SHARCNET™

SHARC BYTES newsletter

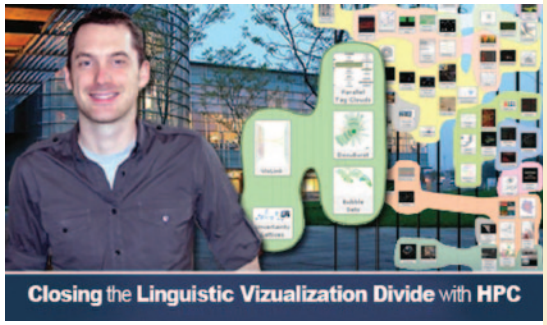
SUMMER 2010 VOL. 8 ISSUE 1

Closing the Linguistic Visualization Divide: HPC Aids in Information Visualization

BY KIMBERLY WHETSTONE, STUDENT COMMUNICATIONS OFFICER, SHARCNET

Modern technology has created a surge in multi-language textual data that can easily overwhelm users when interacting with a daily over abundance of information. This excess has fuelled the increasing demand for information management. News articles have been compressed into compact statements, twitter posts are 140 characters, and shorthand has lent itself to the web; however, the problems for researchers extend beyond the organization of language. Linguistic analysts are expected to create solutions to enhance the communication and understanding of linguistic data. As text visualization continues to permeate our lives we need to improve the way we present textual relationships. High performance computing (HPC) allows for efficient processing of computationally taxing data sets such as digital libraries; a necessary tool when developing new large scale visualization systems.

Dr. Christopher Collins is an assistant professor at the University of Ontario Institute of Technology and was recently named SHARCNET Research Chair in Information Visualization. Using his background in computational linguistics, techniques from information visualization, and research in human-computer interaction, Dr. Collins is working towards creating a toolkit for analysts, scholars, scientists, and everyday internet users to better understand the content and relationships within large scale repositories. Success in previous design studies using smaller data sets include: Uncertainty Lattices; which enable informed decisions about the quality of computational outputs, DocuBurst and Parallel Tag Clouds; which spatially



Christopher Collins
University of Ontario Institute of Technology

organize document content for comparison, and Bubble Sets and VisLink; which address visualization for natural language processing research. Using the smaller design studies as reference, Dr. Collins is branching out to address the problem of large-mixed datasets with the aid of SHARCNET's computational resources.

Dr. Collins' current research focus in visualization for large scale text databases is concerned with bridging the linguistic visualization divide; "The gulf separating sophisticated natural language processing algorithms and data structures from state-of-the-art interactive visualization design". In other words, he feels that the popular visual text analysis programs either use oversimplified algorithms or the linguistically sophisticated algorithms lack a clear visual interface. HPC will allow Dr. Collins to pre-process library information systems, the focus of his study, while the visualization systems are still in the development stages.

Continued on page 6

in this issue:

Message from the Chair of the Board

Paul S. Maxim,
Chair of the Board

"HPC is not a luxury, but a necessity in this era where advancements are short-lived, and 'innovation' and 'competitive advantage' are slogans of the day."

Over the past decade, High Performance Computing has emerged as a game-changing technology that transcends discipline boundaries. It provides competitive advantages in both developed and emerging economies to accelerate research and innovation in academia, business and industry. HPC is a core element of advanced research programs in disciplines as diverse as chemistry, demography and astronomy. It enables research that could not otherwise be done, be done safely, or be done economically. HPC has been used to assess the safety features of automobiles in crashes, the effectiveness of materials for storage systems for radioactive material, or to screen hundreds of drug candidates. We often combine HPC with other research approaches-such as laboratory studies-in areas like genomics or proteomics. It has become essential component for breakthroughs and is an increasingly critical component of competitive business and industry.

SHARCNET continues to be a key part in this equation by providing the essential technology and support services to academia and business partners who require HPC to advance their research. The number of users who rely on high performance computing to advance their research continues to grow. HPC is not a luxury, but a necessity in this era where advancements are short-lived, and "innovation" and "competitive advantage" are slogans of the day.

Over the past several months, we have been working closely with our consortia counterparts to advance the goals and objectives of Compute Canada. A midterm review of Compute Canada was recently conducted, and several recommendations have been put forward. One highlight of the review underscores the need for a longer-term financial sustainability plan for HPC. Indeed, Compute Canada's financial model is arguably its weakest link. Investments in HPC in Canada are between 35% - 50% that of our G8 competitors, putting us at a significant international disadvantage.

