

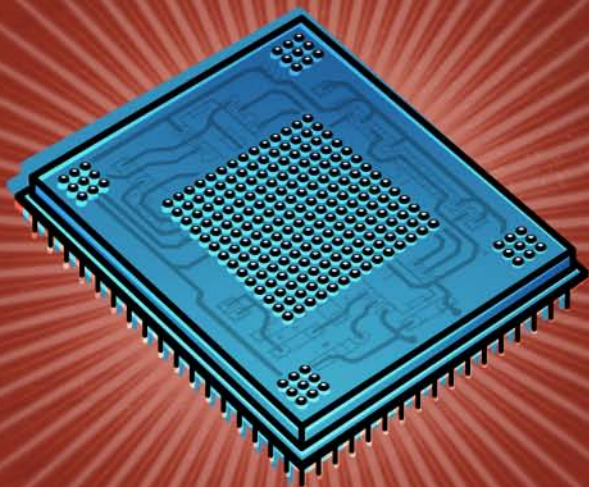
STORAGE

Vol. 8 No. 6 September 2009

THE LOWDOWN ON

Solid-State Storage

Solid-state drives sip power and have lightning fast performance. There's no doubt they'll replace spinning disks—but when? [page 10](#)



ALSO INSIDE

- 5 Do you really know your data?
- 22 How to make your storage more efficient
- 31 Backup apps gain dedupe smarts
- 34 Clearing up cloud-based backup
- 37 Backup still no easy job

STORAGE

sponsors | september 2009

Quantum

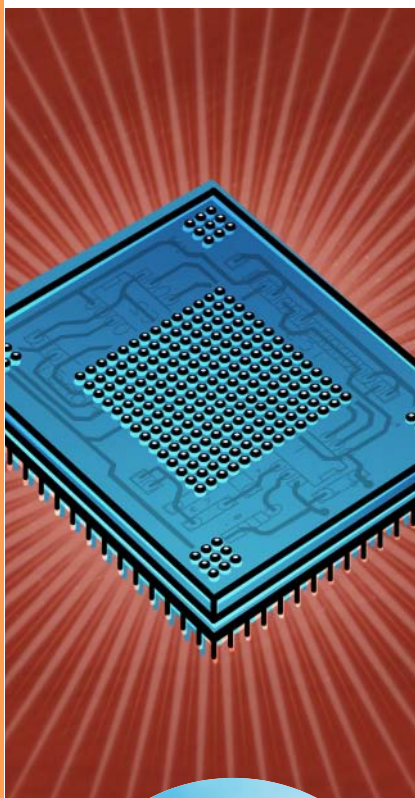
FUJIFILM



maxell



Infotech Ltd



Rule #1: Know Thy Data

- 5 **EDITORIAL** Data classification isn't just for document management systems anymore; it's the key to storage efficiency. Effective storage tiering, data archiving and data protection all depend on knowing the nature of the data you're storing, so the ability to accurately classify that data is the keystone to efficient storage operations. by RICH CASTAGNA

Solid-State Storage Update

- 10 Solid-state storage will eventually nudge spinning disks out of our data centers, but right now it's still mostly for well-heeled shops with power-hungry applications. But with new technology developments addressing some of solid state's shortcomings and larger-scale manufacturing, solid-state storage could come down to earth soon. by JACOB GSOEDL

Use Storage More Efficiently

- 22 Once upon a time budgets were hefty enough to deal with spiraling capacity demands by just adding more iron. But times have changed and getting the most out of what you already have isn't just smart, it might be the only way to keep your storage shop alive and well. by ROBERT L. SCHEIER

Software Vendors Expand Data Deduplication Offerings

- 31 You can add data deduplication to your backup operations with an appliance, virtual tape library (VTL) or target array; however, backup application vendors offer software-only alternatives. by BETH PARISEAU

Is Cloud-Based Backup Right for You?

- 34 **HOT SPOTS** There's been plenty of buzz about cloud-based backup, but most enterprises have shied away from the services. Although putting your data protection in the hands of a company somewhere in the cloud might seem risky, you need to know that there are variations of cloud backup that might be suitable for your company. by LAUREN WHITEHOUSE

Capacity Still Biggest Backup Bugaboo

- 37 **SNAPSHOT** Storage managers are making headway in dealing with their backup pain points, but the ongoing struggle with rising capacities is still headache No. 1, according to our Snapshot survey. How are storage pros dealing with these challenges? More disk, of course. by RICH CASTAGNA

Vendor Resources

- 38 Useful links from our advertisers.





