Cornell University & partners demonstrate seamless access to a software utility with the National Science Foundation-funded “MATLAB on the TeraGrid” project

Benefits
- Enable a broader class of scientific researchers and students who are well versed in MATLAB to reduce the time to solution in a seamless and scalable manner
- Provide transparent access to an HPC software utility from a single interface that hides the complexity of batch submissions from multiple client operating systems
- Deploy IT faster and at a lower cost with Dell blade servers that are easy to deploy and maintain by in-house staff
- Run applications in parallel through Science Gateways that mask the intricacies of parallel computing
- 75% time savings on system procurement through Dell

“This experimental resource is available for remote interactive use by academic researchers and for integration with Science Gateways that serve entire communities.”

David Lifka, director, Cornell Center for Advanced Computing

Customer Profile

Company: Cornell University Center for Advanced Computing
Industry: Education, Scientific Research, Engineering
Country: United States
Employees: 14
Web: www.cac.cornell.edu

Business Need
The Cornell Center for Advanced Computing (CAC) was interested in understanding how best to deploy a software utility as a transparent and responsive user service for scientific researchers and engineers. The goal was to demonstrate a working model for high-performance utility computing with MATLAB®, which might encourage other software vendors to explore and develop similar capabilities.

Solution
The center received an NSF grant to deploy the high-level technical computing language MATLAB to a cluster called “MATLAB on the TeraGrid.” This experimental TeraGrid computing resource was deployed using Dell® PowerEdge™ servers with Intel® Xeon® processors running Microsoft® Windows® HPC Server 2008 along with DataDirect Networks storage and Force10 Networks switches sourced through Dell. MATLAB Distributed Computing Server™ was also installed on the servers.
External researchers remotely accessed all 512 cores of the MATLAB cluster to model a hepatitis C virus, a major cause of liver disease worldwide.

Cornell researchers from earth and atmospheric sciences, civil and environmental engineering and biomedical engineering, and researchers across the nation are preparing MATLAB programs to run on the parallel resource.

What do these research groups have in common?

They are eager to benefit from seamless access to a high-performance computing software utility. Users will access the utility from any desktop operating system—Microsoft Windows, Mac OS X or Linux—using job submission integration code developed by CAC. After programming and prototyping applications on their desktop with MATLAB and The MathWorks Parallel Computing Toolbox™, they will scale up to a Dell PowerEdge cluster using MATLAB Distributed Computing Server.

Why “MATLAB On The TeraGrid”

Thanks to a grant from the National Science Foundation, MATLAB, a popular numerical computing environment, is now available as a high-performance software utility. This experimental TeraGrid resource was deployed in order to investigate the value of software utilities as HPC resources and as computational backends to Science Gateways. Science Gateways provide community-developed tools, applications and data collections for specific research domains through a single portal. This software utility will also serve as a resource for educators who are interested in introducing students to the basic principles of parallel computing. “We can start using the software utility as a tool to teach students about parallel computing basics and message passing constructs,” says Susan Mehringer, assistant director, consulting at CAC.

Opening The Door To More Researchers

Over one million scientists and engineers use MATLAB to develop technical computing applications. It is a pervasive tool in research, essential to scientific discovery in a wide range of disciplines including nanoscience, restoration ecology, microbiology and computational fluid dynamics. MATLAB is also an important tool for manipulating data from scientific experiments.

“MATLAB on the TeraGrid was deployed as a 512-core high performance software utility on Dell PowerEdge blade servers.”