Industry Canada also led a public consultation process to develop a strategy for the digital economy. The details of this can be found at http://de-en.gc.ca. Compute Canada submitted an idea titled "To Compete You Must Compute" which describes the value and necessity of HPC to the digital economy. An on-line forum accepted public responses to the questions and issues raised by the consultation paper. The HPC community made their point by ranking Compute Canada's submission the number one idea overall. We hope this helped to raise the profile of HPC with funders and policy makers, and will lead to greater efforts to enhance Canada's HPC capacity.

Paul S. Maxim
Chair, SHARCNET Board of Directors
Associate Vice-President: Research
Wilfrid Laurier University

Scientific Director's Message

SHARCNET continues to be busy on a number of fronts, most notably working with Compute Canada and the other consortia on national HPC initiatives and lobbying efforts to ensure continued funding for SHARCNET and HPC. This remains our most significant challenge at present and your continued support and advocacy is vital.

We recently announced the results of SHARCNET's Round II Dedicated Programming Support competition, at www.sharcnet.ca/my/research/programming. This programme provides support for computational projects of exceptional potential that will have lasting impact and value. A central goal is to enable projects from disciplines that are traditionally not major users of HPC. We awarded four projects this round, one in support of an application to the G8 Exascale initiative.

SHARCNET users will also notice that the second National Resource Allocation Call for Proposals has been issued by Compute Canada. This program is designed to award both computational time as well as additional storage to researchers and research groups based on scientific merit. Researchers who require more resources than they would normally receive within standard access policies are invited to submit an application. To view the call and for more information please visit the Compute Canada website at www.computecanada.org. This is a companion program to SHARCNET's continuously running Small Dedicated Resources and will replace SHARCNET's Large Dedicated Resources call.

On the equipment side, I am pleased to highlight the following:

- We have ordered our new general-purpose cluster, to be named "Orca", from our share of the CFI-NPF funds and are on track to have the system available later this Fall. This AMD system comes with over 7,680 cores in 320, 24-core nodes, and more than 65 peak Tflops.
- Our new "Kraken" cluster continues to grow. It brings together a number of our systems into one large cluster. This makes it easier to keep all hardware busy and users don't need to select which of our many smaller clusters to log into.
- The new storage location called global work is now available on many clusters and the remaining systems will be converted in the coming weeks. This storage provides 1TB of space to each user and will be available on any cluster much like the home storage. Users should continue to use the cluster specific scratch space for I/O intensive jobs as it is the fastest storage available.

SHARCFest 2010, our annual outreach campaign of workshops, conferences and sponsored events, was held over the month of May. Our premier event, SHARCNET Research Day, drew a record number of attendees and we look forward to offering the event again next year. Our HPC Summer School, held a second time at Sheridan College, was another huge draw and feedback on the events continues to be extremely positive.

As part of our continuing outreach efforts, SHARCNET was pleased to be one of the sponsors and organizers of "MIND THE GAP", co-hosted by the University of Alberta and WestGrid from May 10-14, 2010. This was a multidisciplinary workshop bridging the gap between High Performance Computing and the Digital Humanities.

Another groundbreaking term of Coast-to-Coast seminars has begun, which includes lectures from distinguished researchers and leading scientists from across the country, delivered locally through your AccessGrid facilities. Events are published on our online calendar and you can register for any of these free seminars by e-mailing help@sharcnet.ca.

On a sad note, SHARCNET lost a supporter and champion of HPC with the sudden death of Ross Hallett on August 6th. Ross was a very good friend of SHARCNET, as one of our original Board members, and most recently he served on our Chairs Selection Committee with enthusiasm and dedication. SHARCNET has made a donation to the University of Guelph "F. Ross Hallett Scholarship Fund in Physics" in his memory.

Hugh M.P. Couchman
SHARCNET Scientific Director
Fellow, Canadian Institute for Advanced Research
Professor, Physics and Astronomy, McMaster University

Hugh Couchman,
Scientific Director

Seeking a Greener Turbine

BY NATALIE OSBORNE, STUDENTS PROMOTING AWARENESS OF RESEARCH KNOWLEDGE (SPARK) PROGRAM AT THE UNIVERSITY OF GUELPH

Turbines power our daily lives, whether they're combustion engines helping us travel the world or harnessing wind energy to generate electricity for homes. SHARCNET's computational resources are fueling research aimed at better understanding the dynamics of wind and combustion turbines. The findings could one day lead to more efficient, environmentally friendly designs.