De-boxed in

When it comes to data de-duplication, most companies only offer one kind of solution. But with Quantum, you're in control. Our new DXi7500 offers policy-based de-duplication to let you choose the right de-duplication method for each of your backup jobs. We provide data de-duplication that scales from small sites to the enterprise, all based on a common technology so they can be linked by replication. And our de-duplication solutions integrate easily with tape and encryption to give you everything you need for secure backup and retention. It's this dedication to our customers' range of needs that makes us the smart choice for short-term and long-term data protection. After all, it's your data, and you should get to choose how you protect it.

Find out what Quantum can do for you.
For more information please go to www.quantum.com

Quantum
BACKUP. RECOVERY. ARCHIVE.



Rule #1: Know thy data

Data classification isn't just for document management systems anymore; it's the key to storage efficiency.

FEW YEARS FROM NOW, when you look back on 2009, what will you see? Will it be just a blur of slashed budgets and urgent demands for new capacity, leaving you wondering how your storage shop survived the turmoil? Or will you see it as the turning point that set your shop on the path to greater storage efficiency?

How you reflect on these tough times will depend on your ability to turn misfortune into opportunity.

It's pretty clear that this year's theme is "efficiency." Months ago, I said 2009 was going to be a year of rethinking how we manage our storage. While you've consistently been hearing that you have to do "more with less," the truth is that you have to do more with the same stuff—just better than ever before.

And there are plenty of tools to help you use your storage gear more efficiently. Data deduplication—the poster child for the storage industry for the past two years—is a great way to avoid adding more disk to your backup operation while protecting growing data stores. More and more storage managers are implementing dedupe in their primary storage systems, too. Archivers do a neat job of moving less useful data off primary disk and onto cheaper storage systems. And if you bought or plan to buy new storage arrays this year, there's a good chance you're looking at low-cost, high-capacity systems that can add a tier to your setup and ease the strain on your more expensive primary storage.

These are all very effective ways to make better use of installed storage systems, but they are, in essence, point solutions that tend not to play well with each other. For each one to work well, you need to know the nature of your data and be able to classify that data so you can determine if it should be saved and, if so, where it should go.

While you've consistently been hearing that you have to do "more with less," the truth is that you have to do more with the same stuff—just better than ever before.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

That means, of course, data classification, something that should be on your agenda in lean times and anytime. In some cases, the classification capabilities are fairly sophisticated, with policy creation scenarios and integration with directory services like Microsoft's Active Directory and LDAP. But classification is often limited to the basics like file extensions or time/date stamps. The differences among products' classification features can make it hard (or even impossible) to create consistent data classification policies across multiple services like data archiving and data migration.

So what's missing from this picture? A unified, or federated, data classification method where you can define your corporate, regulatory and maintenance policies once and apply them across the board to all products that move data around your storage environment.

A few years back, several companies brought out data classification products; they managed to form a few alliances with other vendors, but the idea of having a separate box or application to handle classifying data for other applications never caught on. But that was a different time, for sure. Budgets were bulging back then and escalating capacity demands were dealt with by just tossing more iron into the data center.

If the theme is indeed efficiency, wouldn't it be the ultimate in efficiency to create a single set of data disposition policies and apply them across all apps? A consistent approach like that would work for 90% or more of the data you're storing right now—if it's important enough to save in one system, it'll likely be just as important in another.

So what has to happen to reach this kind of data classification nirvana? First, vendors that provide only rudimentary classification criteria in their products need to beef up those capabilities. And, second, vendors have to either create a standard set of APIs that let systems share each other's classification information or they have to create a completely transparent, common classification system that plugs into their apps.

Cynics would say that these things will never happen because the storage industry, while giving lip service to standards, never actually standardizes much at all. But the only way that can change is if you put pressure on them. Storage classification is important right now, and it will get even more important as storage systems become more intelligent—so let your vendors know now that data classification is important to you.

In the meantime, you can create your own internal standard classification system. Make sure the classification and data disposition policies you create for one application are used consistently across all others, too. And don't settle for products that don't provide the level of classification sophistication that your environment requires. ☉

Rich Castagna (rcastagna@storagemagazine.com) is editorial director of the Storage Media Group.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

**Can you spot a data backup problem
before it finds you?**



Fujifilm Tape Environment Analysis. We'll hunt down system problems, saving you time and money...guaranteed!

