BlueArc Accelerates Performance of H.W. Wilson's Vast Online Library

"We're a small company with customers all over the world, and our Titan cluster gives us the peace of mind that our library will remain up and running around the clock."

— Lu Parziale, H.W. Wilson
Vice President, Information Systems

"The quality of support we got from the BlueArc team during testing was unparalleled. Even when the evaluation period expired, they were still calling us to find out how it was going and answer our questions."

— Sat Persaud, H.W. Wilson
Senior Network Engineer

Summary

After two decades of transition to digital media formats, venerable research information provider H.W. Wilson was left with a collection of data storage technology and a decentralized system that could not keep pace with new information acquisition and data volume. With BlueArc Titan storage at the heart of a centralized, tiered storage infrastructure, H.W. Wilson has been able to extend century-old paper archives and legacy storage to a Web-based digital library that offers unparalleled access to more than 20 million digital documents, without stacking new storage management burdens on IT staff's desks.

The Customer

Founded in 1898, H.W. Wilson (www.hwwilson.com) is a leading publishing company providing a variety of full-text and image databases to all types of libraries and institutions worldwide. The company pioneered index materials for researchers in fields ranging from science to art, law, education and general-interest topics, and today maintains a vast document library with millions of digital files accessible through the WilsonWeb database available on the Web, or in print. Schools, universities, public libraries and corporations depend on H.W. Wilson for breadth and depth of material spanning thousands of books, periodicals, reference monographs, images, reviews and profiles.

The Challenge

Over the decades, as H.W. Wilson shifted its library from print to mainframe to CD-Rom to Web-based technology, its storage infrastructure had to evolve along with new data formats. "It was a definite step up from direct-attached storage and Fibre Channel disks to about two dozen devices and a decentralized infrastructure," says Lu Parziale, Vice President, Information Systems at H.W. Wilson. "But hardware obsolescence put us in a bind as our data continued to grow."

Parziale's team built a front end for about two dozen storage devices and NFS protocol support, but huge data growth taxed the server system's performance, backups, and storage management became a challenge.

The Solution

Parziale concluded that centralization would help ensure accessibility, data integrity and security for millions of documents. The criteria for a new storage solution put ease of management—including deduplication and access controls—at the top of the list, followed by performance, scalability and file system limitations. Sat Persaud, Senior Network Engineer, explains that Titan's closest competitor could not address H.W. Wilson's requirement for single storage volume capacity of 20 million files or more.

BlueArc implemented an evaluation system including two clustered Titan servers, 29 terabytes of Fibre Channel storage and 36 terabytes of serial advanced technology attachment (SATA) storage.

BLUEARC®

The Results

Parziale compares H.W. Wilson to another premier research and reference resource that serves the legal profession. "We maintain the data that you can't find or get through even the most popular Web search engines," he explains. "Our data has to be accessed by query, rather than simple links. As a consequence, our storage requirements and information retrieval tasks can be much more data-intensive."

New WilsonWeb users may take for granted the system's fast response to their library searches. Prior to the Titan implementation, even network file system (NFS) servers with gigabytes of capacity couldn't change the fact that WilsonWeb response time would degrade daily during regularly scheduled database releases from seconds to almost a full minute. Titan keeps response time down to seconds—or less. Parziale's team reboots the Web application each morning in a process that is also about 40 percent faster as a result of the Titan implementation.

Parziale observes that H.W. Wilson's data has grown dramatically in just the two decades since the company began offering electronic products. In contrast to most vendors, the research information firm maintains full text of documents dating all the way back to 1898, in ASCII, HTML and/or PDF file formats. Even as the library's document collections continue to expand, the Titan cluster continues to deliver uninterrupted service to customers. With one node supporting customers and the other node supporting internal activity, the cluster also provides failover support. Persaud notes that a failure on a Sunday went unnoticed by staff and users until Wednesday. "We also wanted something that could stay up at all costs. It's the first time technology has worked like it should," says Persaud.

That reliability extends to backup tasks as well. The Titan system manages a tiered storage infrastructure that uses Fibre Channel for applications and SATA to support less frequently accessed material, with data replication from the Fibre Channel to the SATA tier. Where backup took days, the Titan system now backs up millions of files flawlessly in just a few hours, and the information is copied to SATA storage. Where the previous storage infrastructure's limitations meant that Parziale's team only backed up the larger data sets weekly, Titan supports a complete backup, daily.

Finally, H.W. Wilson is benefiting from centralization, including better use of resources. Instead of transferring 120 gigabytes of data on two different systems in a process that could take as long as six hours, the Titan cluster transfers a comparable volume in approximately 20 minutes, without the need for multiple copies.

The Conclusion

As the H.W. Wilson collections grow and new products are developed, Parziale appreciates that Titan's hardware-based architecture and modular design will make it easy for H.W. Wilson to add capacity. "We liked that Titan's performance isn't tied to processor speed, and looking 10 years down the line, we liked that Titan's design wouldn't present the upgrade challenges we knew we'd face with other storage servers," he says.

Instead, Parziale anticipates that "we'll just buy modules or add disks, and we're good to go." That ease of use is especially important to Parziale. "My employees—who have to be network technology experts—won't have to become storage specialists as well."

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