ECAR Research Bulletin

Developing and Extending a Cyberinfrastructure Model

Overview

Increasingly, research and education institutions are realizing the strategic value and challenge of deploying and supporting institutional cyberinfrastructure (CI). Cyberinfrastructure is composed of high performance computing systems, massive storage systems, visualization systems, and advanced networks to interconnect the components within and across institutions and research communities. CI also includes the professionals with expertise in scientific application and algorithm development and parallel systems operation. Unlike “regular” IT infrastructure, the manner in which the components are configured and skills to do so are highly specific and specialized. Planning and coordinating these assets is a fundamental step toward enhancing an institution’s research competitiveness and return on personnel, technology, and facilities investments.

Coordinated deployment of CI assets has implications across the institution. Consider the VC for Research whose new faculty in the Life Sciences are now asking for simulation systems rather than wet labs, or the Provost who lost another faculty candidate to a peer institution that offered computational support for research, or the VC for Administration who has seen a spike in power and cooling demands from many of the labs and office spaces being converted to house systems. These are just some of the issues that research institutions are wrestling with as research becomes increasingly computational, data-intensive and interdisciplinary. This bulletin will discuss these issues and will present an approach for developing a cyberinfrastructure model that was successfully developed at one institution and then deployed across institutions.

Highlights of Cyberinfrastructure

Data-intensive research, interdisciplinary research and inter-institutional research is quickly becoming the standard in most scientific disciplines. Central, if not fundamental, to these types of research is computation. As some have suggested, computation has recently become the third pillar of science, joining both theory and experiment. If we think about how discovery happened in the past, discoveries accrued to those who had access to unique instruments and their data. But the growing costs and complexity of tools prohibit the proliferation of instruments. Instead, we have very few unique instruments producing data, say for example the Large Hadron Collider in Geneva, and that data housed in a few locations across the globe. The result is that discovery is now based on the right questions rather than access to unique tools. We go from hypothesis driven science “I have a question, I will collect or find data” to exploratory driven “I have lots of data, what can I glean from it”. Research in many disciplines has become much more about access, analysis, movement and management of very large amounts of data. Indeed, the availability of data and computational resources has created the rise of new areas of study, such as synthetic biology, genomics and bioinformatics, to name a few. It has prompted the National Science Foundation to issue a report stating that computation is being used “to replace and extend traditional efforts in scientific and engineering research, indeed to create new disciplines.” But computation has also become important in fields not traditionally considered “hard” science such as those in the social sciences as well as the arts and humanities.

Researchers have responded to the need for computation, much like they did in the late 70’s and early 80’s when there was a need for network access but the wiring was slow in coming to buildings and research labs. The common solution then involved science departments running cables, setting up electronics and providing networking services to inhabitants of buildings for whom network access was a fundamental requirement for doing research. Similarly, some twenty-five years later, many researchers have deployed their own clusters of high performance computers. A cluster is a single system comprised of interconnected computers that communicate with one another. There is usually a master node and many (even hundreds) of compute nodes. And while CI encompasses much more than clusters, they have fundamentally transformed research-based computing in the last few years for a few compelling reasons.

Clusters have very important socio-economic effects. They are built from relatively inexpensive commodity PCs, and use Linux and tools available in the public domain. In effect, even high schools can buy and build clusters. Now individuals and laboratories at universities believe (and have shown) that they can assemble and incrementally grow a small to midrange supercomputer. In many instances clusters have provided the computational "staging ground" needed by researchers who have outgrown the power of their desktops but are not quite ready to compete nationally for allocations on larger supercomputers such as those available at the National Center for Supercomputing Applications, the San Diego Supercomputer Center or the National Energy Research Scientific Computing Center. In the mid-range space, researchers are able to use clusters for staging productions runs, parallelizing and optimizing code, and other work that prepares them for competing nationally for supercomputing allocations.

The economics of clusters is the key advantage. Not only is the hardware and software reasonably inexpensive, but these distributed clusters ride "free" on their lab, department or research group's organizational overhead that includes space, networking and personnel. It is no surprise then, that research and education organizations have pockets of computing strewn throughout their research areas. Much like the scenario of a couple of decades when IT divisions began piecing together the hodgepodge of do-it-yourself networks around campus, the landscape of high performance computing presents a similar challenge to research organizations. As perhaps the only comprehensive study on research computing shows, many campuses find that most research computing is highly decentralized (ref from ECAR study). While these distributed labs are able to tailor the configuration of systems to the needs of the specific science discipline and their discovery methods, there is much effort expended that is hidden or implicit for these researchers. Further, these resources are usually confined to a specific department and are therefore unavailable to support other areas of science within the institution.

