



The Woods Hole Research Center

A High-Performance Computing Case Study February 2010

Summary

The Woods Hole Research Center is a nonprofit institute focused on global environmental science, education, and public policy. In 2005, scientists at the Center undertook an ambitious study to map forest biomass of the conterminous United States—important because vegetation sequesters carbon dioxide, one of the biggest contributors to the greenhouse effect.

The Center’s High-Performance Computing (HPC) solution was instrumental in the successful completion of this study in early 2010, but was unable to address the computational and storage requirements of the next phase—mapping the world’s pan-tropical forest cover.

WHRC selected a Sun Microsystems HPC solution provided by Corporate Technologies, Sun’s 2008 Partner of the Year. This solution was designed to meet their stringent computing and storage requirements, in an environment characterized by exponential data growth. »

Customer Results

Corporate Technologies developed a HPC solution designed to meet the Woods Hole Research Center's stringent computing and storage requirements, in an environment characterized by exponential data growth.

Industry	Nonprofit institute
Customer	The Woods Hole Research Center
Problem	Define a new computing infrastructure with sufficient performance and storage to support research featuring satellite imagery to monitor the world’s pan-tropical forest cover and carbon stored in vegetation in the context of energy-efficiency, growth, and budgetary requirements



Solution	A Sun-based High-Performance Computing environment, capable of delivering the needed performance at every stage of the research workflow—image uploads, data computations, visualization, analysis, and backup—utilizing energy-efficient components and designed to scale infinitely
Benefits	<p>10- to 20-fold increase in compute ability and storage capacity versus their current HPC solution</p> <p>Energy-efficient components, which allow these improvements while minimizing power/cooling requirements</p> <p>Reduced time-to-analysis by enabling remote desktop access to data on the blades</p> <p>Enhanced analytics that pinpoint performance problems</p> <p>Infinite scalability, both vertically and horizontally</p> <p>Knowledge transfer to ensure the HPC solution would fully support the research</p> <p>Delivered within the budgetary scope of the WHRC grant</p>
Products	Sun Sun Blade 6000 chassis, X6250 and T6320 blade server modules, 7410 Unified Storage System, SL48 Tape Library, Sun HPC Grid software, and Cisco C2960G switches

Customer Results

The HPC solution delivered a 10- to 20-fold increase in compute ability and storage capacity versus their current HPC solution and reduced time-to-analysis by enabling remote desktop access to data on the blades.

The Challenge

Research

The Woods Hole Research Center (WHRC), located in Falmouth, Massachusetts, is an independent, nonprofit institute focused on environmental science, education, and public policy, with the objective of developing practical approaches to maintaining the planet’s natural assets in support of human well-being. For example, scholars from the Center were significantly involved in crafting the Kyoto Protocol, an international agreement forged in 1997 that seeks to limit world greenhouse gas emissions to slow the progress of global warming. As of February 2005, 141 countries had ratified the Protocol, although scientists agree that even if all parties to the Kyoto conference eventually ratify the Protocol, it would reduce greenhouse gas emissions just 5.2%.

Understanding that additional measures will be required to stabilize the composition of the atmosphere, Dr. Josef Kellndorfer, Associate Scientist at WHRC, engages high-resolution satellite imagery in the pan-tropical mapping and monitoring of forest cover to support the UN’s REDD (Reducing Emissions from Deforestation and Degradation) program to slow tropical deforestation. “About twenty percent of yearly carbon dioxide emissions around the globe are from land-cover change, so if we can compensate countries for ‘avoided deforestation’, we tip the balance in



favor of forest preservation and sustainable management,” KelIndorfer said. “But first we need a baseline of forest cover.”

KelIndorfer’s team is providing that baseline. Their research utilizes cloud-free radar imagery from the Japanese Advanced Land Observing Satellite (ALOS). “The radar sensor is particularly powerful at separating forest from deforested areas, and the Japanese Space Agency (JAXA) has committed to collecting these high-resolution data sets at least once a year,” explained KelIndorfer. His WHRC team of scientists takes these images, runs calculations and data interpolations, and then converts the data into forest cover maps.

Collaborative research with Center colleagues is also producing medium-resolution, pan-tropical biomass/carbon maps based on the fusion of optical (MODIS), radar, and lidar (GLAS) data. The maps create an invaluable reference against which changes in forest cover and carbon stock can be measured.

First results of the study can be found at the WHRC’s website at <http://whrc.org/pantropical>

High-Performance Computing

A High-Performance Computing (HPC) infrastructure was instrumental in the successful completion of KelIndorfer’s earlier research that mapped the biomass of the conterminous United States. However, that system was unable to encompass the computational and storage requirements of the next phase of research—mapping the world’s pan-tropical forest cover.