Resa Alvord, assistant director, systems and operations, Cornell Center for Advanced Computing

### Technology at Work

<table>
<thead>
<tr>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell™ Warranty Parts Direct Program</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataDirect Networks S2A9700 storage</td>
</tr>
<tr>
<td>Dell PowerEdge™ M600 blade servers with Intel® Xeon® 5420 processors</td>
</tr>
<tr>
<td>Dell PowerEdge M1000e modular blade enclosures</td>
</tr>
<tr>
<td>Force10 Networks C300 switches</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft® Windows® HPC Server 2008</td>
</tr>
<tr>
<td>The MathWorks MATLAB Distributed Computing Server</td>
</tr>
</tbody>
</table>
instruments, satellites and sensors in novel ways in order to gain new insights into basic research questions or to solve problems confronting society today. Investigating the value of parallel HPC software utilities such as MATLAB on the TeraGrid is important in order to understand their potential in broadening the high performance computing user community. Ease of use, transparent access via Science Gateways, and the effectiveness of HPC utilities as educational tools are among the project’s metrics of success.

Partnering With Dell
CAC selected a cluster of 64 Dell PowerEdge M600 blade servers for the system configuration with Intel Xeon processors for energy efficient performance. The blades are housed in Dell PowerEdge M1000e modular blade enclosures. DataDirect Networks S2A9700 storage and Force10 Networks C300 switches—both sourced from Dell—provide scalable storage and reliable connectivity. The cluster runs Microsoft Windows HPC Server 2008.

CAC participates in the Dell Warranty Parts Direct Program. “Working with Dell enables us to be certified on the equipment that we use,” says Resa Alvord, assistant director, systems and operations, CAC. “We can support ourselves and get parts from Dell overnight at a price that makes sense.”

“Dell blades provide the performance and reliability a HPC software utility needs and reduce infrastructure costs,” adds Paul Redfern, assistant director. “Dell products integrate leading edge components and have the added value of ongoing maintenance and support that doesn’t break the bank.”

Easy Server Set Up
The Cornell Center for Advanced Computing has been purchasing Dell servers and desktops for more than a decade, and sourcing the investigation of an HPC software utility through Dell was a natural evolution of the Dell/Cornell partnership. “It often makes sense for us to go through Dell for all of the components of an HPC system,” says David Lifka, CAC director. “Dell has strong partnerships with leading technology providers, and we can have a conversation with our Dell sales engineer and readily understand the value propositions of different components in an overall system configuration. We can then order the complete configuration through Dell which is a lot easier and, in this case, saved 75 percent on procurement time. Dell has a positive working relationship with vendors such as Force10 and DataDirect Networks and is able to work seamlessly with those vendors to deliver a complete solution.”

“MATLAB on the TeraGrid was deployed as a 512-core high performance software utility,” explains Alvord. The Dell PowerEdge M600 blade servers in the Dell PowerEdge M1000e modular blade enclosures were easy and fast to set up. “Installing the enclosures was one of the easiest installations yet,” Alvord says. CAC operates 15 Dell PowerEdge HPC clusters. Installations typically take a day or less if no special software is required.

Supporting Science Gateways
This parallel MATLAB utility will also serve as a computational engine for Science Gateways such as nanohub.org, a Web-based resource for research, education and collaboration created by the NSF-

“We can start using the software utility as a tool to teach students about parallel computing basics and message passing constructs.”

Susan Mehringer, assistant director, consulting, Cornell Center for Advanced Computing
funded Network for Computational Nanotechnology and operated by Purdue University, a partner in Cornell’s MATLAB project. Science Gateway users will access discipline-specific Web portals and through simulation inputs via a Web form, launch MATLAB simulations and get timely results without knowledge of the underlying software or computing infrastructure. This will allow researchers to leverage a computational research tool without having to overcome a platform-specific learning curve.

“This experimental resource is available for remote interactive use by academic researchers and for integration with Science Gateways that serve entire communities. We’re confident that scientific users and engineers from across the nation will benefit from the reliability and performance of the Dell blade solution while, on the IT side, Cornell will benefit from reduced power and space requirements and the ability to maintain the blades with in-house staff,” concludes Lifka.

“Additional Science Gateways using the MATLAB resource may follow,” Lifka says. CAC is working with medical college researchers to develop gateways for epidemiology and medical image analysis.

For more information go to: www.DellHPCSolutions.com