Prof. Stephen Tullis, Department of Mechanical Engineering at McMaster University is investigating the physics behind pre-mixed combustion, in which fuel and air are mixed before being ignited.

Pre-mixed combustion is used widely, from natural gas turbines to the spark ignition in a car engine. Safely and effectively reducing the emissions of pre-mixed combustors is a priority for environmentalists and engineers alike.

"Running these combustors leaner or cooler, with more air than is needed to burn the fuel, can reduce the amount of harmful emissions such as nitrogen oxides," says Tullis. "However, it also makes the flames more unstable, and this instability limits the extent to which lean combustors can run."

Tullis is examining the interaction between turbulence and combustion to understand exactly where the instability comes from. Turbulent forces from air "wrinkle" the flame, increasing its surface area and causing it to burn faster, which in turn generates more turbulence. Researchers consider this a multi-scale problem because they must calculate forces ranging from a tenth of a millimetre, such as the burning flame's chemical interactions, to turbulent forces as large as the cylinders in an engine.

Before engineers can design improved pre-mixed combustors, they need reliable computer models of the flame to replicate the flame dynamics in the combustion chamber. However, physicists don't fully understand the dynamics, so they can't estimate the flames' behavior.

That's where Tullis comes in. He's using what's called direct numerical simulations (DNS), which solve the multiple equations related to combustion and turbulence. These complex calculations get very large, but provide a theoretically precise picture of combustion.

"We need SHARCNET's computational horse power because we're not making any simplifying assumptions, but calculating the solutions to the governing equations directly," says Tullis. "We can then develop accurate models of the flame and verify them with the results from the DNS."

Tullis also uses computer simulations in combination with full-scale experiments to examine the aerodynamics of wind turbines for power production. He's examining small scale vertical axis turbines, which are better adapted to the fluctuating speeds and directions of wind in cities and suburbs. He's simulating the variable flow of "dirty" air through and around the turbine to gain a better understanding of the unsteady aerodynamic effects on the turning blades.

Vibration is a concern with vertical axis turbines. It can cause mechanical and installation problems and even turbine failure. Vibrating turbines are also more expensive because they require larger and stronger supports.

"We're trying to understand the loading that causes this vibration and how we can mitigate it," says Tullis. "If we can stop it vibrating, we can design models that are lighter, taller and cheaper."

Making wind turbines a more practical alternative energy source and decreasing greenhouse gasses with more efficient combustors are just some of the eco friendly goals of Tullis's research. A more inherent understanding of the physics behind turbines will propel this technology in years to come.

Research contributors include Dr. Steward Cant of Cambridge University and the Cleanfield Energy Company.

ResearchDay
"HPC Innovation for Research"
York University • May 6, 2010

SHARCNET Research Day 2010

BY MICHAEL HASLAM, SHARCNET SITE LEADER, YORK UNIVERSITY

The annual SHARCNET Research Day was held on May 6th at York University. Research Day is SHARCNET's premier annual event at which professors, postdocs and graduate students meet to learn about each other's high performance computing related research. The theme of this year's meeting was "HPC Innovation for Research".

SHARCNET was pleased to welcome the first keynote speaker, David H. Bailey, Chief Technologist of the Computational Research Department at the Lawrence Berkeley National Laboratory. In his presentation "Computing as the Third Mode of Scientific and Mathematical Discovery," Dr. Bailey gave an impressive overview of current HPC applications and interesting glimpses on the future of high-performance computing.

David Bailey delivering his keynote address

The second keynote address, entitled "Modelling membranes, proteins and biology: studying the fundamentals by computer experiment", was delivered by Mikko Karttunen, Associate Professor of Applied Mathematics at UWO. In this presentation, Dr. Karttunen showed the audience how he uses SHARCNET's HPC equipment to calculate microscopic properties of biological systems.

The event also featured five parallel sessions of contributed talks covering topics in applied mathematics, computer science, fluid dynamics, many-body physics, molecular modeling, and biochemistry. For the second consecutive year, a special session in Digital Humanities was held.

Several awards were presented for best student poster presentations. The conference was a success and much fun was had by all.