WHRC selected a Sun Microsystems HPC solution provided by Corporate Technologies after considering options from a number of HPC solutions providers.

Performance

“The overriding challenge placed on the new infrastructure was performance,” said Peter Galvin, Corporate Technologies Chief Technology Officer and leader of the solutions recommendation team. “[The Center] needed an overall solution, in an environment characterized by exponential data growth, where all components are high-performance: storage for the satellite imagery, fast data feeds to high-performance blade compute servers, a new network to rapidly move the data, and a new back-up infrastructure to save calculations.”

Energy Efficiency

There were several other specific requirements for the High Performance Computing solution, the first being energy-efficiency. “The Woods Hole Research Center campus is committed to shifting from a reliance on fossil fuels towards renewable energy. Our campus relies upon energy produced from a photovoltaic array and a wind

Customer Results

WHRC needed an overall solution where all components are high-performance: storage for the satellite imagery, fast data feeds to high-performance blade compute servers, a new network to rapidly move the data, and a new back-up infrastructure to save calculations.



turbine, and we want all resources, including our data center, to be as efficient as possible. The second is a practical matter—our data center has a small footprint and we must engineer within our power/cooling envelope,” said Michael Ernst, Director of Information Systems.

Data Analysis

The second requirement involved the location for data analysis. KelIndorfer’s team wanted to be able to run analysis on compute servers, eliminating the need to move large data-intensive files over the network. Their current HPC solution required that they process satellite image calculations on the compute server, creating a graphical data set, then transfer those data sets to their desktops, where analysis is done. The time-to-analysis was simply too long.

Storage

The third was the projected storage requirements, currently just under 50 TB. “Right now our computations are in the order of 25 TB and the resulting forest cover images are somewhat lower. However, as soon as we start adding all the temporal data sets, we will need to vastly increase that storage capacity,” KelIndorfer noted.

Growth

Sheer growth in the data set was a significant concern because, in addition to the current research study’s larger geographical area, several other factors were likely to require growth in both performance and storage: higher resolution imagery requirements, more frequent data collection, metrics such as watershed analysis in addition to forest cover, and enhanced data calculations. KelIndorfer remarked, “We also plan to expand the research from the tropics to the temperate rain forest and include multi-temporal data sets. Additionally, we are going to be a major processing house for the Group on Earth Observation for Carbon-Tracking Tasks, so exponential growth in the data sets is probable.”

Budgets

Overarching these challenges was a budgetary issue. As Corporate Technologies’ Peter Galvin observed, “The Center had a fixed budget, and they needed a lot within that budget: storage, networking, compute, software, implementation services, knowledge transfer, and it all had to fit within the grant they received. For a project that serves all of our common interests, we certainly wanted to help them out as much as we could. We leveraged our vendor partnerships to provide aggressive pricing within their available budget.”

Customer Results

Sheer growth in the data set was a significant concern because, in addition to the current research study’s larger geographical area, several other factors were likely to require growth in both performance and storage: higher resolution imagery, more frequent data collection, watershed analysis in addition to forest cover, and enhanced data calculations.



The Solution

Corporate Technologies recommended a tightly integrated High-Performance Computing infrastructure to support KelIndorfer's research initiative. "Given the computational and growth requirements, we knew that each component of the solution had to be extremely high performing, while also being cost- and energy-efficient," remarked Galvin.

At the heart of the solution is a high-performance compute engine comprised of 20 Sun X6250 blade server modules, currently supporting up to 4 sockets (3GHZ Quad Core CPUs) with 128 GB RAM, and Solaris, Linux, and Windows operating systems.

Storage is provided by the Sun 7410 Unified Storage System, which utilizes a large solid state flash memory cache to accelerate reads and writes, and a disk storage pool comprised of 7200 RPM SATA drives, and provides for unlimited disk snapshots, compression, and replication. A Sun SL48 Tape Library provides a 76.8 TB LT04 tape capacity for backups. High-speed networking is provided by Cisco C2960G switches with a 32 Gbps switching fabric. Sun HPC Grid software rounded out the solution, permitting parallel job support and maximum resource utilization.

Performance

"We tried to take a holistic view of their performance needs to make sure our infrastructure would fully support WHRC's research," said Galvin. "This implementation represents a highly integrated HPC solution, designed to deliver the performance the research team needs at every stage of the research workflow, including image uploads, data computations, visualization, analysis, and backup," he said.

