The Power of Scalable Cluster Computing

By Paul Redfern

Cornell University has deployed a 256-processor Windows cluster as the high-performance computing server for the Advanced Cluster Computing Consortium (AC³). AC³ helps corporate enterprises, higher education, and government agencies plan, implement, and maximize the performance of industry-standard software, systems, and tools. This article describes how high-speed switches, Intel architecture, and Windows are redefining servers—from small clusters to supercomputers.

The Cornell Theory Center (CTC) differs from other universities that serve research initiatives only. It has deployed a 256-processor Windows NT cluster to serve as a production high-performance computing resource for corporations, academic institutions, and government agencies. The goal of this working environment is to provide affordable supercomputing of the future and make it easy to launch and control large-scale initiatives on commodity-based solutions.

The Advanced Cluster Computing Consortium

Dell Computer, Intel, and Microsoft recently joined the CTC to establish the Advanced Cluster Computing Consortium, or AC³. The AC³ is an IT service and research consortium to help businesses, educational organizations, and government agencies make the transition to industry-standard clusters and maximize their performance.

The AC³ provides consulting expertise and member access to leading-edge systems. It has grown to 30 members, including Air Products and Chemicals, Argonne National Laboratory, City University of New York, Etnus, Fluent, IBM, Indiana University, JP Morgan, LEXIS-NEXIS™, Reliable Network Solutions, and SAS.

The AC³ also provides UNIX®-to-Windows NT training and IT seminars such as the June 2 "Roadmaps to the Future of Cluster Computing," featuring lab leaders from Dell, Intel, Microsoft, and Cornell. Other services include code porting and parallelization services, benchmark testing, research projects such as future I/O, joint tool development projects, technical briefings, and other custom studies and services.

AC³ Launches Most Powerful Windows Cluster

In August 1999, AC³ installed the world's most powerful Windows cluster, a 256-processor Dell PowerEdge system with Pentium® processors. This was the first cluster of its kind within an academic environment that moved from traditional UNIX to a Windows NT environment.

This cluster included 64 Dell PowerEdge servers, each with a quad Intel Pentium III Xeon™ 500 MHz processor running Microsoft Windows NT, Dell PowerVault storage systems, and Giganet cLAN host adapters and switches for high-speed communications. This entire AC³ Velocity system was up and running in less than 10 hours—and that was just the beginning.

The consortium is now planning its next system, a second 256-processor cluster from Dell—Velocity2—that

will feature Intel's new 64-bit Itanium™ processors.

Why High-Performance Computing?

Many applications today require high-performance computing (HPC) as the compute environment becomes more complex. HPC applications can be both numeric and data intensive. Some require tracking millions of particles over an equally large number of time steps. For example, astrophysicists use HPC to predict the future of the galaxies; biologists focus on understanding structures such as proteins. The U.S. National Foundation for Cancer Research has reported the use of high-performance computers to predict a complex protein structure that is expected to help design new anti-cancer drugs.

Cornell Theory Center deployed a 256-processor Windows cluster to serve as a production high-performance computing resource for corporations, academic institutions, and government agencies.

HPC is used extensively for commercial applications. Banks and insurance companies conduct day-to-day transactions on clusters, then use that data to model the behavior of financial markets. Merrill Lynch® recently installed a 64-processor Dell PowerEdge cluster with Giganet interconnect for its next-generation Broker Information Management System. This system includes Web servers for client access, transaction servers for high-volume query processing, and parallel database servers. Figure 1 shows how clusters are ideal for computing complex data such as the implied volatility of the S&P 500.

HPC Application Requirements

Some cluster applications require high-bandwidth and low-latency communications among the processors doing the work; other applications can be divided among the processors and left to run on their own. One job may require a few minutes to run on 64 processors, while another might run for days across hundreds of processors.

All HPC applications need a large systems infrastructure. This could mean dozens to even thousands of processors, very large cache, high-bandwidth communications, massive storage systems for databases, and more.

