Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies

HPC User Forum

October 5-6, 2009

High Performance Computing Center (HLRS)

University of Stuttgart, Germany

Vijay K. Agarwala

Senior Director, Research Computing and CI

Information Technology Services

The Pennsylvania State University

University Park, PA 16802 USA

HPC USER FORUM

EDUCAUSE uncommon thinking for the common good

C A S C Coalition for Academic Scientific Computation

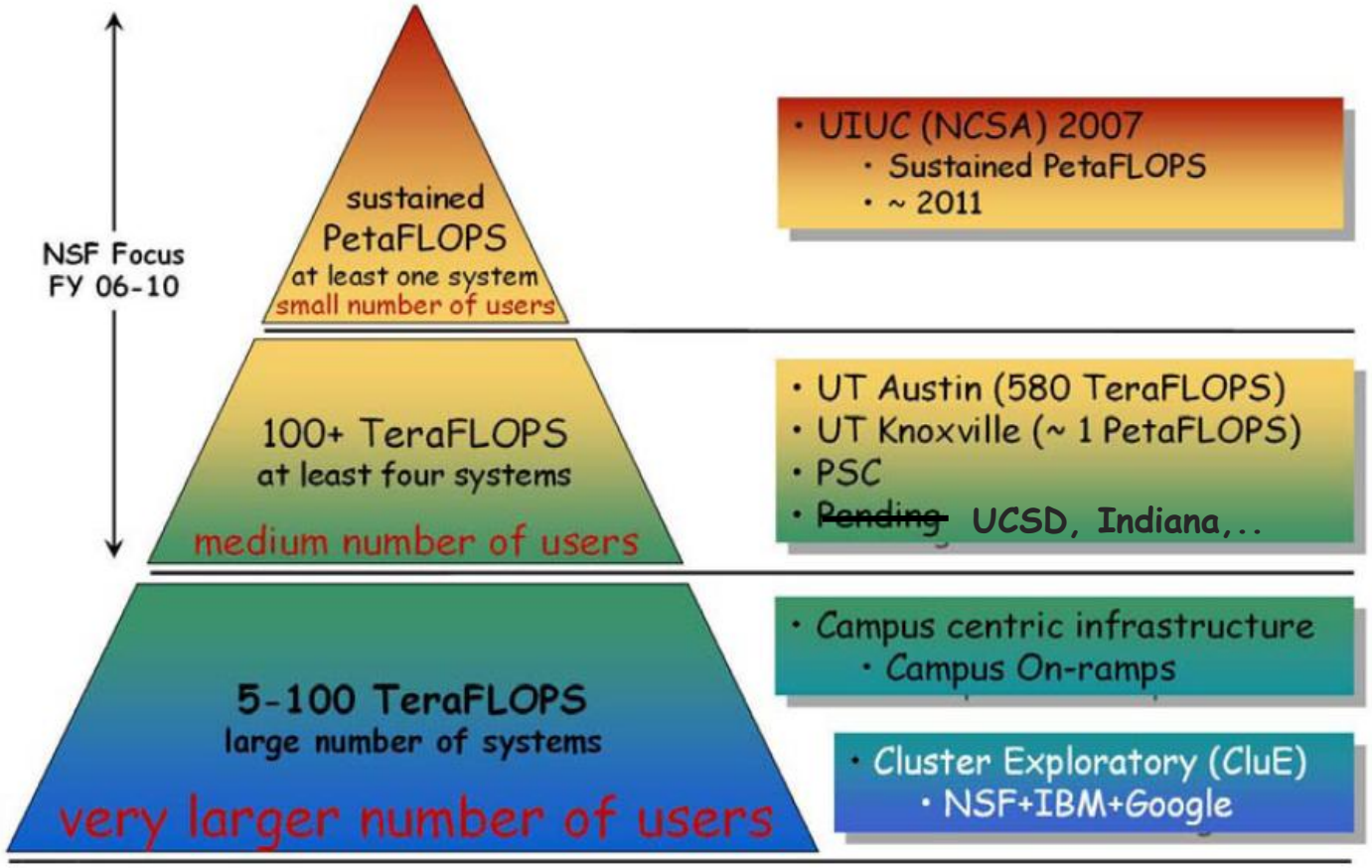
its@PennState information technology services

The full report is available at:

http://www.educause.edu/Resources/DevelopingaCoherentCyberinfras/169441

http://net.educause.edu/ir/library/pdf/EPO0906.pdf

Structure of the existing computing landscape in the context of the NSF-OCI Infrastructure investments



HPC USER FORUM
EDUCAUSE uncommon thinking for the common good
C A S C Coalition for Academic Scientific Computation
its@PennState information technology services

Towards a Coherent Cyberinfrastructure (CI)

- Goal: to remain competitive in the discovery of new knowledge and rapid pursuit of new research directions
- Current issues:
  - Limited ability of an individual researcher to effectively leverage campus and national resources in a seamless way
  - Nationally funded CI centers face challenge in actively engaging and enabling a wider spectrum of potential users
- Comprehensive CI requires a full spectrum of support and resources stretching from labs, through the campus layer, and up to the national centers

