WestGrid

www.westgrid.ca

at
Simon Fraser
University

NOUS·SOMMES·PRETS

www.cecm.sfu.ca

Western Canada Research Grid

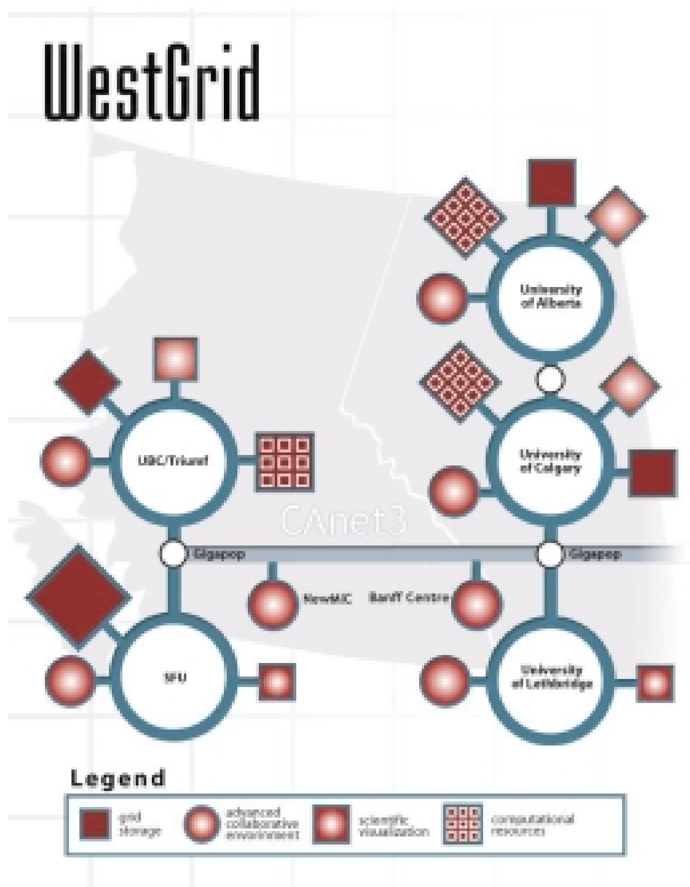
WestGrid

Why is this shared grid of resources needed?
The need for high-performance computing in all disciplines is pervasive and growing. Computational problems once considered intractable are yielding to multiple processors, large memories, large data files, and sophisticated visualization of results. No single university can meet all computing needs of its diverse research communities. WestGrid is creating a shared, geographically dispersed resource using the latest technologies and tools addresses the high-end computing needs of seven institutions in Alberta and British Columbia. The result will meet the needs of a large multidisciplinary research community with minimized duplication and costs, and enhanced access.

Meeting multiple computing needs
WestGrid proposes to install three major computing platforms, each effective for different classes or scales of research problems, and representing different price/performance tradeoffs. The high-end machine has capabilities beyond that of a desk-top machine; the low-end platform represent an inexpensive route to address the computing capacity needs of the progressive research communities in both provinces.

Major components of WestGrid

1 A large cluster of commodity-based computers to be housed at the University of British Columbia and TRIUMF (low-cost computers with a low-speed interconnect; for “coarse-grained” concurrency).
2 A clump of small shared-memory machines at the University of Calgary (a tightly-coupled cluster of multi-processors; “medium-grained” concurrency).
3 A large shared-memory machine at the University of Alberta (high-speed processor intercon-nect; offering “fine-grained” concurrency).
4 A large data management system and graphics server at Simon Fraser University.
5 Collaborative visualization resources at all institutions (used both for scientific visualization and for research into new collaborative technologies).
6 The Grid is the glue that binds this all together. Grid software, careful resource management policies, and a commitment to sharing by all participants enables relatively seamless access to all WestGrid resources.



What is Grid Computing?

Computing grids are geographically separated computers or computer clusters that share applications, data and computational resources. The term grid comes from electricity utility companies, which use a grid architecture in their power distribution systems. The following news clippings describe grid computing for lay audiences.

