

UC Information Technology Guidance Committee
Research Computing Work Group Report
Prepared for the ITGC meeting on September 19, 2006

Draft Position Statement

There have been opportunities to hear and assimilate opinions and comments from a wide spectrum of UC academic researchers, VCRs and CIOs, and industrial practitioners and partners, to track national trends with cyber infrastructure initiatives and to receive feedback that UC is falling behind with respect to competitiveness in cyber infrastructure-supported and IT-enabled research. The ITGC Research oversight group will begin tackling broad-based considerations for UC.

From these discussions, we can begin to pose at least four primary premises that describe areas of potential for UC leadership and national competitiveness:

Premise 1. There is a set of state/national/global problems challenging society that UC should be playing a leadership role in helping to solve. The challenge problems include energy development, use and management, homeland security, environmental impact and management, infectious disease management, food and water supply, health management and earthquake preparedness.

Premise 2. The State/US value share in a global market is expected to dramatically turn toward vision and application of “knowledge” and “information” that can be applied in highly targeted, individual use and use-specific ways. Targeted use encompasses ‘kinds’ of applications like use-specific product designs, just-in-time product production, assembly and/or delivery, highly refined social and economic analyses, just-in-time community analysis, substantially shortened product life cycles and highly refined and tuned technologies and processes, and data and information integration on macro scales.

Premise 3. The State/US mind share and IP impact of knowledge and information for use at macro system levels will stem from understandings and technologies at micro levels i.e. the cell, the molecule or the individual. There will be greatly escalated amounts of data and information used to conduct the research and development.

Premise 4. A transition to global management, e.g. global supply chains and smart manufacturing sites that operate to tighter specifications, more tightly tuned socio-economic decision making, environment impact management, will involve much greater understanding of the processes, greater automation and decision support, more expansive use of automation, data and data interpretation, and a new generation workforce that is trained and oriented toward a knowledge and information mindshare.

Across these, there are significant roles for simulation, modeling, optimization, design, ubiquitous sensing, spatially wide area data aggregation and interpretation and large-scale/wide area data management, visualization and analysis but in significantly evolving

ways that are supported by cyber infrastructure. There are four areas that stand out for cyber infrastructure:

- Data need to be collected, mined and interpreted in increasingly sophisticated ways that involve increasing numbers and/or larger coverage of sensors to achieved more refined understanding, models and information. Massive increases in the need for data storage, management, shared access, standards and security loom large for the near future.
- Modeling, simulation, parametric analysis, concurrent evaluations, data analysis etc. are needed to shorten product life-cycles, achieve next generation products and develop processes for just-in-time, targeted use products and goods.
- Operational balances among automation, decision support and human involvement will need to shift but in a coherent manner that depends concurrently on technology development, operational and regulatory trust of automation and education, human factors development and training.
- Sophisticated understanding and practical development at the level of the molecule, the cell or the individual but used at the macro/global level of a company, a region, a state, or a country require increasingly sophisticated computation resources that need to aim toward petascale capabilities and the mathematical and computer science expertise to move to new plateaus of application code.
- Addressing macro/global problems requires sharing of data, information, data and IP.

A UC-grid is a key enabling capability that takes UC toward the virtualization of resources across geographical locations. It then becomes possible to invest in, aggregate distributed resources and optimize a given geographical location for a particular type or class of service that can then be made available to a wide range researchers regardless of location. Virtualization should not be confused with centralization. Virtualized services become components of an adaptive infrastructure that is responsive to the management and needs of the individual researcher or local research group as well as large interdisciplinary teams. High performance computational, visualization and data services can be made available without any one group or facility being over burdened with providing all services with their own resources. The grid and the exposed resources become a sum-greater-than-the-whole enabling infrastructure for the conduct of the research and grant competitiveness.