It's a jungle out there when it comes to today's complex data storage environments. To survive, you can't afford to simply react to data backup problems—you have to proactively identify and correct them, before it's too late. Fujifilm's Tape Environment Analysis is the answer you've been looking for. It's the first vendor- and platform-neutral comprehensive assessment for tape storage systems powered by the ReadVerify™ Appliance. It automatically monitors tape library, tape drive and media performance, and provides detailed, real-time alerting and reporting, to help you maximize existing resources, improve reliability and avoid unnecessary expenditures—today and tomorrow.

Best of all, we guarantee Total R.O.I.! If we can't save you as much or more of the cost of our system-wide analysis over a year—it's free!

For details on this service and our R.O.I. guarantee, visit www.fujifilmusa.com/TEA

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

STORAGE

COMING IN OCTOBER

Guide to Snapshot Technologies

Snapshots are commonly used to enhance backup systems and to dramatically shorten RTOs and RPOs. We'll explain the different types of snapshots; provide guidelines for how frequently snapshots should be done; and describe how snapshots might fit into your overall data protection environment. We'll also look at the key differences in vendors' snapshot offerings.

Find Out What Your Peers Are Buying This Year

For seven years, *Storage* magazine and SearchStorage.com have fielded bi-annual surveys to determine the purchasing plans of storage managers. We report and analyze the results of the latest edition of the survey and provide insight into emerging trends. Find out how other storage managers are stretching tight budgets to cope with growing capacity demands.

Is It Getting Cloudy Inside?

There's been plenty of talk about storage cloud services recently, but for many firms, the technologies related to cloud storage may be more attractive for building internal clouds than for sending data offsite. Internal clouds may offer a practical approach to virtualization that will finally deliver on the promises of utility storage.

And don't miss our monthly columns and commentary, or the results of our Snapshot reader survey.

