Research Computing Newsletter Vol. 1, no. 4 August 2007

RCNews

August 2007

Year-end Roundup

FY 2007 saw significant progress for the research computing initiative at RIT. Advances were made in the key areas of cyberinfrastructure development, collaboration efforts, and community building. The most critical area was that of support, with the hiring of a new, full-time research systems administrator. And the ultimate measure of successful results will be the number of research faculty, staff, and students who use research computing resources and services to get their work done, perhaps best measured in papers, presentations, and publications generated.

I wish to thank all the researchers, technical staff, students, administrators, and administrative staff who have shown interest and support for advancing the research computing effort at RIT by their participation in its planning, implementation, or critiquing.

-*Gurcharan S. Khanna*, Director

Community

Communications

Communication in all its forms is key to transforming RIT to support a research culture that leverages advanced computing technologies. Research Computing actively disseminates pertinent information via its website, wiki, and quarterly newsletters. In addition, Research Computing holds quarterly Advisory Board meetings, quarterly researcher user group meetings, biweekly technical group meetings, and biweekly faculty lunch get togethers. The RC Seminar Series enables researchers to share their work with the RIT community. Informal email and formal web surveys are conducted throughout the year. These will continue going forward.

Global Connections

Grids

RIT actively participated in the creation and development of NYSGrid, a state-wide initiative to form a research and education cyberinfrastructure for a unified voice for funding, collaboration, and outreach. RIT served on the Board and the Executive Council, and designed and hosted its website. Research Computing was among the first to put a system on the grid. RIT is now part of the Open Science Grid and is exploring ways to connect to the TeraGrid. We are also working with IBM to develop a campus-wide computing grid. Research Computing is also deploying Access Grid collaboration systems across campus.

Internets

Research Computing actively leveraged RIT's membership in the Internet2 Consortium by attending its major meetings to keep us apprised of national research activities, such as the merging of Internet2 and NLR, and gives RIT exposure to the national and international research community. IT faculty and ITS staff joined Research Computing in the Collaboration Special Interest Group at Internet2 last fall. Research Computing's next Internet2 collaborations will focus on grid computing and high definition conferencing to international sites.

Collaboration

Planning

Many hours were devoted to consulting, meeting, and interacting with researchers who had potential projects that might take advantage of Research Computing resources now or in the future. This was part of an extensive effort not just to learn what the research needs are but to make sure researchers knew about the existence of Research Computing and its mandate to support researchers. Projects discussed covered a gamut from virtual theatre to molecular modeling to live stereo 3D broadcasts. [planning examples]

Proposals

Research Computing participated in three major funding initiatives last year.

- RIT: Research Computing (RC), The Center for Imaging Science (CIS), and The Center for Advancing the Study of Cyberinfrastructure (CASCI) jointly submitted an internal capital equipment request for funding to set up 5 Access Grid sites in CIS, Munsell Color Science Lab (MCSL), The IT Collaboratory (ITC) and CASCI to promote connectivity among on-campus departments that were split, and connectivity to external research communities using this technology. Funding was allocated for this project in Feb 2007 and the project is nearing completion in August.

- Microsoft Research: RC was invited to respond to this RFP in December 2006 to set up an Advanced Collaboration Technologies Center focussed on developing and extending ConferenceXP, a research collaboration tool for Windows. Although this $300K proposal was not awarded to RIT, it brought together researchers from Educational Technology Center (ETC), Information Technology (IT), National Technical Institute for the Deaf (NTID), and RC to work toward a common goal, resulting in communications and collaborations that are still going on today.

- CISCO: RC submitted a proposal to a research RFP in June 2007 for equipment donation and salary support to upgrade the research network to 10 Gigabit ethernet equipment for evaluating its performance for video, data-intensive computing, and grid computing. Participating researchers in this $300K proposal came from Networking, Security, and Systems Administration and ITS. The award for this proposal is still pending.

This coming year we will continuing to seek alliances with corporate partners with common interests in building a robust infrastructure for computing and collaboration.

Projects

Here are a sample of the projects that have used research computing systems. A database of researchers' projects is being developed on the Research Computing website.