Energy Efficiency

Energy efficiency, the second requirement placed on the HPC solution, was achieved by selecting components with documented low-power consumption. For example, the Sun blade chassis is up to 45% more efficient than competing rack-mount servers, and the Sun Unified Storage System, which utilizes Solid State Disks, increases performance while cutting power consumption by up to 80% compared to high-speed hard disks. "Power consumption was a foremost concern with a facility trending towards adding power to the electrical grid, rather than removal from the grid," said Ed Hamilton, Corporate Technologies Principal Solutions Architect and WHRC Project Lead. "Our final tally showed a total of 8707 Watts at 100% utilization, which is remarkably low for such a robust infrastructure. Given the small footprint of the data center, we also calculated heat output at 35,820 BTU/HR and conservatively recommended a 4+ ton cooling unit. Assuming the power is equally distributed between the two UPS, this would provide approximately 15-20 minutes of run time in the event of a power failure," Hamilton noted.

Customer Results

At the heart of the solution is a high-performance compute engine comprised of 20 Sun X6250 blade server modules, currently supporting up to 4 sockets (3GHZ Quad Core CPUs) with 128 GB RAM, and Solaris, Linux, and Windows operating systems.



Data Analysis

To make sure that the proposed infrastructure would reduce time-to-analysis, Corporate Technologies undertook a Proof of Concept in its Technology Lab. “Their current HPC solution used the compute server to process data and create graphical images, but then transferred the images to desktops for analysis,” said Galvin, noting that transfer of these massive files over the network was very time-intensive. “We believed that a strategy of keeping the data and images on the blades, then remotely accessing them from the desktop would improve time-to-analysis. So in our Proof-of-Concept we used the proposed storage and blade infrastructure, installed their major application, and enabled remote desktop access to the application running on the blades. Testing response time showed that the solution was quite useable and provided better performance than copying the image files between the blade servers and the desktops,” Galvin said.

This data analysis solution was a key to overall performance of the HPC solution. The data in the storage array had to store and feed the data fast enough to the blades, and the blades had to compute it fast enough, and then the desktops would need fast enough remote access to view the graphical data set as it was stored on the blades.

Enhanced Analytics

In addition to meeting their capacity and performance requirements, Sun’s Unified Storage System is unmatched in its analytics, which allow real-time monitoring of the storage appliance.

“The absence of such insight was one of their biggest challenges with their current HPC environment,” said Corporate Technologies’ Ed Hamilton. “Every other solution has generic reports, which are great for trending, but they do not tell you the source of the problem. With our storage analytics, you can see the time it took for a single request from a single box and track it down to a single spindle. It tells you exactly why storage is running at 50% utilization, and no other solution can tell you that,” he concluded.

Growth

WHRC knew that flexibility for expansion would be an essential requirement if they were not going to outgrow the selected HPC infrastructure. Several months after working with the new infrastructure, they were already considering an additional 20 – 30 TB of storage to handle increased computational needs.

“Our solution infinitely allows growth,” said Hamilton. Horizontally, it allows the purchase of additional blades, or more compute nodes; the switches themselves can be vertically expanded, the storage can be both vertically and horizontally expanded—additional trays on the head, additional heads. We worked hard to ensure that this system would be able to handle future growth in data storage and computational needs.”

Customer Results

To make sure that the proposed infrastructure would reduce time-to-analysis, Corporate Technologies undertook a Proof of Concept in its Technology Lab. Testing response time showed that the solution was quite useable and provided better performance than copying the image files between the blade servers and the desktops.



Budgets

Corporate Technologies was able to deliver its HPC infrastructure meeting both WHRC's stringent technology and budgetary requirements. "We enjoyed a great interactive relationship with the research scientists and technologists on the WHRC team," noted Corporate Technologies' Peter Galvin. We went through 3–4 configurations and architect diagrams, proposed budgets, and finally settled on the solution that felt right to all of us. It is a pleasure to be part of a solution so focused on improving the human condition," he concluded.

Knowledge Transfer

The final phase of Corporate Technologies engagement with the WHRC team was knowledge transfer. "Dr. Kellndorfer's team—both the research scientists and the technologists—were very HPC-savvy," remarked Ed Hamilton, Corporate Technologies' Lead Architect, "but our infrastructure introduced new technologies with which they had no experience."

He continued, "We did 3 days of general knowledge transfer, 6 to 8 hour sessions, with everybody involved and ran through the entire infrastructure soup to nuts. We showed them what was there, how to effectively manage each of the components, basically sharing our tricks of the trade. We also did a 3-day knowledge transfer on the Sun HPC Grid software and how you operate it to maximize performance. We even configured a framework of standard daily, weekly, and monthly Symantec Backup Exec policies, so that that WHRC staff could quickly customize their data protection to match their evolving needs," he concluded.