TechTarget Storage Media Group

STORAGE

Vice President of Editorial
Mark Schlack

Editorial Director
Rich Castagna

Senior Managing Editor
Kim Hefner

Senior Editor
Ellen O'Brien

Creative Director
Maureen Joyce

Editorial Assistant
Rachel Kossman

Contributing Editors
James Damoulakis
Steve Duplessie
Jacob Gsoedl



Site Editor
Ellen O'Brien

Senior News Director
Dave Raffo

Senior News Writer
Beth Pariseau

Features Writer
Carol Sliwa

Senior Managing Editor
Kim Hefner

Associate Site Editor
Chris Griffin



Site Editor
Susan Troy



Site Editor
Andrew Burton

Associate Site Editor
Heather Darcy

Assistant Site Editor
Megan Kellet

Features Writer
Todd Erickson

Executive Editor and Independent Backup Expert
W. Curtis Preston

Storage Decisions

TechTarget Conferences

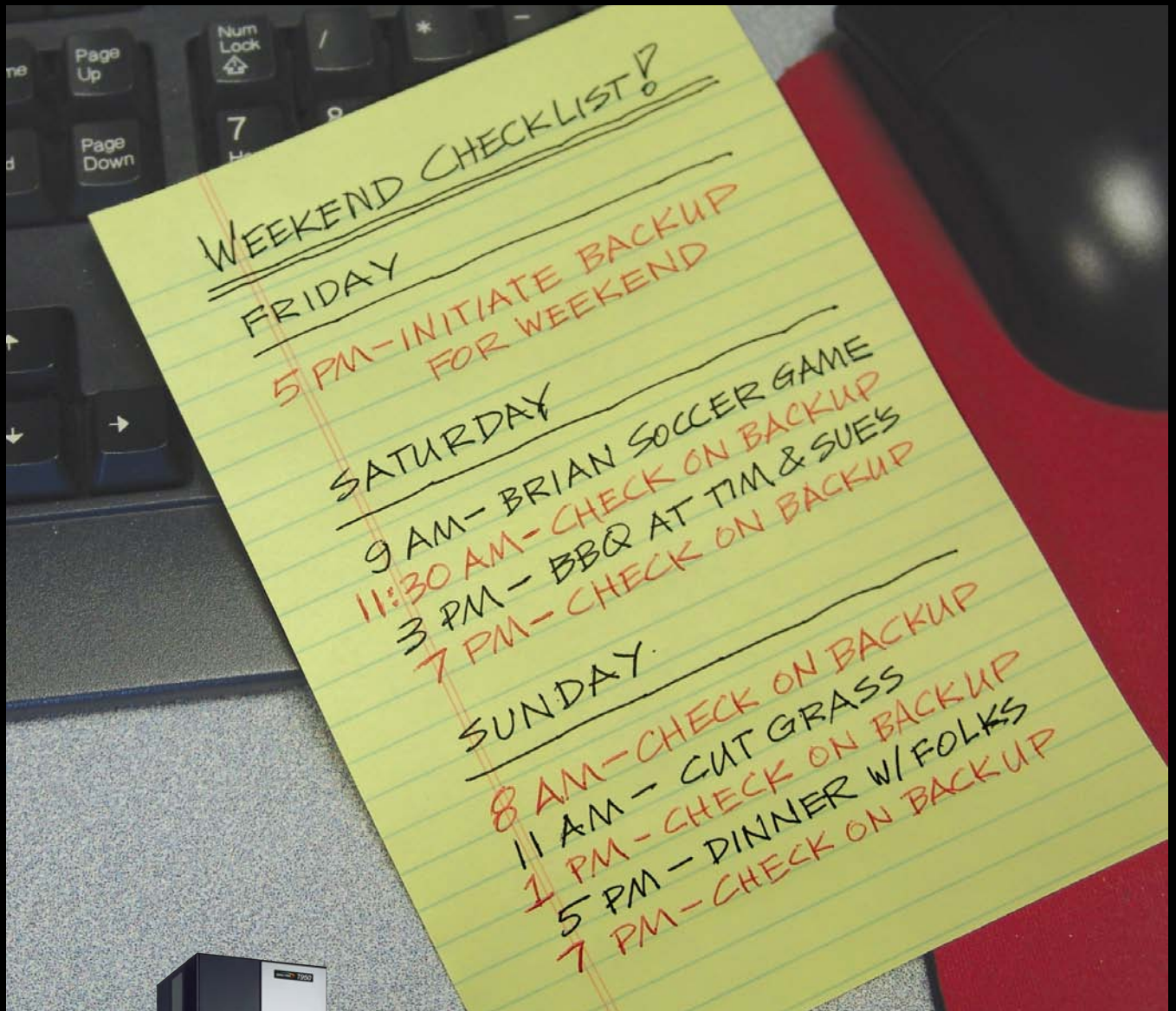
Sr. Editorial Events Manager
Lindsay Mullen

Editorial Events Associate
Nicole Tierney

Storage magazine
117 Kendrick Street,
Suite 800
Needham, MA 02494
editor@storagemagazine.com

Subscriptions:
www.SearchStorage.com

Babysitting your backups?



We're sure you have a list of things you want to do over your weekend, and one of them isn't watching your backups. Spectra's Tape Without Pain™ solutions include the latest tape and disk technologies and features that ensure you can back up with confidence. With high capacity, fast transfer rates plus advanced media and hardware health monitoring, you can be sure your backups get done in time — and without all of the hand-holding you are used to. Now you just might be able to squeeze in that round of golf between those soccer games and barbeques.



What is Tape Without Pain?



Click here to watch our Presentation...

Capacity is biggest backup issue

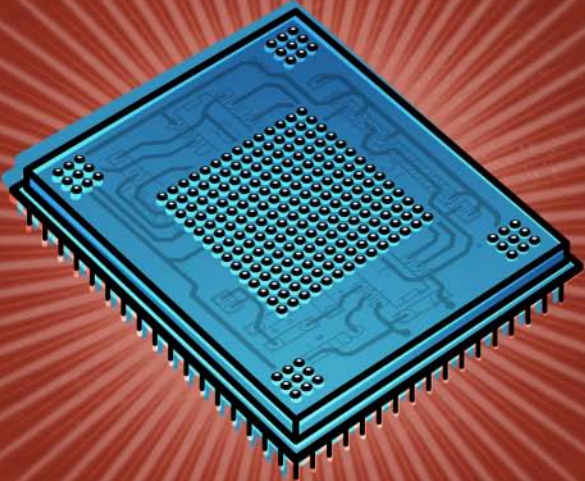
Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Solid-state STORAGE UPDATE



Solid-state storage is still mostly for well-heeled shops with power-hungry apps, but new developments could bring solid state down to earth soon.