Mike Foster, PhD student, Imaging Science (advisor Prof. Schott)

Use of Lidar data to geometrically-constrain spectral spaces for physics-based target detection

"I have used the Condor cluster to run a number of ray tracing algorithms on sparse

point cloud data sets. While I have been able to use the cluster to process these clouds in a number of days, performing the same calculations on my stand-alone PC would take months.

As such, the cluster has been a critical resource in my research, enabling me to defend on schedule. To that end, I will be publishing the results in my dissertation. I have had a paper based on this research accepted to the SPIE Conference on Imaging Spectrometry in San Diego."

Stefan Preble, Microsystems Engineering

Nanophotonic Device Simulations

"We are using the Large Memory Computer (LMC) to simulate nanophotonic devices. The devices confine light to dimensions on the order of 100 nanometers, enabling strong control of the light's properties (such as its speed, bandwidth, etc.). They are simulated using the Finite Difference Time Domain Method, which solves Maxwell's equations on a discrete spatial/time grid.

Our typical nanophotonic device is discretized with 28 million grid points (8 Gbytes of memory/2 Gbytes per processor). And it takes 1.2 seconds (vs. 4.5 seconds with only one processor) for each algorithm step, which corresponds to a simulation time of 1.4 to 14 days (100,000-1,000,000 time steps)."

David Rivshin, MS student, Computer Science (advisor Prof. Radziszowski)

Graph Reconstruction Numbers

"There are currently very few classes of graphs with which reconstruction numbers can be predicted, and therefore determination of them relies on exhaustive search. Each graph is independent from all others, and therefore the problem can be trivially parallelized to as many CPUs as there are graphs under consideration.

Computation of graph reconstruction numbers is very CPU intensive. I estimate that my other primary computation resource would have taken about 87 days to do what the cluster computed in about 4 days. Currently CPU limitations make it infeasible to do calculations on all graphs of 12 or more vertices."

William Basener, Prof., Mathematics

Analysis of Hyperspectral Imaging

"We used LMC to run algorithms on hyperspectral images. The data consists of 10,000 - 1,000,000 points in ~150 dimensional space. For some algorithms we compared all 1012 pairwise distances on LMC in about twelve days. The work involved students from an NSF funded REU at RIT headed by Darren Narayan of the School of Mathematical Sciences."

Publications

Although the Research Computing effort has only been in existence a short time, already a few results can be seen in terms of published research output.

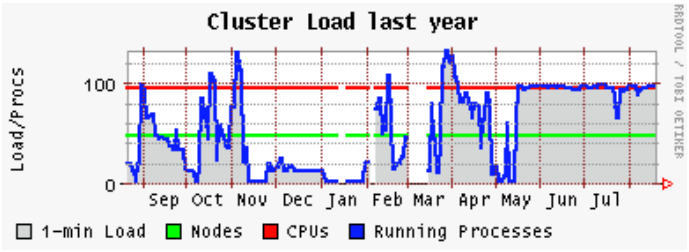
- (Center for Imaging Science) K.G. Baum and M. Helguera (2007) Execution of the SimSET Monte Carlo PET/SPECT Simulator in the Condor Distributed Computing Environment. Journal of Digital Imaging. Technical Note. September. abstract
- (Center for Imaging Science) Michael Foster. Use of Lidar data to geometrically constrain spectral spaces for physics-based target detection. SPIE Conference on Imaging Spectrometry in San Diego. Paper to be presented on 28 August 2007.
- (Mechanical Engineering) Stevens, Robert & Brown, Thomas. (2007) Modeling Thermal Transport at Single Interfaces and in Nanostructured Materials Using Molecular Dynamics and Monte Carlo Techniques. Material Research Society Spring Meeting, San Francisco. abstract ppt
- (University of Rochester) Uzilov, A. V., Keegan, J. M., & Mathews, D. H. (2006) Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. BMC Bioinformatics. 7:173. abstract pdf
- (University of Rochester) Mathews, D. H. (2005). Predicting a set of minimal free energy RNA secondary structures common to two sequences. Bioinformatics. 21, 2246-2253. Supplementary Material. abstract full text

Computation

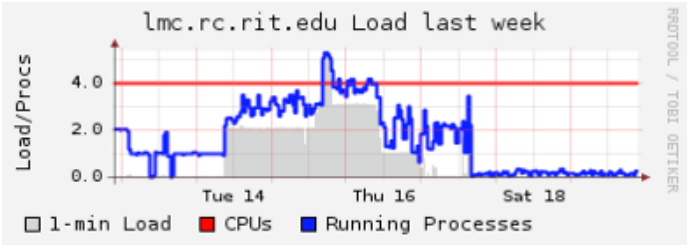
Usage

Research Computing is in the process of building up a research clientele for its supported systems. The number of users has grown to take full advantage of current resources and is driving the planning for expanded system resources.