For example, the AC³ Velocity cluster is 64 Dell PowerEdge 6350 servers. Each symmetric multiprocessing (SMP) node has four Pentium III processors with 2 MB of cache, 4 GB of RAM, and 50 GB of disk space. Velocity has a full Giganet (www.giganet.com) interconnect with very low latency and 100 MB/sec bandwidth.

Large HPC clusters usually indicate the need for large storage systems, often called mass storage systems. Velocity nodes are connected to a 4 terabyte (TB) Dell PowerVault 200S storage system and a 2 TB Dell PowerVault 130T Tape Library. See Figure 2.

Scaling Up to the Challenge

Midrange to high-end cluster computing systems with industry-standard components provide the scalable resources needed for today's technical and business computing requirements. The need for scalability of applications from the desktop to clusters on demand is both appealing and growing. Exceptional economies of scale can be leveraged when high-performance clusters use the same software and systems as the desktop. Preliminary engineering designs or scientific simulations can be seamlessly scaled up to dozens, or even hundreds, of processors.

One example is to effectively predict when an aircraft engine piston might crack. This requires multiscale modeling, or computer simulations, that range from microns to meters. The engineering software and computational methods required to solve this problem come from diverse disciplines such as molecular dynamics and materials science. This in turn creates a particularly complex problem set.

High-End Clusters Enable Real-Time Decision Making

The rapid solution of large-scale data problems to enable real-time or near real-time decision making is becoming a key competitive factor. HPC clusters built with industry-standard components will become more common not only

