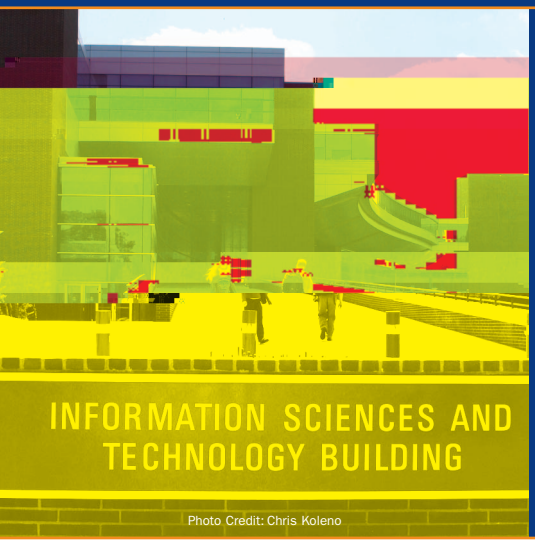

Penn State Turns to BlueArc Titan to Enable Artificial Intelligence Research

"It's impressive that a single BlueArc Titan NAS head is able to handle all of the traffic. It stands between our users and our storage disks and it funnels requests at an incredible, consistent rate."

"When the research team gets an HPC cluster, it means hundreds of computers are working together as one, and each one is mounting and accessing disk space on our BlueArc system. Hundreds of file requests are pounding image files for days at a time, as the research project looks for visual markers."

- John Domico, Penn State Asst. CSE director of academic research computing

Summary

The Pennsylvania State University (Penn State) Department of Computer Science and Engineering (CSE) depends on BlueArc's Titan 2000 storage system to support a wide range of research projects designed to investigate computers' role in medical diagnostics; the behavior of crowds; training computers to recognize the content of images; and other aspects of artificial intelligence and computing. BlueArc's Titan serves as the massively scalable, reliable foundation for CSE at Penn State, providing unimpeded access to data for thousands of concurrent requests per second as faculty and students seek new ways to benefit from artificial intelligence.

The Customer

The Department of Computer Science and Engineering is part of the Penn State College of Engineering, established in 1896 and ranked among the U.S. News & World Report top 25 graduate schools of engineering. CSE is also a recipient of major federal research grants, such as a multi-year, multi-million-dollar award from the National Science Foundation (NSF) Computer & Information Science & Engineering (CISE) for work related to computer infrastructure.

Roughly one in fifty engineers in the United States has earned his or her bachelor's degree in engineering from Penn State. Penn State Engineering strives for national leadership in innovative engineering education, vigorous R&D partnerships with U.S. industry, and service initiatives through continuing education, technology transfer, and support for governmental agencies.

The Challenge

At Penn State, CSE maintains a computing infrastructure that reflects the diverse nature of the research conducted by faculty and students. Genetics research demands high input/output (I/O) performance to process thousands of small files, while research that uses large image files requires high throughput. These projects compete for disk access at the same time that the department's heterogeneous IT presents the risk of incompatibility in system-specific user access.

In order to meet unpredictable demand, John Domico, asst. director of academic and research computing for Penn State CSE, considered purchasing dedicated storage platforms for each project, which would have required significant up-front expense and long-term costs of administration and maintenance.

The Solution

After it received a grant from the National Science Foundation (NSF), Penn State selected a BlueArc Titan 2000 storage system as an alternative to adding more storage-attached networks (SANs) for its growing research needs. Already on Domico's shortlist, his decision was aided by the insight from an independent research team that designs field programmable gate arrays. That team pointed out to Domico that BlueArc Titan's distinctive hardware architecture would likely offer a performance advantage.

Penn State's BlueArc storage system, featuring greater than thirty terabytes of storage, interacts with four high-performance computing (HPC) clusters, each averaging 100 nodes. Following Penn State's initial deployment, BlueArc Professional Services evaluated and fine-tuned it for even greater system performance and usable space.

The Results

An NSF grant for an interdisciplinary study of the behavior of crowds is funding just one of the latest research projects underway at Penn State's College of Engineering. Other initiatives include the Laboratory for Perception Action and Cognition (LPAC), where researchers are training computers to interpret visual sensor data that can in turn drive robotic action, as well as working on systems that can learn from training examples. LPAC projects have a wide range of applications, such as biomedical analysis—training computers to recognize visual characteristics of disease from among thousands of high-resolution medical images.

"When the research team gets an HPC cluster, it means hundreds of computers are working in parallel, and each one is mounting and accessing disk space on our BlueArc system," explains Domico. "Hundreds of file requests are pounding image files for days at a time, as the researchers look for visual markers."

Another project is developing software to help computers recognize qualitative aspects of images. Called ALIPR (Automatic Linguistic Indexing of Pictures), the project is training computers to analyze images pixel by pixel and calculate the probability that a specific word describes its content. The software compares image color and texture to those characteristics in a stored database of 50,000 images that have been tagged with words for particular qualities and characteristics.

"In academic research, you can't plan for a likely usage scenario," says Domico. "We have to support systems all over the map, including Fedora, Red Hat, and Ubuntu Linux, Solaris, Windows, and OS X. We can't really mandate a standard, as users are designing their own research environments and need to get at the kernel level."

Approximately 500 people regularly conduct research that depends on the BlueArc system. Yet as some operating systems don't support all popular methods for accessing file space, Domico and his team have to work to keep user access permissions mapped between systems, enabling a diverse set of users to access projects and work uninterrupted. As Penn State's research projects continued to grow, Domico says that he and his team reached a point where they had to consider the purchase of multiple storage platforms to support users, incurring significant expense to purchase and support new systems over the long term.

Domico is pleased that Penn State never had to make that choice. Instead, BlueArc Professional Services performed an on-site evaluation of Penn State's system and proposed a solution that would support growing demand, without a major capital investment in new hardware and software—or additional IT support.

Today, Penn State's BlueArc system routinely exceeds the Titan 2000's rated I/O capacity. "We didn't have to spend hundreds of thousands of dollars on new platforms and related IT support. Our storage array is processing immense research traffic, diverse I/O profiles—and it's also housing and serving some of our administrative storage needs, as well."

The Conclusion

At Penn State, research projects attract millions of dollars in grants for the pursuit of meaningful breakthroughs in artificial intelligence applications ranging from medicine to national security. To deliver on high-profile research proposals, faculty and students need to be able to count on a computing infrastructure capable of supporting massively data-intensive files and high-volume transactions at blinding speeds. For the College of Engineering, it's essential to have a storage system capable of scaling to handle the unpredictable demands of simultaneous activity by thousands of users.

BlueArc's Titan system lets CSE research teams at Penn State harness the power of cutting-edge storage architecture, to eliminate technology bottlenecks to action and innovation. What's more, BlueArc's forward-looking approach to storage helps Penn State maximize the value potential of its technology investments.

BLUEARC®

BlueArc Corporation
Corporate Headquarters
50 Rio Robles Drive
San Jose, CA 95134
t 408 576 6600
f 408 576 6601
www.bluearc.com

BlueArc UK Ltd.
European Headquarters
Queensgate House
Cookham Road
Bracknell RG12 1RB, United Kingdom
t +44 (0) 1344 408 200
f +44 (0) 1344 408 202