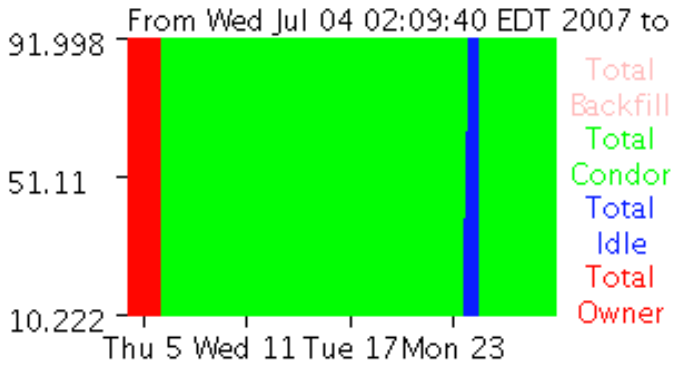
The number of users on the 47 node linux cluster has grown from essentially just one or two last year to over two dozen current users. This graph shows the system load for the last twelve months:



Only online since the end of June, our new large memory computer already has six researchers using it. We hope to support about a dozen or so users on this starter system once adequate software has been acquired, both individually and institutionally funded. This graph show load for one recent week:



The Condor "cluster" in CIS has been running thousands of jobs over the last year and will continue to do so as we expand its footprint in CIS to include Sun, Windows, and Macintosh workstations. We will also be incorporating Condor clusters in other departments under Research Computing support. Condor jobs are now running on the linux cluster. This graph shows cluster jobs in red, condor jobs in green, and idle time in blue:



Support

The most crucial development for Research Computing this past year was the hiring of a full-time staff member, Paul Mezzanini, who is the Senior Systems Administrator/Engineer for Research Computing. He is responsible for all central shared research computing systems and for interfacing with systems that connect to ours via the campus grid and external grids.

Brett Matzke joined the CIS computing support team in December 2006. Along with Jim Bodie, they are responsible for administering the systems in CIS, including lab clusters, the Condor cluster, fileservers, webservers, and mailservers, as well as user support and user systems support. We are integrating hardware, software, and support services where appropriate to leverage the combined resources of both CIS and Research Computing.

Rick Bohn worked part-time for Research Computing since May 2006. He was responsible for bringing the IBM cluster back to usability and since May 2007 was charged with responsibility for the CIS Condor cluster as well. He was also critical in bringing our NYSGrid cluster node online. Rick left us in August to take a permanent, full time position elsewhere in the area. We appreciate all the fine work Rick provided to us.

Software

The central shared systems, such as the linux cluster and the SMP machine, are being stocked with a useful complement of freely available common software as well as special requests for free software or user-provided commercial software applications. The list includes packages such as Blast, MPIBlast, MPI, LAM, IDL/ENVI, MODTRAN, Matlab, Mathematica, Meep, R, NAG, pgf77, Silvaco, JVM, Harpoon, Comsol, etc.

Research Computing is seeking to make highly critical research applications like Matlab more accessible and affordable by researching, documenting, and recommending a site-wide licensing agreement. This and other important software tools will make the hardware they run on more valuable to all researchers.

Systems

Research Computing's plan is to develop and deploy a diverse set of interrelated resources that will provide the range of computational systems necessary to satisfy the various needs of RIT researchers. These include central shared SMP machines and clusters, distributed workstations and access to systems and instrumentation accessible over the grid.

