

COMPUTERWORLD®

Peer Perspective. IT Leadership. Business Results. | COMPUTERWORLD.COM | JUNE 2011

DATA TIERING

Ready, Set, Tier!

THESE EARLY ADOPTERS ARE STACKING UP THE SIMPLE REWARDS OF **AUTOMATED DATA TIERING**, WHICH DETERMINES WHICH STORAGE LAYER GETS WHICH DATA.



F THE 5,000 STUDENTS AND 250 FACULTY MEMBERS at Ohio State University's Fisher College of Business have one thing in common, it's an obsession with saving everything on the university's servers.

In 2005, timely data such as classroom presentations and senior thesis documents mingled with dormant files such as three-year-old homework assignments on storage servers. The amount of stored data was growing 20% each year and creating a huge drag on systems when the important stuff was called up.

"Our storage needs were growing so fast," recalls Brian Wilson, principal system engineer/developer at the business school. "We spent 25% to 50% of [IT staff] time constantly rebuilding servers just to handle our storage needs."

Enter [automated data tiering](#). In 2005, Wilson's team installed an automated tiering system from Compellent Technologies (now Dell Compellent), becoming one of the product's first users. They first experimented by migrating older student files to the system.

"We watched it over the next week or two and saw how it migrated the data from top, fast disks to big, slower disks, and we were convinced. It worked exactly how they told us it would," Wilson says. Over the next four months, the school migrated the rest of its 30 to 40 servers to Compellent — and it's still reaping benefits.

Wilson estimates that the university saves at least \$100,000 annually by not having to buy storage on speculation. What's more, the system helps the school save on payroll: His team has been cut from six full-time operations staffers to three.

"A good portion of their time [was spent] on maintenance," Wilson explains. Server failures can now be addressed by simply plugging cables into a different server.

Ohio State is one of many early adopters of automated tiered storage — a feature that was relatively new to the scene a year ago but is now being touted for its ability to improve performance, save money and enable IT staffers to spend less time fixing things and more time enhancing technology for the business.

BY STACY COLLETT

"I expect [automated data] tiering to be a checklist item relatively soon, and that 'commonality' will drive comfort that will spur usage," says Mark Peters, an analyst at Enterprise Strategy Group. But just because automated tiered storage is now available on existing storage systems, that doesn't mean the feature is being used. "Much of the commentary around tiering focuses on performance, but it should equally be about economics," he adds.

Automated data tiering not only controls the movement of data; it also monitors how data is being used and determines which data should be on

DATA TIERING

which type of storage as demand for it rises or falls. Older or less frequently accessed data is moved to slower, less expensive storage such as SATA drives, while the most frequently accessed data can be transferred to faster, more expensive Fibre Channel drives or even to high-performance [solid-state drives](#).

When Speed Is What You Need

Mount Sinai School of Medicine in New York is home to modern experimental technologies for genetic and genome sequencing projects that involve massive data sets. The research aligns hundreds of millions of fragments of RNA/DNA against the human genome for analysis. Many projects involve searching 30GB to 40GB of genome sequence data against similar data sets — a very disk-intensive job with lots of reads and writes to the storage system that result in a huge bottleneck, says Ravi Sachidanandam, assistant professor of genetics and genomic sciences at Mount Sinai.

The school deployed an automatically tiered NAS appliance from Avere Systems for use in large-scale processing projects to improve read and write times and boost processing speeds.

The school's biggest ROI comes from time saved in analyzing sequences. Processes that once took six to eight hours are now done in half the time. Automated tiered storage also helps optimize the use of existing computers because they're not tied up for long periods of time.

Reads, writes and metadata are allocated to storage via the appliance. Allocation algorithms monitor how often certain patterns are accessed and then place the data on one of multiple internal tiers. This increases performance, distributes the workload in the cluster and minimizes requests to the mass storage server.

Avere acts as a caching device, so data generated at high rates can be sent through it to large-scale storage devices. This process happens automatically and is transparent to the end user of the NAS device.

The NAS device also “has many nodes and internal ways to keep all nodes in sync,” Sachidanandam says. “When you start doing lots of writes more quickly, they get out of sync. That’s another one of the problems we solved using [automated data tiering].” Movement of data now occurs in real time at the file or block level, where the data is isolated and written to a specific area, while keeping the rest of the nodes in sync.

The appliance can be attached or detached at will, which created some angst for Sachidanandam early on. “The worst fear was [the thought of] taking the appliance out of the loop” and what would happen to the data, he says. “But it’s fairly transparent and easy to work around” because it mirrors storage. Having used the appliance since the fall of 2010, he is now at ease with the automation. “It works, so you don’t care exactly what it does,” he says.

At Blood Systems, a Scottsdale, Ariz.-based provider of blood banking and other services, IT was looking to improve the performance of its storage systems, but it couldn’t afford to buy solid-state disks for all of its critical data.

When you start doing lots of writes more quickly, they get out of sync. That’s another one of the problems we solved using [automated data tiering].

RAVI SACHIDANANDAM,
ASSISTANT PROFESSOR OF GENETICS
AND GENOMIC SCIENCES, MOUNT SINAI



Black expects that automated data tiering will help Blood Systems save money on future storage purchases. “I will be able to purchase less of the more expensive tiered storage,” because only the data in highest demand will migrate to expensive storage devices, he says.

Sachidanandam advises companies to be sure they really need automated data tiering before making an investment in the technology. “Does your processor require this type of speed-up? Not many companies may have this need,” he says.

Black advises would-be users to make sure that the automated tiering system they’re considering is thoroughly tested for their environments, and that it meets their specific requirements — as they should with any new technology.

He admits that he was “very apprehensive at first” about ceding control of the data to an automated system. “We want to make sure it’s going to work before we turn it on. So IT staff first set it up and put some test systems into the tiering storage pools,” says Black. “We watched that for a while and made sure there were no adverse effects. Everything happened the way we expected it to. Then we moved the test systems out of Tier 0 disk and started moving our production systems into tiered storage. So far, we’ve been very pleased.” ♦

Collett is a Computerworld contributing writer. You can contact her at stcollett@aol.com.

A V E R E

Avere Systems, Inc. • 5000 McKnight Rd. • Suite 404 • Pittsburgh, PA 15237 USA