There are hidden costs associated with assigning researchers or graduate students to spend considerable time on system administration rather than on scientific inquiry. The cost of lost research productivity of these individuals is not factored into the total costs of owning a cluster system. Not to mention that researchers are not system administrators, so the time expended doing this activity is considerably higher than would be by a professional. There are also hidden costs associated with losing office or teaching space to computers that are placed in areas not designed to for them and therefore do not maximize the use of space. But the exponential increase in power and cooling for these cluster systems is perhaps one factor that while previously hidden to some degree, is making now becoming much more explicit.

For instance, Hacker and Wheeler (2007) provide a cogent example of the hidden electrical and cooling costs incurred by distributed cluster systems. They suggest that a 1 teraflop (TF) system can consume about $52,416 of electrical energy in a year. In the short timeframe since the publication of their article, computing power per compute node of each cluster increased substantially. Evidence from clusters going into production now shows that a 1TF system consists of fewer nodes with more processing cores and will consume approximately one tenth of that amount. However, the demand for computation power for the most part, is constrained by the amount dollars available for purchasing systems. Therefore, 2 and 3 TF systems are not at all uncommon. One can easily find 50 and even perhaps as high as 100 of these in various research labs at a mid-sized research university. Based on this, those mid-sized universities can project that these clusters add anywhere from $750,000 to $1,500,000 to a university's utility bill. If you add to that the cost incurred of inefficiently cooling these systems due to aging or poorly designed computer room facilities(often times with fans or window air conditioners purchased at local retail stores), one can easily double the cost. In the absence of systematic data on the proliferation of clusters, these are rough estimates, but the numbers are large enough to no longer be ignored.

The projections are embellished by the anecdotes of computer disasters. The stories of overheated computers smoking and burning abound. At the Lawrence Berkeley Laboratory, a cluster was housed in a substandard space that had marginal cooling and insufficient monitoring of the environment. Without proper notification systems in place, during a power outage the computers sat in 114 degree temperatures overnight. When discovered, one storage unit was permanently damaged and had to be sent to a firm specializing in last resort data recovery. The cost of research or experimental data loss can be invaluable, not to mention the time of the IT and facilities staff spent on this emergency. And for the most part, ad hoc remedies are patched together in the emergency scenarios rather a more coherent cost and energy efficient plan.

The implicit and explicit costs associated with the deployment of distributed high performance clusters is perhaps one of the more compelling arguments to be made in support of a more planned and rational model. The following section describes how the Lawrence Berkeley National Laboratory developed such a model and then extended in support of scientific research at the University of California Berkeley.

Developing a CI Model

Scientific Cluster Support – Phase I

Computing has been part of scientific research for the last fifty years. At the Lawrence Berkeley National Laboratory scientific research computing had evolved from centralized supercomputers and timesharing system to powerful desktop computing in the mid 1990. However, many scientists' computational needs exceeded the power offered by desktop devices and were finding themselves at a competitive disadvantage as compared to their peers at other institutions. Moreover, many were interested in allocations at the national supercomputing user facility housed at LBNL, but only a select few were chosen, leaving a gap between the high end supercomputing and the lower end desktop computing. This "mid range" computing gap was identified by LBNL and as a consequence a mid range working group, composed of scientists from a variety of fields, was set in motion in early 2000 to identify how to address this gap. After several months of work, the working group put forward a proposal that opposed the idea of purchasing an institutional computing resource and instead supported the idea of a central cluster support program. That is, scientists would obtain funding for their cluster but the services to support the cluster would be made available by the IT Division which had developed a high degree of expertise in supporting researchers over the years. The relationships that the IT Division staff had developed with the scientists, the high degree of interest, need, and support within the research community, and the commitment of key management were factors in getting a centrally-managed cluster support program started.

Specifically, the cluster support program staff would provide the following:

1. Pre-purchase consulting – based on scientific problems to be solved, determine the hardware, interconnect, operating system, compilers and application software.
2. Procurement assistance – develop a budget, specification for RFP and acceptance criteria and evaluate bids.
3. Cluster integration – install and configure hardware, operating system, cluster software, applications and computer security.
4. Ongoing systems administration and cyber security – system maintenance and upgrades, cluster monitoring, hardware troubleshoot, maintenance of cluster software stack, resource management and scheduler support, and user account maintenance.
5. Data center space, networking, power and cooling – host clusters in data center to ensure adequate cooling, power and networking.

