

“Our data just keeps growing no matter what happens on the field,”

SAYS JOSEPH MILONE,
SENIOR DIRECTOR OF INFORMATION SYSTEMS
AND TECHNOLOGY AT STERLING
AMERICAN PROPERTY INC., AN AFFILIATE
OF STERLING EQUITIES, WHICH IS
THE PARENT COMPANY OF THE NEW YORK METS.

Managing the information that drives the enterprise

STORAGE

Vol. 7 No. 1 March 2008



also inside

- Users rate top tape libraries
- The new SRM: Vendors retool management apps
- New York Mets take a swing at deduplication

ARRAYS DO DOUBLE DUTY

Multiprotocol arrays support block- and file-based storage, giving users the best of both worlds: NAS and SAN in a single box.



"The Mets needed to recruit some backup help... Data Domain was drafted for the job."

New York's "other" baseball team installed deduplication appliances in the first stage of retooling its backup processes. By Alan Radding

New York Mets add deduplication to their roster

THE NEW YORK METS, a legendary baseball team, were seemingly on top of the world in 2007 and cruising toward a playoff spot with a seven game lead. At the same time, the organization's IT group was planning to extend its wireless network to accommodate the hundreds of extra reporters who would cover the post-season games. But with just 17 games left to play, the Mets went into a tailspin, losing 12 of their remaining games—an epic collapse. While that dashed the team's hope of going to the World Series and the need to extend its network, it hardly slowed the rush of data the IT group was scrambling to back up.

"Our data just keeps growing no matter what happens on the field," says Joseph Milone, senior director of information systems and technology at Sterling American Property Inc., an affiliate of Sterling Equities, which is the parent company of the Mets. Sterling American Property, primarily a real estate and venture capital firm, maintains two data centers: one at Shea Stadium and another at its corporate headquarters in Great Neck, NY, approximately 12 miles from the stadium.

The glamour of sports aside, the Mets' biggest data management problem, typical of many mid-sized organizations, comes down to one word: backup. "The data is getting harder and harder to back up," says Milone. With huge volumes of photos and video, as well as the usual corporate data, the organization was facing the need to back up terabytes of data.

The company found itself saddled with a cumbersome, error-prone and labor-intensive backup process. A couple of backup failures were enough to get Milone looking for a new approach. By March 2007, just when spring training

was in full swing, he started thinking about new disk-to-disk (D2D) or disk-to-tape (D2T) alternatives to tape backup. With low-cost disk, virtual tape and newer technologies like deduplication, Milone felt he could not only streamline the Mets' backup requirements but take care of Sterling's other ventures as well.

This puts the Mets right in the sweet spot of the D2D backup market. A recent study by the Enterprise Strategy Group (ESG) found that mid-sized organizations are more likely to turn to D2D virtual tape backup solutions to replace physical tape than are large enterprises, reports Lauren Whitehouse, an analyst at the Milford, MA-based research firm.

Media and more

Fueling the Mets' storage growth is media, mainly in the form of photos. "The Mets love to take photos," says Milone. "In 2006, we took 77,000 photos." In 2007 (when we interviewed Milone), the photographer had snapped 83,000 photos when Milone last checked and the year wasn't over yet. Each 15-megapixel photo represents a 25MB to 30MB file.

The Mets run Windows servers, primarily from Hewlett-Packard (HP) Co. They maintain 18 file and application servers at Shea Stadium and 23 at Sterling, which runs a wider range of applications. In general, the servers are used for file serving, Microsoft SQL Server, Microsoft Exchange and Microsoft Office SharePoint Server.

When Milone began to seriously think about the backup problem, the company had 3.5TB of storage at Shea Stadium and a little more than 1TB at Sterling, all DAS, which eight people managed: five at the stadium and three at Sterling. Milone plans to hire four more people in 2008 and, due to the rapid growth of the company's data, is considering implementing two SANs (one for each data center).

The plan

Given that the company already had two data centers and an assortment of other properties, some of which are suitable as remote backup sites, Milone began formulating a multiphase plan to speed the backup process, eliminate tape and address the company's concerns about disaster recovery. "We could get business continuity and disaster recovery for all the Sterling entities," says Milone. It would just be a question of finding the right combination of technologies at the right price.

Different flavors of deduplication

The key to controlling the growth of data is deduplication technology, which comes in the following forms.

APPLICATION SERVER DEDUPE. Performs dedupe through software running on the application server before the data is sent to the backup server. This reduces the amount of data sent over the network to be backed up, but adds processing overhead to the application server.

BLOCK LEVEL. Provides more granular dedupe by looking at blocks within files that have changed. It will save only those blocks that have changed, not the entire file.

INLINE. Intercepts the data on its way to the disk array and performs the deduplication function before writing the data to disk. The stored data is fully deduped and available for immediate replication or other use. This approach can impact performance.

POSTPROCESSING. Performs deduplication after the data has been stored on the array. It avoids a performance hit, but requires additional storage capacity on the backup system to accommodate the data as it's deduplicated.