Key areas requiring focused effort

- Harnessing campus and national resources
- Information life cycle: accessibility, usability and sustainability
- Identity management, authentication, and authorization
- Human resources and broader impact

Harnessing campus and national resources

- Campuses in partnership with national resource providers and government agencies should support, promote, and develop a coherent, comprehensive set of computing and data facilities
  - Integrate national and campus layer resources
  - Funding models should enable integration of resources (data, computing and instrumentation) from lab to national center
  - Implement grant terms that encourage sharing and effective use of resources at all layers
  - Campus IT organizations should promote and develop new capabilities to access resources external to the campus

Harnessing campus and national resources

- Agencies, campuses, national or state network organizations must improve the aggregate national network infrastructure needed to address the data-transfer and remote resource access needs of a coherent CI
  - Campus network design for cybersecutiry, performance and robustness needed by CI
  - Compatible architectures and network interconnection patterns for coherent national CI
- Work together to create architectures which enable researchers and other CI users to make the most effective use of campus and national resources

Information Life Cycle

Accessibility, Usability, and Sustainability

- Funding agencies and institutions must fund:
  - operational implementations of data preservation to meet immediate and future needs
  - preservation and reuse to guide future activities
- Federal agencies, disciplinary communities, institutions, and data management experts should develop, publish, and use standards for provenance, metadata, discoverability, and openness.
  - Develop and adopt standards for data provenance, metadata, discoverability, reusability
  - Define standards for data life-cycle processes

Information Life Cycle

Accessibility, Usability, and Sustainability

- Funding agencies, research institutions, and communities must collaborate to develop a combination of policy and financial frameworks to ensure the maintenance of important data over time scales longer than the career of any individual investigator.
  - Policy development by funding agencies for distribution of data must expand to explicitly address maintenance of data over periods of time to make the data a societal asset
  - In addition to developing policy frameworks, all the agencies must develop financial and management strategies that assure availability of funds for maintenance of data

Identity Management, Authentication, and Authorization

- Agencies, campuses, national and state organizations should adopt a single, open, standards-based system for
  - Identity management
  - Authentication
  - Authorization
- Improve the usability and interoperability of CI resources throughout the nation
  - Global federated system supported by InCommon Federation

Human resources and broader impact

- Agencies and campuses should support a strategic investment in human capital and curricula in order to build a pipeline of qualified experts who can develop the full capacity of Cyberinfrastructure
  - Open-source mechanisms should be used to share curricular materials
- Develop technologies and tools to use the emerging CI for education and scholarship
  - Investigate whether commoditized on-demand computational and storage systems offer practical and economical solution

Human resources and broader impact

- Agencies and campuses should invest in partnerships between industry and academia
  - Work with businesses to adopt the use of CI and assist in transforming sponsored research projects to small and medium-sized businesses
  - Identify the industry needs for new modeling software, adapt software to run effectively on several HPC platforms
  - Academia and industry should adopt a sensible model for sharing intellectual property
  - Develop effective strategies to encourage students from traditionally underrepresented groups to pursue academic careers in computational science and to address workforce needs in industry

Building blocks of CI

Issues and Opportunities

- Relevant technology trends
  - Storage trends : Capacity per dollar is doubling every 12 months
  - Network trends : Increase in the use of owned fiber-optic cables, more slowly falling cost of 10-Gbps opto-electronics in switches and routers
  - Ability to move large scientific data sets is growing at a slower pace than our ability to generate and store data
  - Data-set size is doubling every 12 months, processing capacity every 18 months, but the university-to-university network performance is doubling every 48 months

Building blocks of CI

Issues and Opportunities

- Improvement in performance will require effective use of multi-core processors, hence sustained investment in software and algorithms needed
- Medium-sized campus layer clusters are most efficient “capacity” machines for running high-throughput computations. In this role, they are strong complement to national “capability” resources

Cloud Computing – will it disrupt the existing CI landscape?

Cloud computing : Virtualized/remote computing resource from which users can purchase what they need, when they need it

amazon web services™

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Simple Storage Service (Amazon S3)

Nimbis Services®

A clearinghouse for buyers and sellers of Digital Analysis Computing (DAC) services

r systems
rapid response computing resources

Computing resources for commercial and academic community

CRL
Computational Research Laboratories

Affiliated with TATA CONSULTANCY SERVICES

Offers HPC services to industry, government ,and academia

HPC USER FORUM

EDUCAUSE
uncommon thinking for the common good

C A S C
Coalition for Academic Scientific Computation

its@PennState
information technology services

Cloud Computing – will it disrupt the existing CI landscape?