The Scientific Cluster Support (SCS) program, as it was called, was allocated $1.3M from central funds for the first four years, this would fund the positions of two FTEs. At the end of the four year period SCS was to have developed a full recharge business model. Ten research groups with funds to purchase a cluster were selected for support in the first four years.

The SCS program developed a cost-effective methodology with hardware and software standards that facilitated the scaling of systems administration support. They also utilized open source software, such as Linux. Because the ten pilot projects came from a variety of scientific fields, the computational needs of some varied from others so the requests for exceptions to the standards were not uncommon. In these instances, a small steering committee composed of stakeholders and outside technical expertise proved invaluable. They helped enforce the standards with their peers, by vetoing requests for these exceptions, understanding that if too many were allowed the costs of the program would increase for the scientists that would pay later after the pilot period.

For the key area of cluster management, the IT Division was not able, at that time, to find a toolkit that would allow for a scalable method of supporting the clusters. Therefore, the IT Division developed a cluster management toolkit called Warewulf that greatly simplified installation and management of clusters. Simply put, Warewulf allows compute nodes to boot from a shared image on the master node of each cluster so that a system administrator needs to support only a master node. In effect, the compute nodes are “stateless”. In cases where some clusters can have a hundred or more compute nodes but only one master, this was immensely valuable in labor savings.

Towards the end of the four year pilot, several business models were developed and vetted with various advisory boards and members of the scientific community at LBNL. The resulting model that was found to be acceptable to both the administration and researchers was a partially funded SCS program. An allocation of $350,000 was given to the program (with promised cost of living increases) and all power and cooling costs would also be subsidized centrally. This dramatically dropped the cost for researchers. The program continued to grow rapidly. In an 18 month period, the number of clusters supported by the increased by 58%. By Fall 2007 the SCS program was supporting approximately 30 clusters representing approximately 3000 processors.

Scientific Cluster Support – Phase II

As successful as the SCS program was, it had its challenges. Intra-cluster management was efficient, but inter-cluster management was not. That is, each cluster was a system onto itself which was based on similar standards but even the slightest variation required the unique configuration of each master node. Also, many researchers who outgrew an initial cluster would purchase another but inter-cluster sharing of computation or other resources was not possible for them. As a consequence, some of the clusters were running at 50-60% utilization. Consequently, a member of the IT Division developed a new cluster management toolkit called Perceus that enabled much larger scalability by creating a “meta-cluster” and in a sense, flattening the cluster management.

The new toolkit builds on the existing management systems’ approach, but goes further in that master nodes are also stateless. Most of the management is moved to a central Perceus appliance that contains standard software images from which all the clusters boot. In effect, the clusters in this model were aggregated in such a way so that they appear as one meta-cluster requiring dramatically less time to manage. Rather than managing some 30 plus unique systems, the system administrators now manage one meta-system consisting of groups of nodes. Because the clusters were integrated in this way, it also allows researchers with two or more clusters to now move seamlessly across them when executing code or moving data. This could result in higher utilization of cluster resources overall. Additionally, spare cycles on any cluster can be harvested and allocated to other researchers. In a sense, this new toolkit functioned much like a local grid system that allows researchers to utilize computational resources across the grid.

An additional component of the second phase of this CI model includes the implementation of a shared institutional cluster - even though four years earlier it had been considered too ambitious and voted down by the working group. The IT Division surveyed scientists and discovered that 38% of them depend on clusters for their research and of this group, 70% said that they would be interested in purchasing cycles from an institutional cluster. A market analysis of commercial offerings, when normalized to a standard level of performance, showed that the cost to run an average 400 cpu-hr job ranged from $148 to $400.. The IT Division once again assembled a science team to help them assess the viability of various business models for an institutional cluster. They analyzed the break point of buying cycles versus buying a cluster. After several financial iterations, they settled on $.10 a cpu-hour. At this price, the comparable 400cpu-job would run $40. This prices was a compromise between the more expensive commercial offerings and the free if-you-can-get it allocations from the National supercomputing facilities. At this price point, it was shown to be more cost-effective to buy time on the institutional cluster unless one was prepared to purchase a 50-node cluster or larger. So for researchers with needs under 50-nodes (the average size of a cluster at that time was about 40) it was better to buy time. After putting forth a strong business case, the IT Division received $1M to purchase the institutional cluster. LBNL also provides competitive research funding which plans to purchase time on the cluster for researchers who are awarded funds and require computational support. There is also discussion of purchasing time on the cluster to award to newly recruited scientists who have needs for computation.