SINGLE-INSTANCE STORAGE. Performs dedupe at the file level, which limits the amount of potential reduction by looking only at the file and not drilling down to the block level for duplication. For instance, if the contents of a file are left unchanged and only its name is changed, file-level dedupe won't recognize the duplication. In other words, the file-level dedupe will see it as a new file and not eliminate it as a duplicate.



Compression is often used in conjunction with dedupe to further reduce the volume of stored data. In most cases, companies first dedupe and then compress.

Source: Lauren Whitehouse, Enterprise Strategy Group

For backup, the Mets use Backup Exec from Symantec Corp. Each server contains two network cards; the second card is directly linked to a backup server to which Quantum Corp. DLT tapes are attached. However, the constant tape handling, sometimes eight tapes per backup, was taking a toll on the IT staff. The backup process entailed nightly incremental backups and a full weekly backup, so tapes were constantly shuffled around, leading to occasional backup failures.

The Mets may be Sterling's highest profile investment, but the company has other business units, each with unique backup needs. For example, for Sterling's real estate and investment businesses, Milone started a project last

year to move away from paper by scanning leases and other documents. The company scans as many as 12,000 documents each month and stores them as PDF images. The document images are accessed and searched through a Microsoft SharePoint server. The Mets also have a number of minor league facilities. Over time, whatever backup solution Milone comes up with for the Mets he expects to extend to the minor league teams.

Milone's plan called for D2D backup automation combined with WAN replication, by which each data center could replicate backups to the other with the ultimate goal of eliminating tape completely. While Milone envisioned such daily and weekly backup as part of a comprehensive business continuity/disaster recovery initiative, the initial effort focused on Sterling's need to automate backup. Management agreed and approved approximately \$200,000 for the initial effort, part of what would be a multiphase initiative.

Picking technology

There are many vendors that address the problems faced by the Mets. "D2T has a reliability problem. D2D costs more, but is more reliable," says Mike Karp, senior analyst at Enterprise Management Associates Inc., Boulder, CO.

The Mets' problem was finding the right vendor or combination of vendors. The team's needs involved not only low-cost disk, but deduplication and compression to reduce the overall amount of data to be backed up, as well as WAN optimization/acceleration to speed the replication process. "But by combining too many technologies you risk complicating the solution and introducing potential problems," says Greg Schulz, founder and senior analyst at StorageIO Group, Stillwater, MN.

Most of the D2D backup solutions for midsized companies are packaged as virtual tape appliances, notes ESG's Whitehouse. As virtual tape, the backup product can drop right into the existing tape backup environment without disrupting applications and backup software. "Midmarket companies don't want to get sophisticated about backup strategy. They want to do what's easy, and the easiest [thing to do] is to just drop in an appliance," she says.

Working through a VAR, ePlus Technology, Milone narrowed his search to three vendors: Data Domain Inc., EMC Corp. and Quantum. Milone quickly rejected EMC as too costly. Quantum was the Mets' incumbent backup vendor, having provided the DLT tape system. Data Domain was the newcomer brought in by ePlus Technology.

The final selection came down to Data Domain and Quantum. Both offered similar products and had deduplication, which Milone by that point considered essential. And each offered comparable pricing.

In the end, the Mets opted for two Data Domain appliances. The DD565 came with 7.5TB (raw) of disk storage

and was installed at Shea Stadium. A smaller unit, the DD510, came with 2.25TB (raw) for Sterling. The DD510 lists for \$19,000, while the DD565 sells for \$95,000. The servers would retain their DAS until the SANs were in place, and Backup Exec would remain the backup software.

Both units feature compression and deduplication, which Milone figures would reduce data volume, on average, at a 25x rate. "In some cases, we've gotten as high as 80x data reduction," he notes.

Vendors bicker about which product delivers the greatest rate of data reduction. "A 20x reduction is pretty common, 50x is reasonable," says ESG's Whitehouse. Beyond that, you need to look carefully at the data and how the vendor is calculating the reduction rate, she advises. Even the length of time the data is retained can impact the reduction rate.

The Data Domain appliance uses inline deduplication, which performs data reduction before the data is stored on the disk. This means the data can be replicated or otherwise managed immediately on hitting the disk. But it takes a performance hit in the process (see "Different flavors of deduplication").

Once Milone chose Data Domain, the implementation went without a hitch. "The Data Domain appliance just attached to our backup server," he says. The IT staff handled most of the deployment with the help of a Data Domain engineer, who spent a day preparing the environment and returned a few days later to verify that everything went in correctly.

Each D2D backup appliance handles the servers at its location. In addition, data at Sterling is replicated to Shea Stadium. The organization, however, hasn't eliminated tape completely. "We still do tapes at Shea," says Milone. That will end with the next phase, which involves either replicating Shea Stadium backups to Sterling or, more likely, to a third Sterling property that will house another Data Domain appliance. At that point, both data centers will replicate to the third site and tape will disappear.



**"I sleep a lot better now"
that data backups are faster
and more reliable, says Milone.**

For now, data backups are happening faster and are more reliable than ever. "My staff loves it," says Milone. Whether the Mets win or lose, "I sleep a lot better now," he says. ☉

Alan Radding is a frequent contributor to Storage magazine.