By Jacob Gsoedl

S

OLID-STATE STORAGE received a big boost in 2009, with a large majority of storage vendors adding solid-state drive (SSD) options to their product lists. As a result, we've seen a sharp increase in the total number of enterprise-grade SSD components traded. A meager 59,000 units were sold worldwide in 2008, according to Stamford, Conn.-based Gartner Inc., but the total is expected to reach 5.1 million units and \$2 billion in revenue by 2013. Although the price for NAND flash has come down by

approximately 30% since last year—with expectations that it will continue to decline annually at that rate—it's still an order of magnitude more expensive than high-end disk drives. Because of its premium price, customers continue to deploy NAND flash judiciously, mostly for applications adverse to latency and requiring a high number of IOPS; in the past, expensive bulky arrays with a large number of spindles were the only alternative.

That's where solid-state storage shines today: A single enclosure of SSDs can displace a rack of high-end Fibre Channel (FC) drives at an overall lower cost, provide better performance, require significantly less power and space, and greatly reduce data center and operational complexity. Solid-state drives can also supplement disk arrays with a small amount of solid-state storage for frequently accessed data to boost array and application performance. OLTP apps like SAP and Oracle ERP, databases, email servers, high-transaction websites and even virtualization platforms are the great beneficiaries of solid-state storage. Whenever hard disk I/O and latency become the limiting factors, solid-state storage is an alternative. Conversely, and as a result of the high per-gigabyte cost of SSD, whenever large capacity is needed, hard disks continue to be the storage medium of choice.

Even though solid-state storage can be implemented with DRAM, NAND flash and other memory technologies, NAND flash is the prevailing solid-state drive memory technology in use today. In addition to non-volatile memory, enterprise-grade SSD products typically come with a small amount of DRAM that acts as write-buffer and cache, a controller with storage interfaces (FC, SATA or SAS) and software. Today, it's mostly the intelligence and proprietary algorithms in controllers that overcome the limitations of NAND flash, making it viable in the enterprise space. "Because of its better controller technology and algorithms to manage NAND flash, STEC [Inc.] has by far the largest number of design wins in the enterprise storage space today," said Joe Unsworth, research director in Gartner's Technology and Service Provider Group.

A single enclosure of SSDs can displace a rack of high-end Fibre Channel (FC) drives at an overall lower cost, provide better performance, require significantly less power and space, and greatly reduce data center and operational complexity.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

CHALLENGES OF NAND FLASH AND SSDs

Unlike DRAM, NAND flash is non-volatile, capable of retaining data as hard disks do, but without depending on vulnerable mechanical parts and requiring significantly less power. But these benefits are offset by shortcomings the storage industry has been attempting to address for several years:

- Durability issues with NAND flash
- Low write performance of NAND flash
- Inadequate software to efficiently support solid-state drives
- Architectural shortcomings of storage systems that have been designed for mechanical disks

DURABILITY ISSUES OF NAND FLASH

The most severe issue with NAND flash is the wear-out of cells, which limits a cell's life span to a very finite number of writes. While consumer-grade multi-level cell (MLC) flash permits approximately 10,000 writes per cell, enterprise-grade single-level cell (SLC) flash supports about 100,000 writes per cell before becoming unusable. The wear-out problem worsens as density increases. The 10,000 number of supported writes of 2-bit per cell MLC flash used in consumer-grade products looks generous vs. the newer 3-bit per cell offerings with their 1,000 to 5,000 supported write cycles and the few hundred supported writes of 4-bit per cell flash. The data storage industry has been contending with this simple rule of NAND flash: as density increases, both cost and durability decrease.

The most severe issue with NAND flash is the wear-out of cells, which limits a cell's life span to a very finite number of writes.

With SLC NAND flash now capable of meeting enterprise storage requirements and accepted in the enterprise space, storage vendors are trying to further decrease its cost by bringing MLC flash into the enterprise realm. Specifically, they're looking to use 2-bit per cell multi-level cell flash to compete with single-level cell flash, and 3-bit and 4-bit per cell flash for read-intense applications with scant write requirements, such as data archival.

"It's not a question if MLC flash can be used in enterprise storage systems, but a question of what it takes to make it happen at an acceptable cost," said Mark Peters, an analyst at Enterprise Strategy Group (ESG) in Milford, Mass. There are already a few instances where MLC flash is used in the enterprise space today. The most prominent example is the Hewlett-Packard (HP) Co. StorageWorks IO Accelerator for HP BladeSystem c-Class, a direct-attached, solid-state storage array mezzanine card; the HP product is based on Fusion-io's ioDrive, and

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

uses both SLC and MLC flash depending on capacity.