COMPUTER NETWORKS COMPUTING POWER ON TAP

The Economist, 21 June 2001

IMAGINE that every time you plugged in a toaster, you had to decide which power station should supply the electricity. Worse still, you could select only from those power stations that were built by the company that made the toaster. If the power station chosen happened to be running at full capacity, no toast. Replace the toaster with a personal computer and electrical power with processing power, and this gives a measure of the frustration facing those who dream of distributing large computing problems to dozens, hundreds or even millions of computers via the Internet. A growing band of computer engineers and scientists want to take the toaster analogy to its logical conclusion with a proposal they call the Grid. Although much of it is still theoretical, the Grid is, in effect, a set of software tools which, when combined with clever hardware, would let users tap processing power off the Internet as easily as electrical power can be drawn from the electricity grid. Many scientific problems that require truly massive amounts of computation – designing drugs from their protein blueprints, forecasting local weather patterns months ahead, simulating the airflow around an aircraft – could benefit hugely from the Grid. And as the Grid bandwagon gathers speed, the commercial pay-off could be handsome.

BREAKING NEWS

Globe and Mail, 5 August 2001

International Business Machines Corp. of Armonk, N.Y., has thrown its research muscle behind grid computing, a new way of using the Internet to create virtual supercomputers capable of cheaply solving complex mathematical problems such as weather forecasting or semiconductor chip design.

Canadian companies and researchers are also playing a lead role in developing grid computing systems and applications.

Computing grids are geographically separated computers or computer clusters that share applications, data and computational resources. The term grid comes from electricity utility companies, which use a grid architecture in their power distribution systems.

IBM is not alone in devoting research resources to grid computing. Hewlett-Packard Co., Compaq Computer Corp., Intel Corp., and other computing industry heavyweights are also devoting an increasing amount of their research resources to grid computing. CANARIE, the CA*net3 Internet network and the National Research Council have formed Grid Canada to pursue grid computing projects.

> Canada does not have supercomputer centres like in the United States, [CANARIE's Director of Networks] Mr. St. Arnaud said, so its grid computing potential lies primarily with the many small computer users country-wide. "Since the Americans really are focused on putting their supercomputers together, we think a bigger opportunity is how we connect all our computing resources in communities, schools, and homes," he added.
>
> Grid computing would have immediate commercial applications in a variety of industries, including aircraft engine design, crash test simulation, computer animation, calculation of fluid dynamics for the oil and gas industry and semiconductor chip design.
>
> Many life sciences research projects are also trying to exploit the potential of grid computing. It is now possible, for example, for people to help search for a cure for AIDS by participating in a grid computing project known as FightAIDSatHome. From the organization's web site (www.fightaidsathome.org), participants download software that uses the processing power of the participant's computer to perform molecular modelling.

IN YOUR FUTURE: COMPUTING POWER ON DEMAND
E-Commerce Times, 2 August 2001

> Grid technology is being pioneered by such computing leaders as NASA, the Los Alamos National Laboratory, and the San Diego Supercomputing Center. Among corporations, IBM is working in the Netherlands to connect five universities with a computing grid system. The company uses grid computing to link its research centers. IBM's development is being done through the open-source model, which scientists hope will ensure seamless interoperability. In the future, especially with widespread implementation of Internet2 technology, computer users will be able to order computing power from the grid. The combination of the super-high-bandwidth Internet2 and readily available supercomputing power promises to deliver applications like streamed high-definition video, which could be used in telemedicine, research collaboration, and distance learning.

How is "grid computing" different from WestGrid?

"Grid computing" is taking on new meaning in two ways. In a strictly scientific sense, it refers to a specialization in computing science that is exploring ways to use distributed computers to process large volumes of information. For example, a personal computer on a network can be involved in a massive calculation that is drawing on the CPU power of thousands of PCs.

The term WestGRid describes the interconnected set of diverse information resources accessible over the research networks of BC, Alberta and Canada. While grid computing may be one activity taking place over the network, it is not entirely defined by this. For example, the WestGrid connects high performance computers, collaboration facilities, and data storage facilites.

Economic and Social Impact

The most important economic product of research programs is a stream of trained researchers who carry their expertise with them into a wide variety of fields, often quite different from their original discipline.

