Research Information Technology Proposal

Introduction

UMass Lowell continues to expand its focus efforts ~~on~~ in research, the Sciences, Engineering and Health areas continue be receive the largest funding support, while all other colleges are experiencing steady growth. As the participation in and complexity of research increases and the number of awarded grants grows, we have seen an expanding need for high performance computing on campus. The current academic technology infrastructure related to research is highly decentralized. It has been this way since the late-90s. This model has served the university well and in several departments decentralization is likely to be a better approach. There is however, a clear and growing need for a more compute-intensive infrastructure on campus and there is confidence that a shared high performance computing environment would be extensively used. Not only is there a need for this type of intense computing but there is a requirement for network infrastructure that can return the results to the desktop or share results with other institutions in which the university is collaborating.

Over the next 2 to 3 years, the network infrastructure will be upgraded to address this growing need. In fiscal year 2007 the university will be upgrading the outside plant fiber, doubling the speed of the backbone, adding a second internet connection and connecting to the nation-wide research network (Internet 2). The last component is using software tools to display the results. This is called visualization. Several faculty members have addressed this already but many others will need the tools in which to display the results of large computations or vast amounts of data in a format that improves overall research productivity. The Computer Science department has a number of tools in place that may be used to jump start this process.

UMass Lowell will need to improve its ability for compute-intense research in order to sustain the current research focus and better compete with other institutions for research grants. Increasingly, agencies prefer to award grants to higher education institutions that can demonstrate ongoing, centralized computing management and support rather than to those that use an individual approach. Funding agencies see it as inefficient.

Current Central Environment

The Olsen computer center renovation was completed in spring of 2006. This included dedicated, sensitive HVAC controls, UPS, power generators, electrical upgrades and structural renovations along with adding to the raised floor space for future growth.

Mission critical enterprise systems are housed in Olsen; in addition all servers for Computer Science and Physics also located there. Most recently, a small 48 node computing cluster for Physics was deployed.

Network upgrades planned for FY 2007 include: increasing the backbone bandwidth from 1 GB to 2GB (in 3 years that bandwidth will be up to 10GB), adding a second

Internet connection on South campus for redundancy and load balancing, and adding an Internet 2 connection to the Northern Crossroads (NOX) for research and collaboration.

Needs for academic research technology infrastructure:

1. Continue to fill out matrix especially the recommended solutions section for each area.
2. Download Condor software that pools and distributes idle computing resources through out the university. The first phase would be to test out the tool in a controlled environment to see if it is workable for UMass Lowell.
3. Funding to purchase hardware and software to startup the environment based on the computing needs outlined by the HPC workgroup.
4. Hire 2 academic technology personnel who understand research requirements. We need to define the skill sets needs and the organizational reporting lines.
5. Develop a visualization lab with tools to display results of large computations or data sets. CS has a widely recognized visualization – in business for over 20 years – and a lot of home grown and public domain visualization tools available.
6. Develop service level agreements that addresses availability and support for faculty researchers.
7. Define an ongoing funding model which includes funds from grants awarded to faculty using the infrastructure and governance that will sustain the environment.

Item description	Initial cost	Ongoing cost
Condor software	free	
Initial HP Cluster		
Additional research storage		
Cluster Backup storage, media, offsite		
Cluster software		
Operating system and maintenance		
Cluster management applications		
Non-research specific applications		
Visualization Tools		
Academic information technology staff		
Collaboration Tools		

Condor[1]

A key to high throughput is the efficient use of available resources. Years ago, the scientific community relied on large mainframe computers to do computational work. A large number of individuals and groups would have to pool their financial resources to afford such a computer. It was not uncommon to find just one such machine at even the largest research institutions. Scientists would wait their turn for mainframe time, and they would be allocated a specific amount of time. Scientists limited the size and scope of their problems to ensure completion. While this environment was inconvenient for the users, it was very efficient, because the mainframe was busy nearly all the time.

As computers became smaller, faster and less expensive, scientists moved away from mainframes and purchased personal computers or workstations. An individual or a small group could afford a computing resource that was available whenever they wanted it. The resource might be slower than the mainframe, but it provided exclusive access. Recently, instead of one large computer for an institution, there are many workstations. Each workstation is owned by its user. This is distributed ownership. While distributed ownership is more convenient for the users, it is also less efficient. Machines sit idle for long periods of time, often while their users are busy doing other things. **Condor takes this wasted computation time and puts it to good use.** The situation today matches that of yesterday, with the addition of clusters in the list of resources. These machines are often dedicated to tasks. Condor manages a cluster's effort efficiently, as well as handling other resources.

To achieve the highest throughput, Condor provides two important functions. First, it makes available resources more efficient by putting idle machines to work. Second, it expands the resources available to users, by functioning well in an environment of distributed ownership.

[1] Copied from Condor website at University of Wisconsin : http://www.cs.wisc.edu/condor/overview/

HPC Workgroup Requirements

Types of Applications (indicate real-time, parallel processing)	EE&AS	Work Environment	Computer Science	Physics and Applied Physics	Biology	Mgmt	Center for Atmospheric Research	Sub-millimeter wave lab	Recommended Solutions	Comments
Real-Time Weather Data flow – teaching and research										
High Resolution Weather Modeling – on our own cluster for now.										
Movement (ftp, scp) of large data sets (~1-2 gb largest for now)										
Computational Physics										
Intensive parallel processing with a lot of communication between nodes: MPI environment										
Both CPU and memory intensive			24G Min							
Using local Linux cluster and supercomputers at several supercomputer centers, such as NERSC, SDSC, etc.										
Data mining and imaging										
Cooperative computing work with LSU and other real-time and video conferencing in conjunction with working with other universities.										
Distributed SANS environment.										
Statistical simulations, not real time.										

Types of Applications (indicate real-time, parallel processing)	EE&AS	Work Environment	Computer Science	Physics and Applied Physics	Biology	Mgmt	Center for Atmospheric Research	Sub-millimeter wave lab	Recommended Solutions	Comments
Can utilize parallel processes										
Social networking										
Real-time video and imaging										
Remote Sensing										
Software requirements										
Backup software (and hardware?)										
SAS										
C, C++										
R (shareware)										
Matlab										
Fortran 90										
Fortran 95										
SIESTA										
VASP										
Gaussian										
MPICH										
FFTW										
SCALAPACK										
PGPlot										
ZPL (Shareware)										
Python										
Perl										
GNU compilers and binary util										

Research Information Technology Proposal

Types of Applications (indicate real-time, parallel processing)	EE&AS	Work Environment	Computer Science	Physics and Applied Physics	Biology	Mgmt	Center for Atmospheric Research	Sub-millimeter wave lab	Recommended Solutions	Comments
Java										
SPSS										
Estimated storage needs (terabytes)	1-2	1-2	30	30	10+					
Internet 2 requirement										
Data could come via I-2 (e.g. Weather)										
Real-time collaboration with other universities.										
Data transfer from the Linux cluster and also from computers at supercomputer centers to local machines										
Other										
Static IP addresses	5	5	10	15						
Open Source										
Gig connections to Olsen										
UPS for switch rooms and departments to failover to generators										
UNIX/LINUX support in IT										
Uniform access/wireless in big lecture halls										
Remote data entry/questionnaires										