SHARCNET Welcomes New Research Chair

Christopher Fletcher

"My research at Waterloo will use computer model simulations to better understand how climate change will impact Canada, the US and Europe on seasonal-to-decadal timescales."

SHARCNET is pleased to welcome its newest Research Chair, Dr. Christopher Fletcher, who joined the Department of Geography and Environmental Management at the University of Waterloo, effective August 1, 2010.

Dr. Fletcher brings excellent qualifications and experience to the role of SHARCNET Research Chair, having established an excellent research reputation in the application of high performance computing to large-scale climate models for climate change research. We anticipate Dr. Fletcher making a significant contribution to the research and outreach activities of the newly launched Interdisciplinary Centre on Climate Change at the University of Waterloo.

Funding for SHARCNET Research Chairs is provided through the generous financial contributions of our private sector partners, Hewlett Packard and Silicon Graphics.

Closing the Linguistic Visualization Divide, continued from page 1

Library systems are the backbones of libraries that would cease to exist without an organized search and retrieval function. However, most complex search queries that return low results require research librarians to physically search through pages to find the desired information. Despite recent digitization of texts, there are still discrepancies in the standardization of tags and summaries. Dr. Collins plans to address the visualization challenges of large scale data sets through development and deployment of web applications. In addition, he is interested in creating new ways to analyze and visualize documents containing both graphics and text. By coupling text analysis systems with interactive visualizations, library patrons and research librarians will be able to find the relevant materials they need more efficiently.

Another fundamental direction of Dr. Collins' research is to learn how visualization fits (or doesn't fit) into the scientific workflow. The first place he will look is to other SHARCNET researchers to determine how visualization is used in high-performance computing. This study will determine the usability of HPC resources and how we can lower the barriers to make HPC techniques more accessible to non-computer science oriented researchers. Not only is this study important to the understanding of human-computer interaction for visualization purposes, but the usability of HPC in other disciplines is significant for the growth and deployment of SHARCNET in future high performance computing endeavours.

Moving forward Dr. Collins remains interested in the ephemeral and growing world of social media. His work in visualization systems using HPC could later be applied to the internet to illustrate correlations within a dataset as large as the World Wide Web. For instance, what are people saying about the latest government report? How many people are talking about it? Are the comments positive? What types of mediums are being used? These are all questions that could be answered using visual links through the development of Dr. Collins' work in large scale visualization systems. It is through the visualization of information that we can exemplify connections between texts to alleviate information overload; and with clearer information visualizations will come a greater understanding of context.

For more information on Dr. Christopher Collins or his research please visit http://faculty.uoit.ca/collins/

software carpentry

Software Carpentry Partners with SHARCNET

BY GREG WILSON, SOFTWARE CARPENTRY

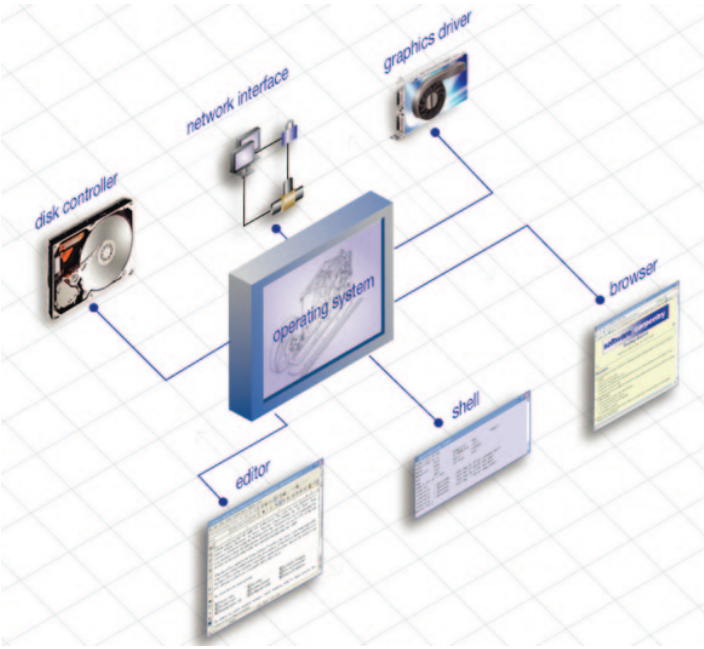
Computers are as important to modern science as telescopes and test tubes. Unfortunately, most scientists are never taught how to use them effectively. After a generic first-year programming course, most scientists have to figure out for themselves how to build, validate, maintain, and share complex programs. This is about as fair as teaching someone arithmetic and then expecting them to figure out calculus on their own, and about as likely to succeed.

