education Challenge

• Petascale computing is more complex than previous computing paradigms.
  – Wholly new approaches to computing are required.
  – No single university has the expertise and experience needed to fully exploit this extraordinary capability.

• Preparation for petascale computing requires solid grounding in CSE, especially HPC and HPC-related curricula, and these are still evolving.
  – CSE education not keeping up with Moore’s Law.
  – We can leverage expertise to establish best practices, fill gaps, and modernize the CSE & HPC curriculum.
Extending the value of the CSERD-NSDL Infrastructure to HPC Professionals

CSERD-NSDL (http://cserd.nsdl.org) has created a strong foundation for the teaching and learning of computational science resources, tools, and methods among K-12 teachers, undergraduate faculty, and their students.

We are extending this foundation to directly benefit graduate students, post-docs and scientists applying high performance computing (HPC) and the emerging national cyberinfrastructure including TeraGrid to advance scientific discovery.

The ambitious efforts of NSF’s Office of Cyberinfrastructure to significantly expand high-end Digital Services in support of computational science and engineering creates a critical need for education and training opportunities at all levels of learning (from K-12 to professionals) in all fields of science.

In the recent Computing in Science and Engineering (CiSE) journal special edition on High Performance Computing Education, nearly every article references CSERD-NSDL.
Sampling of Training

• **HPC Computing**
  – Introduction to Parallel Computing
  – Toward Multicore Petascale Applications
  – Scaling Workshop - Scaling to Petaflops
  – Effective Use of Multi-core Technology

• **Domain-specific Sessions**
  – Petascale Computing in the Biosciences
  – Workshop on Infectious Disease Informatics
  – Computational “X” to address multiple fields

• **Visualization**
  – Introduction to Scientific Visualization
  – Intermediate Visualization
  – Remote/Collaborative Visualization
HPC University: On-Line Access

• Goal is to reach significantly larger audiences
  - To reach people that can’t attend live in-person sessions

• Synchronous sessions
  - Training sessions
  - Seminars

• Over 30 on-line asynchronous tutorials
  - CI-Tutor launched to expand access
  - 4,570 accesses in 2007
Development

Programming Languages
• Serial Programming Languages
  • C/C++, FORTRAN, Scripting Languages
• Parallel Programming Languages

Development Tools
• Compilers
• Development environments
• Building Tools
• Debuggers
• Code Management

Software Engineering
Code optimization

Technology
- Processor
- Multi-core
- Network
- Routers
- Network Topologies
- Memory

Parallel Architectures
- Fault Tolerance
- Parallel I/O

Performance Analysis
- Serial Performance
- Parallel Performance
- Performance Modeling

Verification and Validation
- Mathematical Precision
- Numerical Stability of Algorithms

Canned (Third-Party) Codes
Workflow Management
• Data Management
• Grid Technology
• Queuing systems

Machine Interface
• SSH
• Unix
• Gateways

Visualization
Allocations process
• How-to’s
• Grant writing guidelines

Support
Backup Slides

1. What’s a Concept Map?
2. What’s an ontology?
What's a Concept Map?

Java Community Process™ begins with members joining the JCP by signing the Java logo, which is held by Sun. Sun supports the development of Java by providing documentation and making versions of Java available to the public. Expert groups consist of members who do the work of the Java Community Process (JCP) and are overseen by the Executive Committee. Companies become members of the JCP by signing within the context of the JCP. Developers support the development of Java and make SDKs. Program Management Office develops new Java specifications, which are instantiated in a public draft. A Java Specification Request (JSR) is a proposal to develop a new specification or significant revision to an existing spec. The JSR leads to the creation of a new Java specification. Development process may include tools such as development environments, which are used to write, test, and debug. Java developers often use specific development environments called integrated development environments (IDEs).

http://java.sun.com/new2java/javamap/intro.html
<table>
<thead>
<tr>
<th>Novice Undergrad</th>
<th>Apprentice Master</th>
<th>Journeyman Ph.D.</th>
<th>Master Post-doc/ Prof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/simulation</td>
<td>HPC Technology (hardware)</td>
<td>Domains</td>
<td>Software engineering</td>
</tr>
<tr>
<td>Operational issues</td>
<td>Architectures</td>
<td>Performance analysis</td>
<td>Scalable computing</td>
</tr>
<tr>
<td>Application packages</td>
<td>Programming/Algorithms</td>
<td>Code optimization</td>
<td></td>
</tr>
<tr>
<td>Science gateways</td>
<td>Development tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workflow management</td>
<td>Visualization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data analysis/Post-processing</td>
<td>Verification/validation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
<td>At least 10 offerings</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Less than 10 offerings</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No offerings</td>
<td></td>
</tr>
</tbody>
</table>
# Delivery Methodologies

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number of Entries</th>
<th>Delivery Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Green] Operational Issues</td>
<td>60</td>
<td>Asynchronous ✓</td>
</tr>
<tr>
<td>Programming &amp; Algorithms</td>
<td>40</td>
<td>Synchronous ✓</td>
</tr>
<tr>
<td>Development Tools</td>
<td>37</td>
<td>Face-to-Face ✓</td>
</tr>
<tr>
<td>Architectures (Parallel, Dist, Grid)</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Science gateways &amp; resources</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Performance Analysis</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Data Considerations</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Scalable Computing</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>HPC Application packages</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Modeling &amp; Simulation</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Code Optimization</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>HPC Technology (hardware)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Domains (physics, chem, etc.)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Workflow management</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Software Engineering</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Verification &amp; Validation</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Data analysis/post processing</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Total number of offerings: 239

This is a hole.