Most of Alberta and British Columbia's major economic sectors – including energy, agriculture, forestry, manufacturing, ICT, and distribution – use high-performance computing to enhance their businesses.

WestGrid further supports graduate students, postdoctoral fellows, corporate partners, and others. It is estimated that each new graduate adds $250,000 per year to the economy, aggregating to growth of $75,000,000 annually. This in turn increases the demand for new graduates, new research and new products.

Economic Impact: Global Competitiveness
In a 2000 survey, Canada ranked last in the G8 on a per capita basis for top computing facilities. High performance computing is required to advance the development of new science and engineering that leads to the development of new products, technologies and processes necessary to improve Canada's international ranking.

Economic Impact: Commercialization of Technologies
WestGrid has built an applied research bridge to the corporate world. Research findings in many areas can be tested, applied, and commercialized through corporate partnerships. These range from new products in industries ranging from pharmaceutical to engineering tools to entertainment.

Social Impact: Environmental
Advanced computing supports modelling and simulation research in areas ranging from particles and atoms, to chemicals, geophysical systems, environment, and astrophysics. These significant research developments lead to better, cleaner, and more efficient use and management of Canada's resources.

Social Impact: Health and Quality of Life
Health informatics requires design and testing of distributed, secure, reliable, massive data systems. MAGPIE, the Multipurpose Automated Genome Investigation Environment, is now used to study more than 20 genomes worldwide. Genome Canada will increase the demand for the computational resources for genomic analysis and annotation tools dramatically.

Social Impact: Education and Training
For Canada's health and learning sectors, the continued research and development of new collaborative technologies means continued excellence in tele-health and tele-learning for both public and private enterprises.

Examples of Use

To the Heart of Matter

The goal of subatomic physics is to understand the structure of matter and the fundamental forces of nature at the most fundamental level. Experiments involve the collection of enormous amounts of data, and large-scale computing facilities are required for their analysis. As well as being used to analyze current subatomic physics experiments, the WestGrid facilities will provide a test bed to determine the optimal computing configuration for the analysis of data from the ATLAS experiment at the CERN accelerator in Switzerland. ATLAS is a large international collaborative project to study among other things the mechanism that generates the mass of elementary particles. Twelve senior physicists from TRIUMF and the Universities of Alberta, British Columbia, and Simon Fraser are involved, including Mike Vetterli and other researchers at TRIUMF, and Jim Pinfold at the University of Alberta.

Where No One Has Gone Before

Numerical relativity is concerned with the computational solution of the field equations of general relativity. Among the most spectacular predictions of this theory are the existence of black holes (regions of space with extraordinarily strong gravitational fields), and gravitational waves (ripples in the curvature of space-time). While new experiments are seeking evidence for gravitational waves, physicists Matthew Choptuik at UBC and David Hobill at U of C are performing large-scale numerical simulations to identify the signals that should appear in both terrestrial and space-based detectors from gravitational waves produced from black-hole/black-hole or neutron-star/black-hole interactions. This is an enormous computational task. Another group is studying the details of cosmic microwave background fluctuations, and the large-scale clustering of galaxies, testing important test of models of the universe's origin.

Space Weather: When it Rains, It Pours Particles

Space physics theory and modelling by professors John Samson, Robert Rankin, and Richard Marchand at the University of Alberta is focused on understanding the dynamics of the interaction between the solar wind and the Earth's magnetosphere. These studies are of critical importance to understanding, and eventually forecasting, near-Earth space environment conditions. The economic impact of these studies is considerable, since disruptive events in the magnetosphere can have adverse consequences on ground-based infrastructures such as large power grids and pipelines, as well as on satellites and global communications.

One Billionth of a Meter Matters

Research in condensed matter physics is crucial to the development of advanced materials used in the high-tech industry (nanotechnology). Work by Jeff Young at UBC on a new class of nanostructured optical materials, referred to as "photonic crystals," could lead to important applications in the optical communications industry. This research involves the development of linear and non-linear electromagnetic simulation tools, and uses high-performance computing to accelerate progress in the laboratory.