Scientists trying to take advantage of tera- and peta-scale computing pay a particularly high price for their lack of background skill. How to use parallelism effectively is still an open research problem in computer science; when it's combined with leading-edge problems in physics, mechanical engineering, or biochemistry, the complexities compound more often than they cancel out.

It doesn't have to be like this. Since its first run at Los Alamos National Laboratory in 1997, the Software Carpentry course has taught scientists and engineers the basic skills they need to use computers more effectively. Its topics are those things left standing after various bandwagons have rolled through: version control, automated builds, unit testing, how to crunch common data formats, small-scale software project management, and so on.

This training has consistently had an immediate impact on productivity by making what people currently do less onerous, and new kinds of work feasible. The materials, which are available under an open license, have been viewed by over 140,000 people from 70 countries, and have been used at Cal Tech, the Space Telescope Science Institute, and other universities, labs, and companies around the world.

Software Carpentry is now undergoing its fourth major revision. Thanks to funding from SHARCNET and other sources, Dr. Greg Wilson started working full-time in May 2010 to translate the online notes into short videos, and to update and expand the examples and exercises. The new format will be beta tested starting in October 2010, when 40 students from Ontario universities will have the chance to work through the course online. If you, your colleagues, or your graduate students are interested, please check out the content at http://software-carpentry.org/blog/, or contact the course administrator at info@software-carpentry.org.



SHARCNET Summer School 2010

BY BAOLAI GE, HPTC CONSULTANT, SHARCNET

The Trafalgar campus of Sheridan College was this year's host site for SHARCNET's annual High Performance Computing Summer School. The event brought close to forty students together to spend a week studying and learning about supercomputing tools and techniques.

The Summer School ran from May 31-June 4, and offered intensive courses on various HPC subjects. As before, message passing interface (MPI) and threaded programming with OpenMP were provided as core courses. The increasingly popular GPU programming courses gave the attendees a choice of either using CUDA and/or using OpenCL. As part of the tradition, Summer School continues to offer courses on the theory of parallel computing, debugging, and Matlab in a distributed environment. As in previous years, participants had an overwhelming interest in MPI for programming distributed systems (clusters) and OpenMP, which is the industrial standard API for multi-threaded programming on shared memory systems.

Programming supercomputers – distributed memory systems or clusters, shared memory systems, and GPUs – is essentially implementing numerical algorithms with data movement and coordination among computing units. If designing an algorithm is about translating mathematical formulae into numerical procedures at a high level, implementing the algorithms on supercomputers is like wiring bits and pieces together to make them work synchronously.

Unlike the classic programs of the past few decades, the writing process for programs that can run on multiple CPUs together requires a good understanding of the science, the ability to solve problems in parallel on sub-problems, and the technique of coordinating computing units in a controlled, synchronous manner. This requires the programmer to understand how data is accessed in memory, moved from one location to another, and from one computer to another, over the network. Due to the current nature of programming and data communication, it is now expected that scientists and engineers understand data communications in order to write computer code that can operate on multiple CPUs concurrently.

The learning curve for programming parallel systems can be quite steep. The purpose of Summer School is to provide students with a learning environment in which they can concentrate on learning programming concepts while practicing their skills under the guidance of on-site experts. The courses were intensive and materials that were normally delivered in several lectures were covered within two days. The cores courses, MPI and OpenMP, were taught by instructors who have taught the courses for years. Dr. Jonathan Dursi, from SciNet, has taught the MPI course on many occasions both nationally and internationally. Dr. Hartmut Schmidt and Dr. Liu Gang, from HPCVL, have taught the course on threaded programming for years at many institutions throughout Ontario. The courses received overwhelming complimentary comments from the attendees.

Previously, the event was called the Fall Workshop. In order to focus on major subjects and give participants enhanced hands-on training, the annual workshop was extended to a full week. SHARCNET changed the name of the program to Summer School in 2007. As a demonstration of its commitment to training and support, SHARCNET has provided a limited number of financial subsidies for lodging and transportation for those students travelling from distant institutions.

The hosting institution of Summer School used to be rotated amongst SHARCNET's partner schools. However, over recent years, Sheridan College has become the "go to" site, primarily because of its central location, excellent amenities, and the beauty of the campus.