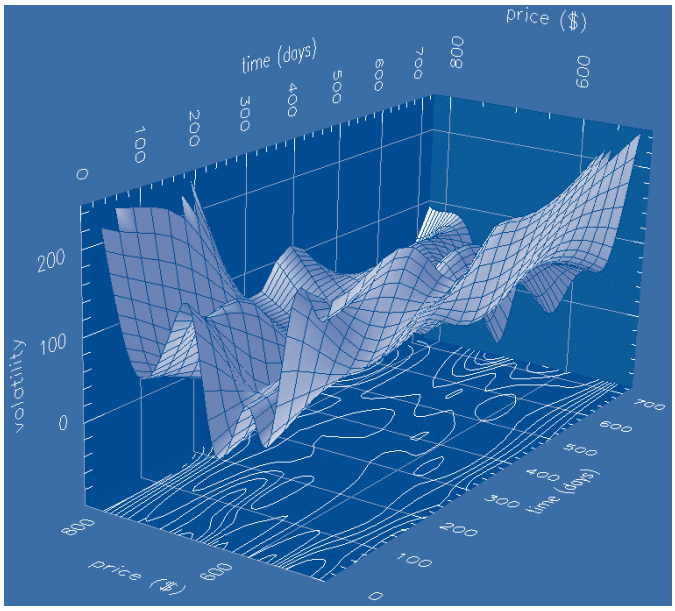

Figure 1. Computing Complex Data for the S&P 500

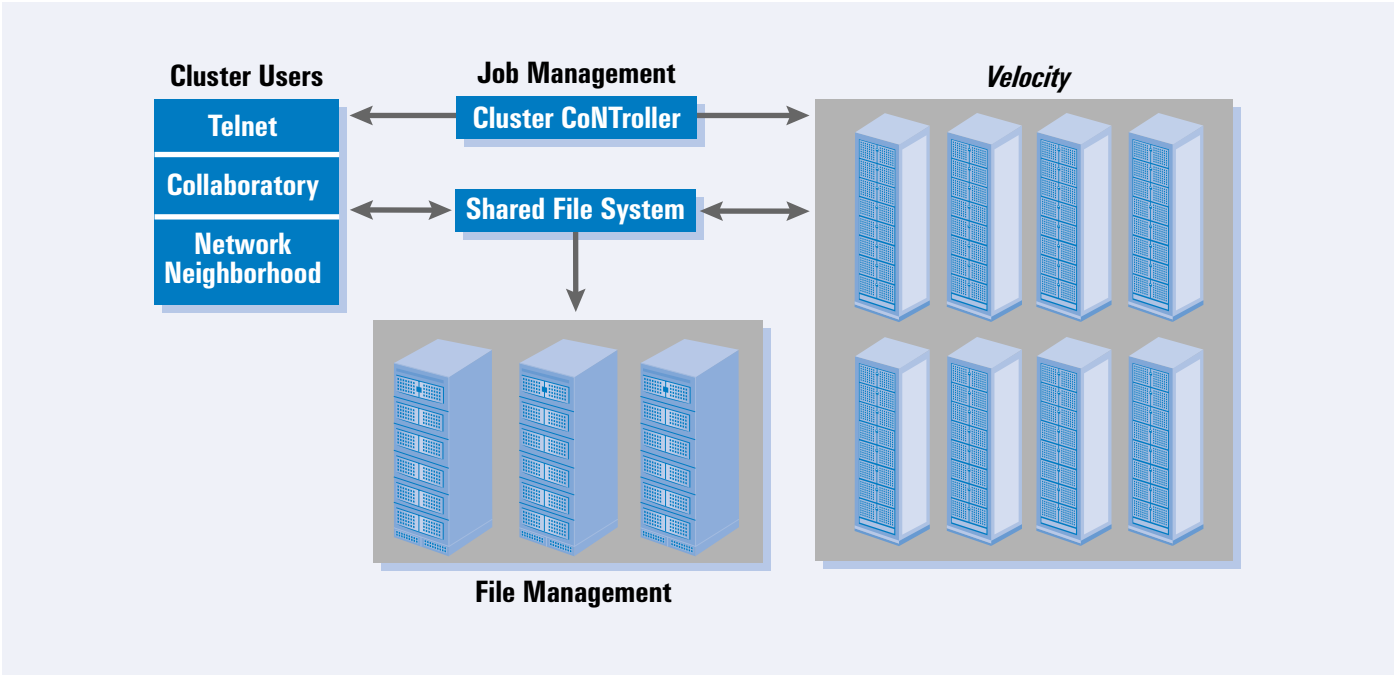

Figure 2. Users Access Velocity from Their Desktops

because they are affordable, but because business and industry want to move closer to real-time computation in a "double click" environment.

The AC³ has particular expertise in developing and testing parallel, scalable applications. CTC has provided parallel computing training, consulting, and advanced systems access to Ford, Schlumberger, SAIC®, Corning, Eastman Kodak®, and over 100 universities and government agencies worldwide.

Most recently, CTC was contracted to operate a 48-processor Dell PowerEdge Windows cluster with Oracle8i for the U.S. Department of Agriculture (USDA) National Bioinformatics and Comparative Genomics Center. CTC also has been selected to operate a 32-processor Dell PowerEdge cluster running SAS for a major social science and economics research center.

Exceptional economies of scale can be leveraged when high-performance clusters use the same software and systems as the desktop.

Parallel Processing Relies on Passing Messages

Processors running parallel programs call for data and instructions, then perform calculations. Each processor checks back periodically with other nodes or a master node to plan its next move or to synchronize the delivery of results. These activities rely on message-passing software, such as industry-standard MPI.

MPI Software Technology, Inc. (www.mpi-softtech.com) has developed a commercial implementation for Windows NT of their software, MPI/Pro™, for both clients and servers. This software has been enhanced to accommodate the needs of shared, massively parallel systems.

Next Step in Cluster Evolution

The AC³ is currently working on the practical realities of integrating clusters of clusters. AC³ consultants are integrating small, application-specific clusters with Velocity to enable users to scale up on demand. This is the forefront of distributed cluster research, and this computing model has captured the interest of many commercial customers.

Other AC³ research also is pushing the envelope, including reliability, I/O performance, resource management, and database testing. CTC and the AC³ also develop new tools such as Cluster CoNTroller™, a serial and parallel job scheduler now available from MPI Software Technology for Windows NT clusters.

To discuss AC³ memberships, services, and customer briefings, call Paul Redfern at 607-254-8693 or send e-mail to red@tc.cornell.edu.◆

***Paul Redfern** is associate director for the Cornell Theory Center (www.tc.cornell.edu).*