Giving a (Chemical) Friend a Helping Hand

Tom Ziegler's ambitious goal at the University of Calgary is the complete ab initio design of catalysts for industrial application. Chemical catalysis is the art of inducing desirable reactions by "gentle" means. The microscopic details of catalytic action are often unknown because such processes are by virtue fast, and hence difficult to follow experimentally. Thus, accurate detailed theoretical calculations are crucial in this area.

Replacing the Test Tube with a Computer

Grenfell Patey's research at the University of British Columbia is concerned with the microscopic theory of liquids and solutions, with the primary aim to develop realistic models and accurate theories for physically important systems. Such systems often involve long-range electrostatic forces, and Patey has contributed significantly to the present, hard-won, qualitative understanding of their equilibrium behaviour. There are several young scientists working in this area, and WestGrid will have a dramatic impact on their research and on their ability to compete internationally.

Lessons from the Ice Age

Concern about climate stability, and the climate of the near future, has drawn attention to the dramatic instability of Earth's climate during the most recent Ice Age (10,000 years ago). Gary Clarke's ongoing research at UBC involves developing computer models to simulate the growth and decay of continental-scale ice sheets over one or more glacial cycles (100,000 years per cycle). Ideally these models would be global in scale and fully coupled to dynamic models of the atmosphere and ocean. Such a coupled global model is beyond the capabilities of computers currently available to Clarke's group. Similarly, Kendal Bushe at UBC has built a coupled glacial-ocean-atmosphere computer model. This research will be used to better understand the processes responsible for the extremely rapid climate changes that have occurred over the last 100,000 years, and to improve our predictions of future climate change. The more realistic the model, the more computations are required.

Exploring the Abyss

Gordon Swaters at the University of Alberta is focused on understanding the dynamics of large-scale deep, or abyssal, circulation in the ocean. Abyssal currents are the principal means by which cold, dense water, produced in high latitudes due to cooling by the atmosphere, flows to the equator. Due to the enormous technical difficulties of obtaining data from a region some five kilometres below the surface, we have limited knowledge of these currents. Nonetheless, the dynamics of these flows has a significant impact on the convective overturning of the ocean and hence the global climate system. Extensive access to a large computer is critical to the success of this research as the researchers are modelling three time-dependent atmospheric.

WestGrid also Includes the Yukon!

Gary Clarke at UBC, David Hik at the U of A, and Glen Marshall at TRIUMF will use the proposed WestGrid commodity cluster in a new approach to the study of climate change impacts on land surface processes and ecosystem dynamics in the south-west Yukon. Their aim is to develop coupled spatio-temporal models of landscape evolution and ecosystem dynamics and to drive these models with a regional climate model.

Decoding the Secrets of Life

The complete sequencing of the human genome is almost finished, posing great challenges for genome analysis and annotation. The sheer size of the human genome (which is about 1,000 times larger than bacterial genomes) is an indicator of the task at hand. Christoph Sensen at the University of Calgary and collaborators have built completely automated systems that can analyze microbial genomes within a few weeks. These systems are considered leading edge in terms of accuracy, performance and wealth of information generated, and they are currently in use by over 20 genome projects.

Would You Give Your Credit Card Number Over the Internet?
The research groups of Jonathan Borwein, Peter Borwein and David Boyd in Vancouver, and Hugh Williams in Calgary bring deep algorithmic number theoretic expertise to WestGrid. Related research on cryptography requires solving difficult mathematical problems. Hugh Williams has developed one of the most widely used public key cryptography systems.

Chaotic Systems: Pushing the Frontiers of Classical and Quantum Dynamics
At the University of Lethbridge, Dr Keramat Ali is working on developing new approaches in dynamics, both classical and quantum, to broaden the efficiency and scope of artificial recurrent neural networks (RNNs). The need to study chaotic neural networks is crucial, as these networks are unavoidable and necessary in real biological systems. They present a significant challenge in physics though, because there is no persistence in dynamics. A breakthrough in the area of control and synchronization of chaos in recurrent neural networks could potentially transform the development of artificial neural networks and improve the efficiency of neural network systemization in robotics applications. Dr Ali's research is pushing the frontiers of classical and quantum dynamics, and requires access to high performance computational resources to continue with research breakthroughs.

