



CISER

CISER: Cornell Researchers Team Itanium® 2-Based Unisys ES7000* Server with SAS* 9 and Windows* Server 2003 to Boost Performance and Streamline Data Center

SOLUTION SUMMARY

Challenge The Cornell Institute for Social and Economic Research (CISER) urgently needed to boost performance for processing- and memory-intensive SAS* applications. CISER needed to acquire the resources to run 64-bit SAS 9 when it became available for the Windows* Server 2003 operating system. Yet, more than 600 researchers still needed broad-ranging 32-bit support. CISER needed to find an economical way to manage both 64-bit and 32-bit user communities demanding more power.

Solution Rather than continue to add 2- and 4-way servers, CISER decided to purchase a Unisys ES7000* Enterprise Server containing eight Intel® Itanium® 2 processors and 16 Intel Xeon™ processors MP. This mainframe-class machine provides tremendous processor and memory scalability in the familiar Windows environment. CISER will be able to host 32-bit and 64-bit applications in the same system while consolidating its server holdings from 12 to 1.

Business value By moving to the Unisys ES7000, CISER will boost the performance of some particularly demanding SAS applications by as much as 70 percent. The additional processing power allows researchers to crunch through demanding data analyses much faster than before and to work with much larger data sets to answer questions they couldn't even ask before. Faster answers lead to more work accomplished in a given grant period and an improved chance of renewing grant funding. With the ability to host both 32-bit and 64-bit applications in the same environment, CISER users will have access to the perfect mix of processing power for every job. CISER has found Itanium 2 processors to be a better fit for complex jobs and more users, while the Intel Xeon processor MP provides superior performance for less demanding jobs and/or fewer users. By consolidating on the one highly scalable Unisys ES7000, CISER will spend between 30 and 50 percent less time installing and maintaining software.

Database and application server Unisys ES7000 with 8 Intel Itanium 2 processors and 16 Intel Xeon processors MP

Operating system Microsoft Windows Server 2003, Datacenter Edition

Applications SAS 9, SPSS, Stata and other statistical applications

Storage 5-TB EMC SAN

Business Challenge

RESEARCHERS CRAVE SPEED

Until recently, social scientists at the Cornell Institute for Social and Economic Research (CISER) had to be as creative in figuring out how to feed data into the institute's computers as in designing and analyzing their research. Some data sets were so huge – up to 15 GB – that researchers had to carve up code into multiple jobs and string them across multiple machines, because no single computer could carry the load. Even when configured as multiple mini-jobs, processing-intensive analyses would take days to run.

“Slow computers really gum up the research process,” explains Carol Murphree, CISER Computing Consulting Manager. “They hold up individual researchers, whole projects, and can even endanger a grant’s renewal. The longer research takes, the sooner grant money runs out and the fewer results researchers have to show for their efforts.”

“As the demands for particular resources increase, so do the advantages of using the Itanium architecture. Additional processors also boost overall performance for these high-end applications.”

Carol Murphree
Computing Consulting Manager,
CISER

“The ability to consolidate and manage many application types on one system is a huge benefit, as is the ability to support both 32-bit and 64-bit applications in the same environment. This should simplify my life a great deal.”

Janet Heslop

Computing Systems Manager,
CISER

Computer run-times at CISER have recently picked up considerably thanks to reinforcement by Itanium 2-based servers. After extensive benchmarking on two loaner systems from Intel, CISER has decided to purchase a Unisys ES7000 Enterprise Server containing eight Itanium 2 processors and 16 Intel Xeon processors MP to run demanding SAS 9 programs and other statistical applications.

Says Janet Heslop, computing systems manager for CISER, “Our new Itanium 2 processor-based servers’ capabilities will not only allow researchers to speed through data analysis faster but tackle new categories of problems. Researchers can work with much larger data sets that use more CPU and memory resources and therefore ask questions they couldn’t ask before.”

RIDING THE INTEL PERFORMANCE CURVE

CISER was founded in 1981 to provide research support services for social scientists at Cornell University. CISER is supported jointly by seven colleges and comprises more than 400 faculty and senior researchers who represent a broad spectrum of interests, from basic theory to direct application.

The computational needs of Cornell researchers vary, from individual researchers running relatively modest programs to whole teams running exceedingly complex processor- and memory-intensive statistical programs. One such high-end project has linked Cornell researchers with the U.S. Census Bureau, the Urban Institute and the University of Maryland in a \$4.1 million National Science Foundation (NSF) social data infrastructure grant. From masked, real-world data, the team will glean various relationships between employers, employees and specific jobs.

CISER originally created a cluster of eight 2-way Intel Pentium® III Xeon processor-based servers to meet the computational needs of this particular research team. CISER already had a cluster of eight 4-way Pentium III Xeon processor-based servers meeting the needs of its 600+ other users. The Cornell Theory Center (CTC) standardized on Intel architecture-based systems back in 1997 and encourages other research groups at Cornell to embrace the advantages of the open Intel architecture and Microsoft Windows* operating system.