For more information and pictures from this year's event please visit: http://www.sharcnet.ca/events/ss2010/

Shanghai Supercomputer Center Visits SHARCNET

BY KIMBERLY WHETSTONE, STUDENT COMMUNICATIONS OFFICER, SHARCNET

This past July, SHARCNET hosted Director, Xi Zili and HPC Engineer, Kou Dazhi from the Shanghai Supercomputer Center (SSC) on a reciprocal visit to Canada. They spent two days touring SHARCNET institutions including The University of Western Ontario, University of Waterloo, and McMaster University. Building on existing collaboration, the SSC visitors were informed of new advancements within HPC, as well as, the work of cutting edge researchers who are using SHARCNET's technology in their fields.

Of particular interest to Director Zili is the field of computational finance, as the SSC is working to reposition and expand their role within the regional and national economy of China. Specifically, Zili wants to know how researchers obtain data from financial institutions, to what degree financial data is made available to the public, and how much interest there is for financial institutions to invest in research. These questions will help the SSC as it further explores the idea of creating a financial data centre to facilitate data processing and provide business consulting services that would complement existing agencies. Through SHARCNET, they would like to engage with people who have experience, insight, and connections to the financial industry to help them with their answers.

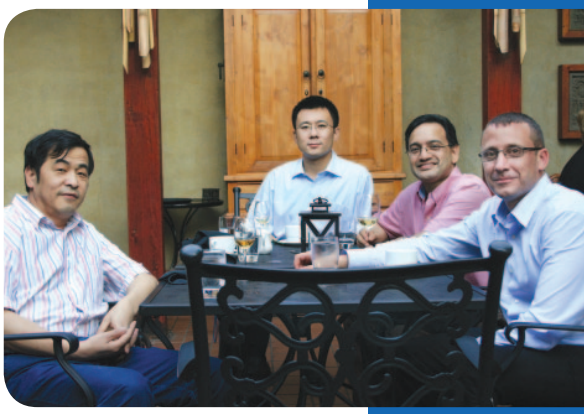
Knowing these interests prior to their visit, Baolai Ge was able to set up various meetings and social events for Zili and Dazhi to network and communicate with leading researchers. Mark Ressor of The University of Western Ontario's Department of Applied Mathematics was at the official welcoming function for the SSC visitors and was on hand to answer some of the financial questions they had.

The following day, Director Zili visited the University of Waterloo to meet with a group of researcher at the Waterloo Research Institute of Insurance, Securities and Quantitative Finance (WatRISQ). Dr. Ken Seng Tan, Associate Director of WatRISQ, gave a brief introduction to the multidiscipline research institute and research activities followed by a round table discussion with leading researchers in computational finance.

After two full days of site visits and meetings, Director Zili finished his visit with SHARCNET, while Kou Dazhi remained behind with Baolai to continue discussing technical aspects and potential areas for collaboration.

Kou demonstrated SSC's "Xfinity", a web portal they have developed inhouse, with a desktop environment that features user file management for the transfer between desktop and remote HPC systems, account management, job submission, system information, and discussion forum. The development started a few years ago, aiming to provide users, especially industry users, with an environment that allows SSC to easily manage their projects and enables commercial software users to submit jobs from a desktop environment they are familiar with. Aside from the features that are seen in some open source online desktop environments, the HPC jobs submission feature appears to be appealing to those who do not need interactive command line access. This enables users of commercial software packages to submit jobs from the web portal without third party software such as file transfer client, which is extremely useful and convenient for Windows users.

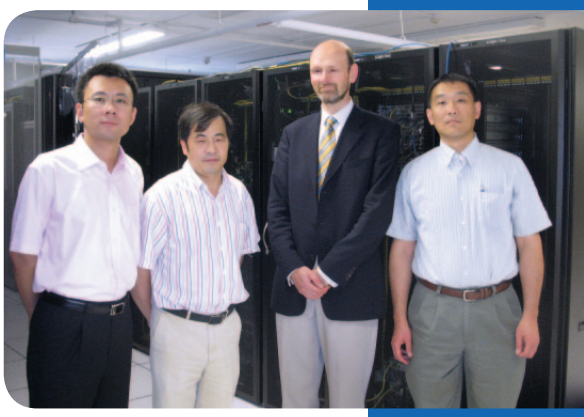
"Xfinity" was created by the research and development department of SSC and could potentially be useful for SHARCNET as well. After a few years, the product is becoming mature and has gained the acceptance of users. The SSC is keen on seeking out more usage of the product. This leaves the option for immediate collaboration; however, if we are to pursue this, some tailoring work needs to be done before we would attempt to run the portal on SHARCNET.