Google IBM National Science Foundation WHERE DISCOVERIES BEGIN

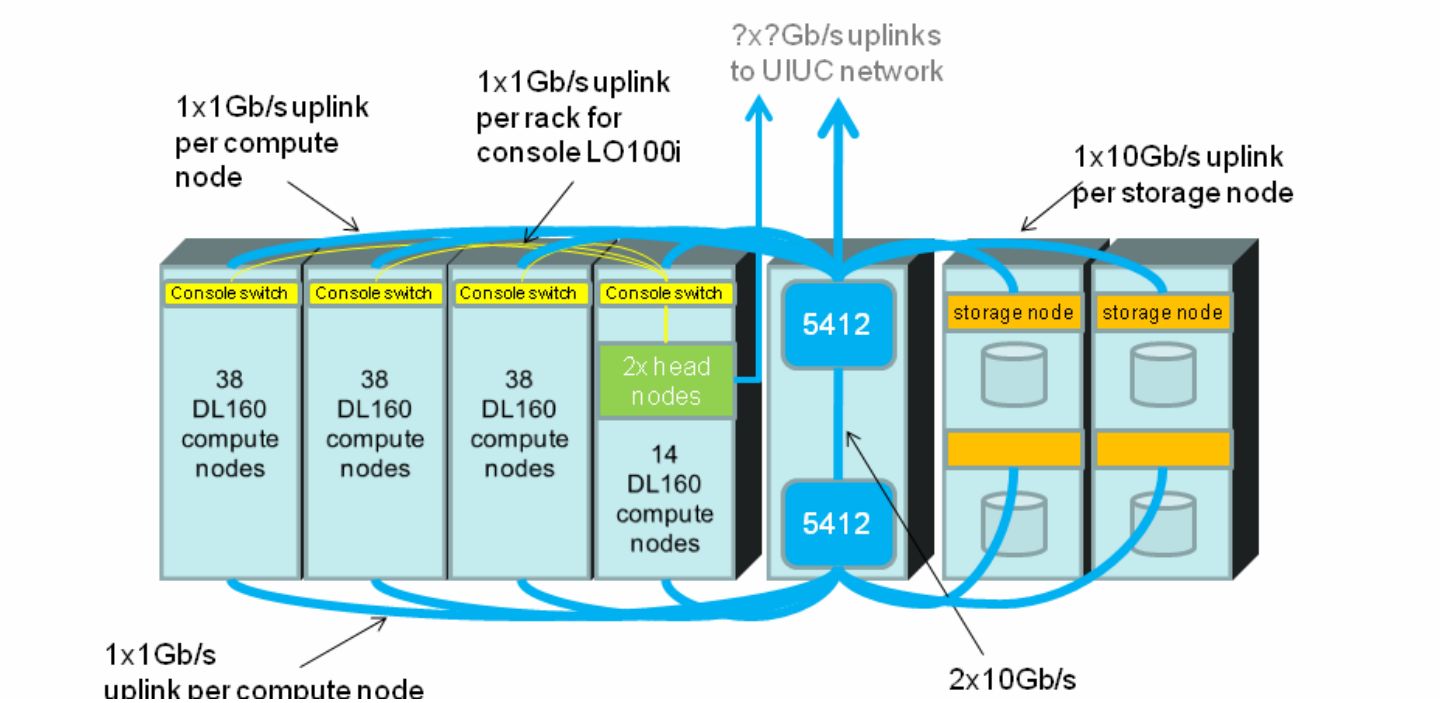
- Google+IBM cloud to provide a platform to computer science researchers for building cloud applications
- Combination of Google machines, IBM blade center and system x servers
- NSF recently awarded $5.5 million to fourteen universities through its Cluster Exploratory (CluE) program

The universities receiving funding include: Carnegie-Mellon, Florida International, MIT, University of Wisconsin, Yale, Purdue, UC-Irvine, UC-San Diego, UC-Santa Barbara, University of Maryland, University of Massachusetts-Amherst, University of Virginia, University of Washington, University of Utah. This is part of “Cloud Computing University Initiative” which serves as a spear-head group and focuses results on industry needs.

Cloud Computing – will it disrupt the existing CI landscape?

→ A global open source test bed, Open Cirrus for the advancement of cloud computing research

→ University of Illinois, KIT (Germany), iDA (Singapore), ETRI (Korea), MIMOS (Malaysia), Russian Academy of Sciences



HPC USER FORUM
EDUCAUSE uncommon thinking for the common good
C A S C Coalition for Academic Scientific Computation
its@PennState information technology services

Cloud Computing – will it disrupt the existing CI landscape?

- What impact buying of commercial cloud services will have on future development of campus, regional and national CI?
- How tightly should the campus CI be integrated with that of private companies in a geographic region? Would it promote stronger partnership between academia and industry?
- Would new policies and guidelines emerge from federal funding agencies on incorporating cloud services in the grants? Would cloud service providers have to be US-based for data and security reasons?

Demand for cloud computing services by campus based researchers will depend on how research computation related CI is funded and deployed in the future.

Expanded Role of University based Centers

- *help make industry and business globally competitive*

- Leveraging CI assets to support business use of advanced modeling technologies and educating future workers in computational science
- Focused initiatives by universities on industrial applications of supercomputing to overcome barriers to widespread adoption of advanced CI
- Implementing and demonstrating virtualization and VPN technologies to ensure confidentiality of customer's data. This will increase the use of shared resources by industry.
- Demonstrating ROI of adopting supercomputing – this is essential to overcome the barrier of 'high perceived cost' among medium-sized businesses.