Many of the social science researchers supported by CISER rely on sophisticated statistical software packages that run under a 32-bit architecture and the Windows 2000 operating system. There are times, however, when 32-bit hardware and software do not provide enough horsepower for working with extremely large data sets and/or performing highly processor-intensive operations. The NSF researchers, for example, were regularly hitting the memory and processor limits of their 32-bit multi-processor servers using SAS 8.2.

“We wanted a system that could support the CPU and memory needs of our very high-end users, as well as carry the load of our general users who prefer Windows-based applications and mostly still rely on 32-bit applications,” Heslop explains.

BUSINESS SOLUTION

ITANIUM® 2 PROCESSORS YIELD IMMEDIATE SPEED-UP

When SAS led the statistical software pack by announcing its intention to release SAS 9 for 64-bit Windows Server 2003, the NSF team was ready with code to be ported to just such a system but without the technology resources to do so.

Initially, CISER staff arranged with Intel and SAS to beta-test a combination of SAS 9, Windows Server 2003 Enterprise Edition and the first generation of Itanium 2 processors. Heslop oversaw the installation of this system behind a firewall at CISER and the machine was turned over to the NSF team researchers, who immediately ported their code and restructured it to take advantage of the increased resources available. Right away, they were able to run multiple simultaneous procedures with code that previously had been run in a series of smaller steps with subsets of the data.

Encouraged by this experience, Murphree and Heslop set out to do more methodical testing to determine what part of this performance improvement could be attributed to the new multi-threaded capabilities of SAS and what could be attributed to the upgrade to 64-bit architecture. They were also interested in whether and at what point, still more 64-bit processors would be beneficial to its most demanding SAS users.

Heslop says, “A related issue of concern for CISER was whether and to what extent, the 64-bit architecture might serve our 600 general users with equally important, but less resource-demanding applications. We had heard that the computational overhead required for a 64-bit system can actually slow down tasks that are less CPU- or memory-intensive. We needed to determine the point at which use of these resources becomes practical, as well as examine the ability of the system to handle multiple users with a broad range of applications.”

ENTER THE UNISYS ES7000 ENTERPRISE SERVER

At this point, the Cornell Theory Center offered CISER the opportunity to test its code on a Unisys ES7000 with 16 Itanium 2 processors. Again, SAS provided a license for the testing and CISER had the opportunity to explore the effect of increasing the number of processors from 4 to 16.

Murphree and Heslop worked with the NSF research team to extract critical pieces of SAS code from their research and modify it to use with carefully simulated data. The test code used the MP/CONNECT feature of SAS to generate from 1 to 20 simultaneous “remote” SAS sessions that sorted and merged up to 20 sets of 3 related files. Each set of 3 files contained synthetic integrated labor and product market information for 16 million jobs. Among other things, the test code takes advantage of some of the new multi-threaded procedures available for the first time with SAS 9.

“Running the tests iteratively while increasing the number of simultaneous sessions provided a proxy for testing the ability of the system to handle multiple users,” Murphree says. “In addition, accumulating detailed performance data for each sub-part of the test code also provided useful information about SAS procedures and steps that are commonly used by many CISER researchers, not just the NSF group. We ran these same tests on our own 32-bit systems with both SAS 8.2 and SAS 9.0 and Windows 2000 Advanced Server to address our concerns about how performance compared across architectures, disk configurations and operating systems. We focused on the ‘real time’ to complete the code as a measure of performance, since this is an important factor for our users.”

Their findings: “As the demands for particular resources increase, so do the advantages of using the Itanium architecture,” Murphree says. “Those demands include the need for multi-threaded processing, multiple simultaneous tasks (by one or more users), memory bus load and data set size. What we learned from our comparison tests on the Theory Center’s 16-processor ES7000 versus our own 4-way nodes was that as these resource demands increase, the additional processors mean better overall performance.”

For CISER’s test code, the 4-way white box was able to process seven or fewer simultaneous sessions the fastest. However, once the number of simultaneous sessions exceeded seven, the 16-way was not only faster but would support a total of 18 sessions (38 percent more than the 4-way).

However, the testing also showed that for less demanding jobs and/or fewer users, the Intel Xeon processor MP delivered superior performance. “Most

of our general social science users are better suited to Intel Xeon processor MP-based systems,” Heslop says. “Their applications don’t take advantage of as much memory or CPU resources. Plus, most of the statistical applications they use are already optimized for Intel Xeon processor MP-based systems.”