From left, Xi Zili (SSC), Kou Dazhi (SSC), Shantanu Basu (UWO), Dan Sinai (UWO)

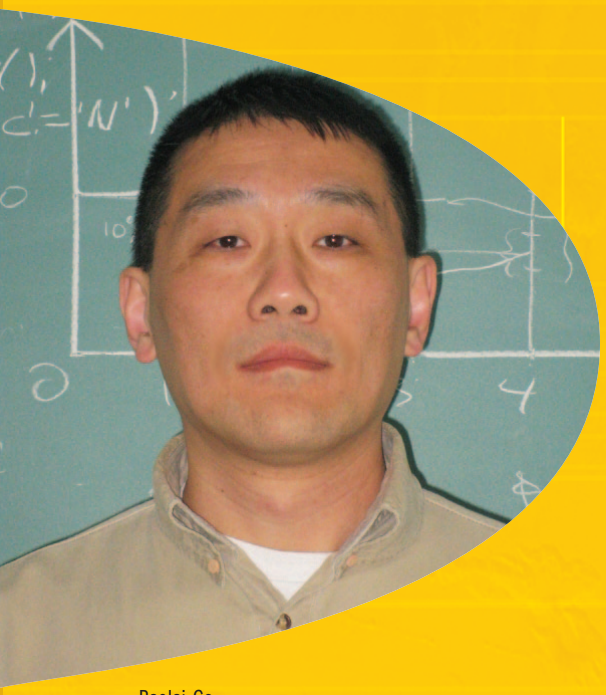
Kou and Baolai also discussed other areas where collaboration may be possible, such as creating libraries and tools for GPUs, and user facing project management with a common format for graphs and tables.

Another significant high point of the SSC visit came from a meeting with Stewart McIntyre, co-PI of the Science Studio project. Science Studio wants to establish collaborations with its Chinese colleagues at the Shanghai Synchrotron, a client of SSC. This bridge model would bring researchers together from The University of Western Ontario through the SHARCNET/SSC partnership and connect them with researchers in China.



From left, Kou Dazhi (SSC), Xi Zili (SSC), Hugh Couchman (SHARCNET), Baolai Ge (SHARCNET)

After all of the meetings and networking opportunities, our guests returned to Shanghai on July 16. Collaboration will continue between SHARCNET and the Shanghai Supercomputer Center as we work with them in developing HPC services.



Baolai Ge

Swimming with SHARCS: Spotlight on Baolai Ge

Baolai Ge is one of the first high performance technical computing consultants hired by SHARCNET. With close to ten years experience working with SHARCNET, along with a background in applied mathematics, scientific and technical computing, and distributive systems, Baolai is able to assist many researchers in multiple aspects of HPC.

From the beginning, Baolai has been leading and organizing SHARCNET's annual training workshops, including the annual summer school. Through these events, Baolai has been able to establish relationships with many of our researchers to provide them with the technical skills and support they need to continue their work.

Since 2007, Baolai has been managing SHARCNET's dedicated resource allocations and is the main technical contact for all awarded projects. Working with the researchers constitutes a large portion of Baolai's time and many researchers are grateful for the assistance he is able to provide.

Lance Lochner, a Professor in Economics at The University of Western Ontario, notes, "Baolai has been very helpful in responding to a variety of issues that have arisen, from helping us adapt to system updates and changes to providing personalized tutorials on MPI debuggers and the use of Makefiles. With such a large organization and numerous machines located throughout Ontario, it is important to have someone local who can help with both the little things that pop up regularly and larger concerns that determine the success or failure of a project. Baolai fits that role perfectly."

"With such a large organization and numerous machines located throughout Ontario, it is important to have someone local who can help."

In addition to providing support for academic researchers, Baolai liaises with our private sector users. He has been working with a number of SHARCNET's industrial researchers including: The Bank of Canada, General Motors Canada, Shanghai Supercomputer Center and Trojan Technologies.

Baolai, along with a team of highly skilled technical consultants deployed across SHARCNET, are the front line in providing user support, tools and training for the HPC community. Feel free to contact one of our consultants, or browse the extensive amount of online materials they have developed at www.sharcnet.ca/help.