Not unlike many other major research universities, UC Berkeley was facing an increase in the proliferation of clusters and demand for cluster support services. With a long tradition of collaboration (indeed LBNL was born at UCB as an organized research unit before becoming a Department of Energy funded laboratory), the two institutions partnered to provide cluster services to UCB faculty. The partnership was supported both in spirit and financially by the Vice Chancellor for Research and the Provost at UCB. They had witnessed the proliferation of requests to NSF and other funding agencies with budgets for clusters that later required infrastructure support that was a drain on their budgets. Drawing on the strengths of each institution, UCB's IT Division will house the clusters in their newly built data center and LBNL IT Division will provide support services similar to those offered at the Lab. The two institutions struggled with developing a financial model that would adhere to UC financial standards as well as DOE standards and oversight. As the two institutions wrestled with their disparate financial systems and overhead rates, researchers at UCB lined up for cluster services. When high performance cluster services at UCB was kicked off in the fall of 2007, three research projects were already in the queue with another awaiting notice from a grant proposal.

What It Means to Higher Education

There are many aspects of CI, as mentioned earlier, but certainly high performance computational clusters are one key component that is transforming research and impinging upon the physical infrastructure of universities in doing so. To some degree clusters are the "last mile" in CI. That is, much like broadband technology (cable or DSL modems) have been the last mile into the home, cluster technology is the last mile into the research lab. As research and education institutions become increasingly aware of this, there are some important trends that should be of concern to Higher Education.

There is a dearth of comprehensive aggregated data on CI in Higher Education that can adequately paint a picture of the state of CI. The exception is ECARs seminal study "IT Engagement in Research: A Baseline Study" (Vol. 5, 2006). Some interesting findings from this study show us that for the institutions surveyed, roughly 2 out of 3 project an increase in high performance computing and high performance networking whereas only 1 out of 2 said that they saw an increase over the past three years. Clearly, CIOs and the like across campuses are expecting growth like they have not had in the past and should be preparing their organizations to meet this projected growth. Yet, approximately half of those surveyed currently have less than 1

FTE assigned to provide research computing support. And 60% project that staffing will remain the same in that area.

Another area of concern to Higher Education is funding of CI. Slightly more than half of the institutions surveyed spend less than $100,000/year on support for research computing. Not surprisingly, only 35% believe that they have a sustainable budget to support research IT and 57% believe that the biggest barrier to funding is lack of institutional commitment. Other institutional priorities take precedence over CI in the eyes of university budget decision makers. Yet, underfunding CI has very costly consequences, as mentioned above, in terms of utilities and loss research productivity for both existing faculty and the lost recruits that may have gone to competing university or research organizations where CI support is an institutional priority.

Key Questions to Ask

There is no one-size-fits-all model for supporting CI on any campus. A strategy and program must be contextual and participatory with the goal of making a direct impact on research. Below are some questions that can help in beginning to assess CI interest and developing a strategy to address those interests:

- What is the magnitude of the current local situation-- how many clusters are run in local centers and what other type of CI services are offered?
- What is the potential immediate growth by reviewing how many research grants are going out with requests for computational funds or computing allocations?
- How much and what type of CI support does central IT offer? What are the potential "gaps"?
- Does IT have in/formal relationships with researchers that can be leveraged to develop and implement a CI plan?
- What is the commitment of senior management and key decision makers to support CI?

With this type of information in hand, university administrators can begin to engage researchers interested in computation in developing a plan that leverages the local (discipline specific) needs with central core IT services and will also make financial and technical sense for the institution.

Where to Learn More

IT Engagement in Research: A Baseline Study, July 2006 at www.educuase.edu (ID; ERS0605)

Final Report: A Workshop on Effective Approaches to Campus Research Computing Cyberinfrastructure, April 25-27, 2006, at http://www.internet2.edu, Document: internet2-ccrc-report-200607.html

"Building the Campus Cyberinfrastructure Roadmap" August 2006, at http://www.educause/edu/cci

"The Challenge of Campus Cyberinfrastructure" October 2006, James Bottom, Patrick Dreher and Bonnie Neas at http://www.educause.edu/Library/DetailPage/666?ID=EDU06063

Gordon Bell and Jim Gray, "High Performance Computing: Crays, Clusters and Centers. What's Next?" Communications of the ACM, January 2002

Scientific Cluster Support at LBNL website http://scs.lbl.gov
Perceus website http://www.perceus.org

Upcoming ECAR study commissioned by Educause and due to be published in 2008 that will assess the level of CI deployment on campuses as well as current financial and strategic models used.

About the Author

Rosio Alvarez (ralvarez@lbl.gov) is CIO at the Lawrence Berkeley National Laboratory.
.

Acknowledgement

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.