EASIER MANAGEABILITY, SCALABILITY

The Unisys ES7000 offers a unique combination of both 32-bit Intel Xeon processors MP and 64-bit Itanium 2 processors in one integrated system running the Windows Server 2003, Datacenter Edition operating system. Unisys employs a mainframe crossbar architecture, rather than the traditional system bus architecture, to deliver huge I/O and memory capacity. The server’s 32 processors can be partitioned into multiple virtual servers, each running independently of the others but sharing system storage and networking resources.

“The biggest attractions of the ES7000 are its easier manageability and its potential for expansion,” Heslop says. “Over time, we’ll be able to phase out our twelve 2-way and 4-way servers and have just one ES7000 system to manage. The ability to consolidate and manage many application types on one system is a huge benefit, as is the ability to support both 32-bit and 64-bit applications in the same environment. This should simplify my life a great deal.”

Heslop will no longer need to deploy applications, operating systems and upgrades across 12 servers but only one. “With funding at the university essentially frozen, it’s very important to us to be able to manage an increasing amount of computing power with the same number of people,” she says.

By consolidating its server holdings down to one generous-sized machine, CISER will not only have more processing power and easier manageability but tremendous scalability within a single-system paradigm. “We can go up to 16 processors on the Itanium 2 processor side of the system that we have purchased,” Heslop says. “That kind of flexibility allows us to purchase a very powerful system within our current budget constraints yet have plenty of growing room as our processing needs expand and as more software vendors

“We can go up to 16 processors on the Itanium 2 processor side of the system that we have purchased. That kind of flexibility allows us to purchase a very powerful system within our current budget constraints yet have plenty of growing room as our processing needs expand and as more software vendors port their code to the 64-bit architecture.”

Janet Heslop
Computing Systems Manager,
CISER

"Our new Itanium® 2 processor-based servers' capabilities will not only allow researchers to speed through data analysis faster but ask questions they couldn't ask before."

Janet Heslop
Computing Systems Manager,
CISER

port their code to the 64-bit architecture." As Intel announces new versions of its processors, the ES7000 can be easily upgraded without disruption and without having to replace the entire system.

The added horsepower also works to attract funding to the university. "Having access to an enterprise server like the ES7000 will help researchers land new projects," Murphree says. "It looks good on your grant

application to say you have access to robust computing resources. Faster computers allow researchers to get more work done in a given research period. Ultimately, it helps Cornell increase its funding levels."

ITANIUM 2 PROCESSORS OPEN UP MASSIVE MEMORY

Offering performance typically associated with RISC processors, the Intel Itanium 2 microarchitecture includes a 64-bit address base, access to greater memory, the ability to work with larger numbers and data sets and higher precision in performance for floating-point operations. The Itanium 2 microarchitecture is designed to meet the computational needs of those working with integer and floating point applications as well as people working with large-scale data sets.

"The ability to get at more than 2 GB of memory is key for us, made possible by the combination of Intel's 64-bit chip, 64-bit Windows 2003 and SAS 9.1," Heslop says. "SAS takes advantage of large memory addressability

in its SORT routines, allowing our researchers to sort in memory significant data sets without processing to disk."

LESSONS LEARNED

- **An Intel® architecture-based data center is more user-friendly for social science research.** Intel-processor-based servers are more accessible to CISER's general users, who are more Windows-oriented, than previous UNIX-based systems. Usage rose dramatically (from 200 to 500+ active users) when CISER switched to its current Intel architecture-based system.
- **Consolidate servers to lower TCO.** With its acquisition of a 24-processor Unisys ES7000, CISER expects to phase out 12 smaller servers and spend between 30 and 50 percent less time installing and maintaining software.
- **Long-term investment protection.** With the ES7000 platform, CISER will be able to easily upgrade to the latest Intel processor technologies without replacing the entire system.
- **Match processors with application needs.** CISER's extensive testing showed that, for less demanding jobs and fewer users, the Intel Xeon™ processor MP was the perfect match, providing exceptional performance and breadth of applications. However, for large data sets and demanding users who want a Windows-based system, nothing beats the performance of Itanium® 2-based solutions. Benchmark your high-end applications carefully to determine the right high-performance Intel processor and server configuration for your needs.

*Solution Provided
by SAS and Unisys*



Intel works with the world's largest community of technology leaders and solution providers—from software and hardware to systems integration and services companies—that are all using Intel® products, technologies and services with a common goal of providing better, more agile, cost-effective business solutions for you.

Find out more about a business solution that is right for your company by contacting your Intel representative, or visit the Intel® Business Computing Web site at: intel.com/ebusiness or its industry solutions specific sites: intel.com/go/retail, intel.com/go/manufacturing, intel.com/go/digitalmedia, intel.com/go/finance, intel.com/go/telco, intel.com/go/hpc, intel.com/go/energy.



Intel, Intel Centrino, Intel Xeon, Intel Itanium and Intel XScale are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

* Other names and brands may be claimed as the property of others.