A UC-grid has the potential to significantly alter the playing field for UC's research investment:

1. What is the investment and incentive strategy for computational and storage resources that can reside at one campus and be used by another.
2. What education and behavioral incentives are needed to engender trust and demonstrate capability such that faculty notions of facility ownership can change.
3. How can infrastructure investments be positioned for grant competitiveness

4. How can UC leverage its collective stake in IP associated with application software of IT enabled research and how can it efficiently facilitate the strategic development of software to solve bigger problems.
5. How can UC foster collaboration and facility development and invest to solve large problems of state and national interest.
6. How can UC foster the education and training needed to draw benefits from an investment in cyber infrastructure resources.

UC Grid Project Overview

The demands on the UC Research Cyberinfrastructure are growing at a phenomenal rate as a result of the increase in computational power available, data being collected and generated, and the requirement for more inter-disciplinary collaboration. This growth is taxing the capability of existing networks, support staffs, and storage and backup services along with creating an ever-increasing complexity for researchers needing access to these resources.

To deal with this demand, a new set of tools and - more importantly - a new and adaptive infrastructure must be developed. To this end, the design goal of the UC Grid Project is to give researchers a unified view of, and access to, their computational resources, regardless of their physical location. From any desktop equipped with a web browser and a network connection (PDAs can also be used), a researcher can perform all the tasks associated with the submittal of serial, parallel and pre-defined commercial application computing jobs to any computational resources they have access to. The UC Grid will give a researcher the ability to view the overall status of a cluster, number and types of jobs in various queues, busy and free nodes, and the ability to move data and program files within and between clusters utilizing a secure, certificate-based access model. It will also provide access to a rich set of visualization tools that work in conjunction with UC Grid resources. The UC Grid will be built on the Open Source Globus Toolkit. The UCLA-built Grid Portal and Appliance overlay software, on which the UC Grid will be built, has already been released to the Open Source community.

The ultimate vision of UC Grid Project is the complete virtualization of resources across geographical locations, where any given location could be optimized for a particular type or class of service and then be made available to a wide range of researchers through a unified, web-based Grid interface. In this way extremely high performance computational, visualization and data services can be made available without any one group or facility over-burdened by attempting to provide all services with their own resources.

The first task is to provide a Grid infrastructure in order to expose existing computing resources to the UC research community and to facilitate the use of those resources as appropriate to existing research needs and funding. The UCRCG proposes to provide this infrastructure by:

1. Creating a Campus Grid at each University of California campus

2. Creating a UC-wide Grid called the UC Grid.
3. Deploying the Grid Certificate Authority (CA) for all the Campus Grids and the UC Grid at the UC Grid level.
4. Using resource pools to provide researchers with the most appropriate compute resources and software anywhere in the UC system according to their compute requirements

This initial deployment will fulfill the following mandates:

- To develop a secure Grid of computing and storage resources in support of research.
- To augment UC's research technology infrastructure in an interoperable fashion that will facilitate the sharing of resources.
- To optimize performance and utilization.
- To deliver services that address researchers' needs while encouraging behavior that benefits the common good.

It will also provide easy access to a very large number of users without having to create individual user login ids for them on all of the clusters.

High Level Project Next Steps

- Establish UC Grid test bed which includes:
 - UCG, MyProxy and Simple CA servers
 - Three test campuses with local MyProxy servers
 - Writing the Grid Registration Service (which is already in progress)
- Install and test replication and fail-over configurations on the test bed
- Perform beta test of UC Grid with three test campuses
- Test with the production UCLA Grid Portal
- Add three additional production campus Grids
- Continue until fully deployed

After deployment, parameters and codes can be optimized and tuned to maximize performance, stability, security, and resource utilization while at the same time ensuring the fastest turnaround for users.

Future Phases

Currently the emphasis in developing the UCLA Grid Architecture and the UGP software has been to join computational clusters together into a Grid. To extend the UC Grid concept and enable the creation of a California Grid for use within K-12 education, will require that Grid services be expanded to provide for other services in addition to batch computing. This will require an assessment of needs, as well as the development necessary to a) connect the kinds of compute resources that meet those needs into the Grid and b) add the user interfaces for those kinds of resources into the UGP software. This will be addressed in a later phase of Grid development.