Reducing Automotive Pollution
Robert Hayes' group at the University of Alberta models the automotive three-way catalytic converter to model accurately the converter behaviour during transient operation, which is necessary for understanding the crucial warm-up period. This period accounts for 80% of typical vehicle emissions, and this drives the design of the converter. Because of the complexity of the underlying model, there are large number of unknowns to calculate (over 30), which, when coupled to the requirement of a fairly fine discretization, leads to a very large computational problem.

What is e-science?
A new concept, known as e-science, takes research and development beyond their traditional "lab" environments and, through the use of advanced networks and applications, into classrooms and small business environments across the provinces. E-science refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualization tools that can show the results back to the individual scientists.

What You See, Hear and Feel is What You Get
With its hundreds of researchers and graduate students, WestGrid provides a unique and fertile test-bed for exploring digitally mediated collaborations between institutions, disciplines and individual researchers. Collaboration isn't just about hardware and software. Significant research still remains to be done in extending current tools, such as video-conferencing, into truly powerful technologies for meaningful collaboration. Jonathan Borwein in Vancouver, Pierre Boulanger in Edmonton, and Saul Greenberg in Calgary will collaborate to address fundamental questions such as: What are the technical and social factors that create high-quality collaboration between distributed users? What can we learn from social science and cultural research in creating appropriate interfaces for collaborative tools? What is the best way to promote casual, serendipitous collaboration, as well as formal, planned activities?

Visualization is becoming increasingly important as our information-dominated society shifts away from primarily verbal modes of communication. Pierre Boulanger at the U of A, Sheila Carpendale at U of C, Kelly Booth at the New Media Innovation Centre, and Sara Diamond at The Banff Centre will investigate ways to better understand the essential aspects of information visualization. Both the Banff Centre and NewMIC will act as centres for applied research.

Oil Sands

Pierre Boulanger at the University of Alberta has a joint research project with Syncrude to built an immersive environment that remotely monitors and controls mining in the Alberta tar sands. The idea is to build a virtual control room located in Edmonton, allowing for the management and control of equipment at the mining site several hundred kilometres away.

Simulating the Heart

Joshua Leon at the University of Calgary is researching atrial fibrillation (AF), a condition which affects up to 10% of the population over 70 years of age. Although AF is not in and of itself life threatening, it has extremely serious sequelae, including stroke and heart failure. Leon's research group uses realistic models of the electrical activity of the heart to examine the dynamics of AF. The models allow them to test hypotheses about how AF might be managed pharmaceutically. They can easily introduce the effects of a particular drug into the simulations and examine whether or not it has prophylactic effects. This approach is becoming a very attractive alternative for the early phases of drug design, significantly reducing the time to test a hypothesis and the amount of extremely costly animal testing. The availability of the computation facilities allows the group to scale up their modelling to include all known ion channel dynamics and to create models of a realistic size. Because of these computational resources, in silico drug development is a now a reality, and is seen as the next stage of research.

Related Projects

SFU CoLab ***(www.colab.sfu.ca)***

CoLab is collaborating with the New Media Innovation Centre (NewMIC) and has access to NewMIC's emersive reality centre. For more information, visit *(www.newmic.com/)*.

HPC@SFU ***(www.hpc.sfu.ca)***

The project consists of three computing architectures:

1 A pre-existsing 8 processor Silicon Graphics Origin 2000. This system is setup and running.
2 A 32 processor, 8 x 4way (ES40), Alpha system that has a Compaq proprietary "Memory Channel" interconnect. This system in running in a beta test mode. Some users are using the system for real research.
3 A 192 processor, 96 x dual (AMD Athlon MP), system that uses bonded 100Mbps Ethernet (3 channels = 300Mbps, fdx).
This system is post-alpha, pre-beta. It's configured and has had a benchmark test done with 160 processors. A few heat related and memory related problems are being ironed out.

After all three systems are fully functional we'll move toward integrating them into a single infrastructure where all will share a common home directory and login system. Users will simply select which architecture meets their needs best for any given problem and then use that system.