The Result

After using the HPC solution for several months, the WHRC scientists and technologists report that their requirements are met.

"I believe that one of the many reasons our HPC infrastructure is successful is that it was architected expressly to meet their specific needs," observed Hamilton.

"Before we did any work, we held architectural discovery meetings with everyone involved. We concentrated up front on things like the application data flow, the existing networking infrastructure, what components in the new solution would be shared with the existing HPC solution, the functionality they needed from the grid, and fundamental speeds and feeds. There are very different ways of implementing the infrastructure, we just wanted to make certain that the one we deployed perfectly met their requirements," he concluded.

Customer Results

After using the HPC solution for several months, the WHRC scientists and technologists report that their requirements are met. One of the many reasons attributed for this success is that the HPC infrastructure was architected expressly to meet the Center's specific needs, determined through architectural discovery meetings.



KellIndorfer commented, “The Sun infrastructure was selected because we liked the fast data access. It is data-intensive research we are doing here, and hence we were quite keen on having that high-performance, as illustrated by the Sun Unified Storage System, with solid state disks and Readzilla Flash that give us great raw performance. We have programmed up a really nice system of very efficient processing now—processing is really, really fast.”

He continued, “When we really fill up the RAM for these processes and move the data between the compute nodes and the storage, we find some situations where the access can slow somewhat, but we are finding ways to work with the Sun Grid engine to help us with prioritization if we have interactive jobs that need to run faster. Sun Grid Engine is a very nice tool, which was an additional reason to go with the Sun infrastructure because the job scheduling and farming it all out makes a lot of sense to us.”

Sun Blade Server’s Modular system, with multiple operating systems support, fits WHRC’s needs for flexibility. If new requirements call for more instances of an operating system, or other operating systems, the blades can be reconfigured to run that software while still having access to the shared storage and being part of the HPC Grid. “The support for multiple operating system platforms—the ability to have Solaris blades mixed in with Windows blades and Linux blades—fits in perfectly with what we know we need to do,” said Michael Ernst, Director of Information Systems at the Woods Hole Research Center.

KellIndorfer also expressed satisfaction with the improvements in “time-to-analysis” through enablement of remote desktop access to the application running on the blades, versus copying image files between blade servers and desktops. “It is definitely much faster,” he observed. “I can’t put a number on it really but that’s been thought from the design beginning. Rather than moving files across the network to individual workstations throughout the building, we have a thin client set-up where data and compute nodes remain very close together with a fast interchange.”

The Woods Hole Research Center is also very pleased with a final important requirement, energy-efficiency. Ernst explained, “We are measuring actual kilowatts going into the data center and are surprised to see how much less energy is being used by this new equipment than we had expected. We expected an increase of 5 – 7 kilowatts, and it is more like 2 – 3 kilowatts over what we already had installed.”

This reduced power consumption makes the Sun computing infrastructure a suitable complement to a facility run largely on renewable energy generated by an array of 88 photovoltaic panels and a newly installed 100kW wind turbine. “This is probably one of the best systems we could have found in the current market to allow us to accomplish the work we

Customer Results

The Sun infrastructure was selected because WHRC liked the fast data access. Its data-intensive research required high-performance, as illustrated by the Sun Unified Storage System, with solid state disks and Readzilla Flash that delivered great raw performance.



need to do without completely compromising our efforts to drive our operations to carbon-neutrality.”

About Corporate Technologies, Inc.

Founded in 1994, Corporate Technologies, Inc. is one of the largest providers of complete IT solutions and services to enterprises in the Northeast. We provide clients with professional services expertise in the areas of IT Strategy, Business Intelligence, and Data Centers, supported by a selective product line, support contracts and services, IT staffing solutions, and financing options.

The company is a top partner with many technology leaders including Sun Microsystems, NetApp, SAP/Business Objects, Oracle, Symantec, VMware, Juniper, and F5 Networks. Its annual investment in R&D and the company's market-leading Technology Lab allow Corporate Technologies to continually evaluate and recommend the best current and emerging technologies.

Customer Results

The reduced power consumption makes the Sun computing infrastructure a suitable complement to a facility run largely on renewable energy generated by an array of 88 photovoltaic panels and a newly installed 100kW wind turbine. The solution allow WHRC to accomplish the work it need to do without completely compromising their efforts to drive their operations to carbon-neutrality.