Funding Models

Funding models and infrastructure at the campus level will have to be addressed in order to create the UC and Campus Grids. The proposed UC Grid architecture is such that discrete identification of resources related to a given Grid component can be easily identified and funded. Additionally an individual campus Grid capability can be grown in an evolutionary fashion based on utilization. An estimate of cost, suggested architecture, and funding sources associated with deployment of the UC Grid and Campus Grids along with the recommended funding are:

1. UC Grid - The UC Grid architecture requires four systems as a minimum to ensure uptime and performance. The equipment consists of an Apache web system, two Tomcat/MySQL/UGP systems and one MyProxy system. Approximate cost is \$22,000-\$25,000 depending on vendor and configuration. *UC or pooled campus funding.*
2. Campus Grid – A campus Grid deployment consists of a minimum of three systems. Note that there is no redundancy in this configuration. The equipment consists of an Apache web/Tomcat/MySQL/UGP system, a MyProxy system and a Storage Server. Approximate cost is \$12,000-\$25,000 depending on vendor and configuration (particularly the storage server). *Campus level funding.*
3. Appliance Node – Each cluster connected to a Campus Grid (and the UC Grid) requires an appliance node. Approximate cost is \$900-\$1,100 depending on vendor and configuration. *Campus or grant funding.*

In addition to the system recommendations above, a small percentage of FTE is required to handle UC and Campus Grid administrative tasks that would be funded by the respective organization. It is expected that the initial Grid start-up will require a larger time commitment than the sustaining effort necessary to maintain the Grid once it is in operation. The FTE time required varies according to the number of users, usage of the Grid, robustness of the Grid systems and other factors and can vary from .1 to .25 FTE. Start-up time would be on the higher end of this range but should only last for two - four weeks. Additionally there is a requirement for individual cluster administrators to enable Grid access for their cluster users. The time requirement to perform this task is on the order of minutes per user. The assumption is that a Programmer Analyst level III or IV will perform these tasks depending on the complexity and usage of the Grid installation.

After startup funding is secured, a long-term funding model will have to be developed that can sustain the Grid services and the technologies behind them.

The full details of the UC Grid Project can be found in the UCRCG Grid Project white paper.

Post-meeting Supplement

A. Key questions presented to ITGC for discussion

- How can UC best facilitate collaboration among researchers to address the "grand challenges" by transforming cyberinfrastructure from hardware and software owned by individuals to services offered by the University?
- How can UC facilitate/support the build-up of "cyberinfrastructure" expertise for the conduct of cyberinfrastructure dependent research.
- What funding and allocation strategies should UC pursue for cyberinfrastructure?

B. Outcomes from ITGC discussion

Scope of Support

- We need facilities (*e.g.*, computers, space, power) *and* support services.
- We need to support all disciplines, not just science and engineering. UC may, however, want to identify certain application areas that are emphasized
- We need the ability for students (graduate and undergraduate) to access the research IT cyberinfrastructure.
-
- Cyberinfrastructure Support Services
 - We need to leverage centers of excellence among the campuses within UC.
 - Expertise in academic disciplines, as well as cyberinfrastructure tools, will be required.
- Changing the Culture
 - Our challenge is how to change the culture to be more favorable toward sharing of IT resources.
 - Researchers are motivated to share resources when there is added value, such as sharing in a larger whole or receiving support services.
 - We should identify those "value adds."
 - Is there a base level of resource that should be provided to all faculty?
- Resource Acquisition
 - We should consider outsourcing components of UC's high-performance computing infrastructure, being careful to consider all costs (*e.g.*, space and power).
- Alignment with National Strategies
 - We need to meet with representatives from NSF, DOE, *etc.* to ensure alignment with national cyberinfrastructure strategies and to position and make UC's efforts known by key national players, i.e. market the substance of our efforts.
 - We need to visibly position UC cyberinfrastructure initiatives nationally.
- UC Grid
 - The ITGC supported advancement of the proposed UC Grid pilot.

○ **C. *Next Steps***

- The High Performance Research Computing Steering Group will address the issues raised in this discussion.
- The Research Computing Group will continue with its efforts to create a pilot UC Grid.