- "LMC" is our newest computer system, just coming online in June 2007. It is an SMP (Symmetric Multi-Processor) machine, similar to a dual processor workstation, but larger. It is targeted to run interactive and single threaded jobs that may be memory, cpu, or disk bound on a researcher's office workstation. It has two dual core AMD Opteron processors rated at 2.8 GHz, 32 GB of main memory, and 1.2 TB of raw disk. Currently it uses a 1 Gigabit network interface but will be upgraded to a 10 Gigabit interface soon.

- "Cluster" is the revitalized IBM bioinformatics cluster. It is now a cluster of 47 nodes, each built from dual P3 1.4 GHz processors and 512 MB RAM. This linux cluster was designed to run parallel computing jobs that are tightly coupled and use the MPI over high speed interconnects (currently a 1 Gigabit ethernet switched network).

  It also runs multiple serial jobs that do not need to communicate with each other but expect to run on dedicated processors. Completely indeterministic serial jobs will also run here in the background under Condor when other queued jobs are not utilizing all of the processors. Upgrade plans include replacing the interconnect switch with a 10 Gigabit one.

- Condor flocks are groups of workstations that individually volunteer to run pieces of jobs that don't have to communicate with each other while running. The Mac Lab in CIS and the Research Computing IBM cluster recently joined the Sun/Solaris workstations in CIS and CS to create about a 300 workstation pool of machines to perform High Throughput Computing tasks. The RC Condor1000 Project aims to create a 1000 workstation strong pool of machines at RIT by Fall 2007.

- "Gcluster" is RIT's five node testbed system that is linked to NYSGrid. Test job submissions have successfully been made to the University at Albany test cluster, and other sites will be available via NYSGrid.

- Future systems will support webservers, wikis, statistics collection, media servers, other operating systems, storage, and backup servers.

Network

The already robust network infrastructure at RIT has been upgraded in parts to support advancing research needs. The entire network in CASCI (Bldg. 74) has been upgraded to Gigabit throughout, as has most of CIS (Bldg. 76) and parts of Color Science (Bldg. 18). In addition, CASCI has been partially enabled for IPv6 protocols and a few routers across campus have been enabled for future 10 Gigabit network connections.

Future support will include high quality video streaming services, high bandwidth data transfers, low-latency server interconnects, and external connectivity to the TeraGrid that will be enabled during the coming year.

Collaboration Technologies

The Interactive Collaboration Environments Lab (ICE Lab), a lab in CASCI, is the aegis for research and development activities especially focussed on advanced collaboration and network technologies. It is designed to provide a center for teaching and learning, research and development, practical applications, and evaluative studies. The overall objective is to create an advanced collaborative infrastructure for the RIT campus.

"Kids on the Grid" use the Access Grid site in CASCI for Take Your Daughters and Sons to Work Day on April 26, 2007.

The desire to highlight and showcase advanced collaboration technologies that are also practical for connecting collaborators both on campus and off is widespread at RIT. More than half the Colleges have specific projects or plans in place for some form of advanced collaboration technologies. The ICE Lab has had discussions about coordinating such efforts with entities in all 8 colleges and CIMS, and is actively working on projects with NTID, CoB, CoS, and GCCIS that will use ICE Lab expertise and technologies.

This past year's focus was on practical applications of collaboration technologies, in the setting up of five enhanced Access Grid nodes in four buildings. That goal is about 80% complete and will be 100% complete by Fall 2007. The ICE Lab was fortunate to be able to employ two co-op students through Research Computing funding during the latter part of the fiscal year to help carry out these projects.

During the past year, the ICE Lab hosted over 20 events, including:

- Demos for Kodak, IBM, Ohio Center for Technology & Science, Digital Rochester, U. Albany, Golisano Dean's Council
- Broadcasts of RC Seminars from CASCI
- Seminars from UK Manchester, NSF workshops, lectures on Art on the Grid
- Kids on the Grid (for Take Your Daughters and Sons to Work Day)
- CS Class final for Virtual Theatre

The ICE Lab received supplementary funding from an NSF grant and from NTID/PEN International for two outside projects focusing on issues for the deaf and hard of hearing.

Next year's focus will be on two research and development projects:

- High Definition live video over IP for connecting public kiosks on campus.
- 10 Gigabit ethernet networks for HD video and for data-intensive and high performance computing.

Funding is currently being sought to accomplish these goals.