 BY GARY MOLENKAMP, SYSTEMS ADMINISTRATOR, SHARCNET

What is Global Work?

At SHARCNET, global work is the implementation of a single, consistent filesystem for all /work storage. Currently, each cluster has three separate storage pools: /home, /work, and /scratch. Each storage pool provides different performance characteristics and capacity so each has different resource limit policies to reflect it. At one end of the spectrum, /home is consistent and available across all clusters but is a limited size and performance. As such, it has a strict quota limit and should not be used for running input/output (IO) intensive jobs. At the other end of the spectrum, /scratch is a filesystem per cluster and is usually a parallel filesystem directly attached to that cluster's high speed interconnect. Scratch space is usually large and quota free but has an expiration policy to keep the resources available. /work attempts to fill the middle ground between these two. It has a larger quota than /home but still limited, longer expiration times than /scratch, and higher performance than /home storage. To date, /work has also been built as a per cluster storage pool occasional using the same filesystem hardware as /scratch. Global work adds a single, consistent view of /work files per user across all clusters while still filling the middle ground between /home and /scratch storage pools.

Global work is composed of two geographically separate 300TB lustre 1.8based parallel filesystems. Each of the filesystems is composed of 32, 10 terabyte fileservers and two administrative servers for metadata and management. Each SHARCNET user is assigned to one of the two storage pools, but from the user perspective, work data will always be mounted as /work/<username>. The initial quota limit per user is 1TB, with exceptions managed on a case by case basis (similar to home quotas). Users are able to independently manage the degree of parallelism of their files by controlling the number of fileservers (stripesize) that contain a single file. By default, the stripesize for a user's work space is set to one; however, by using the command 'lfs setstripe -c # <filename>' the stripesize may be increased for a file or directory. A similar command 'lfs getstripe' will display the degree of parallelism for a file. A higher stripe size increases the throughput of the IO with a trade off in how the data is stored. A reasonable value for parallel writes is on the order of three to four for our current systems.

Users may notice varying degrees of performance with the new global work depending on the IO patterns that are used for a compute job. As with most Lustre based parallel filesystem, lustre tends to favour large files that are striped across multiple file servers versus large quantities of small files. For example, a 100GB file striped across four fileservers will have good performance while hundreds or thousands of small 100byte files will not. If data requirements dictate the need for many small files, then it is recommended to archive the data using a utility such as tar when the data is not needed in the short term. As well, consider using /scratch or /home where appropriate.

Due to the disruptive nature of removing one filesystem and replacing it with another, cluster conversion to global work is occurring during cluster software maintenance schedules. With each cluster converted, the existing cluster's work storage is relocated to a /work/<username>/<clustername> directory to retain the data. For some users who use many clusters, this may mean a temporary duplication of some data and they will need to manage it properly to stay under quota limits. The conversion of several smaller clusters such as: hound, mako, angel, and kraken, have already been completed. Work has begun to start converting some of the larger clusters such as whale and saw with additional clusters in the conversion planning stage. Please watch the system notices for updates of future global work conversions.

SHARCNET™

Shared Hierarchical Academic Research Computing Network

SHARCNET Funding

Canada Foundation for Innovation (CFI)
Ontario Innovation Trust (OIT)
Ontario Research Fund (ORF)

Private Sector Partners

Bell Canada • www.bell.ca
Hewlett Packard • www.hp.com
Nortel Networks • www.nortel.com
Optical Regional Advanced Network of Ontario (ORANO) • www.orion.on.ca
Platform Computing • www.platform.com
Quadrics Supercomputing World • www.quadrics.com
Silicon Graphics • www.sgi.com

Academic Partner Institutions

The University of Western Ontario
University of Guelph
McMaster University
Wilfrid Laurier University
University of Windsor
Fanshawe College
Sheridan College
University of Waterloo
Brock University
University of Ontario Institute of Technology
York University
Lakehead University
Laurentian University
Trent University
Ontario College of Art & Design University
Perimeter Institute
Nipissing University

compute • calcul CANADA

SHARCNET is one of seven HPC consortia in Canada that operates under the umbrella of Compute/Calcul Canada.

SHARCNET™

The University of Western Ontario
Western Science Centre
London, ON Canada
N6A 5B7

T. (519) 661-4000
F. (519) 850-2500
E. info@sharcnet.ca
www.sharcnet.ca