Enterprise-grade solid-state drive vendors have employed a variety of techniques that enable their products to match and even exceed the life span and durability of mechanical disk drives. With SSD drives warranted for three to five years—depending on the SSD vendor—and mean time between failures (MTBF) north of 1 million, enterprise-grade SSD drives are at least as durable as high-end disk drives.

“By now, we consider SSD drives as reliable as high-end FC drives,” noted Claus Mikkelsen, chief technology officer, storage architectures at Hitachi Data Systems. To attain this degree of durability, sophisticated wear-level algorithms that reduce the number of writes and distribute writes evenly among flash cells have been devised and implemented in solid-state drive controllers. The use of spare capacity, which typically ranges from 20% to 100% of usable capacity, extends the life span of SSDs by reducing the number of times cells are written to during a given time period and providing the extra capacity to replace defunct cells. Compression and data deduplication algorithms are used to maximize efficiency and reduce the number of writes per cell. And similar to high-end mechanical disks, enhanced error-correction algorithms are used to find, fix and isolate bad blocks. “Error-correction codes [ECCs] used to occupy four or five bits per 512 byte block; now six to eight bits are common and we’re seeing it move to 12 bits,” Gartner’s Unsworth explained.

LOW WRITE PERFORMANCE OF NAND FLASH

The other severe handicap of NAND flash is its slanted read-write performance ratio (see “[NAND flash solid-state drives vs. disk](#),” below). While enterprise SSDs are capable of delivering a sustained read performance greater than 40,000 I/Os per second, write performance typi-

NAND FLASH SOLID-STATE DRIVES vs. DISK

	NAND flash SSD	Disk drive
I/O per second (sustained)	Read: 45,000+ Write: 15,000+	Few hundreds
Latency in milliseconds	Read: 0.2+ Write: < 1	4+
Cost/GB	High	Low
Cost/IOPS	Low	High
Resilience	High	Lower, because of mechanical components
Power consumption	Low	Higher

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

cally lags by a factor of three or four. This discrepancy is caused by NAND flash's requirement to erase blocks before they can be written, adding substantial overhead. That's also the reason why NAND flash storage shows significantly higher write performance as long as erased cells are available, but declines by a factor of two to three thereafter.

"Since NAND flash-based SSD products show great write performance the first 15 to 20 minutes, it's pertinent to compare their sustainable performance rather than their inflated burst performance," cautioned Woody Hutsell, president at Texas Memory Systems.

The STEC Zeus IOPS solid-state drive with its maximum 52,000 sustainable read IOPS and 17,000 write IOPS, according to the company, currently dominates in the enterprise storage space and has established a baseline that other SSD offerings are compared to. "With the latest STEC drives, write performance is enterprise ready, but clearly not on par with read performance," said Kyle Fitze, marketing director, HP StorageWorks Storage Platforms Division. Unfortunately, no independent third-party tests for enterprise-grade SSD products are available at this point and, as a result, performance numbers cited by vendors should be taken with a grain of salt.

To overcome the read-write performance gap, most vendors are deploying a small DRAM cache that acts as a write buffer; that is, data is written to the cache first and then to NAND flash. "A DRAM write buffer doesn't quite get you to read performance, but it gets you closer," noted Clod Barrera, chief technical strategist for IBM System Storage. Even though DRAM helps close the gap, it adds the complexity of having to back up the volatile data in cache in case of a power failure and consumes valuable real estate. "As we move to 1.8-inch drives and custom form factors, you want to enable the highest density, and DRAM clearly is prohibitive," said Thad Omura, vice president of marketing at SandForce Inc., a developer of SSD processors.

Because of these shortcomings and the inability to completely make up for the write handicap of NAND flash, newer and more innovative solid-state drive developments shun the DRAM write buffer approach. Pliant Technology Inc. claims to achieve more than 100,000 IOPS for both reads and writes. Commercially available, the Texas Memory Systems

"Since NAND flash-based SSD products show great write performance the first 15 to 20 minutes, it's pertinent to compare their sustainable performance rather than their inflated burst performance."

—Woody Hutsell, president,
Texas Memory Systems

maxell

**IS YOUR
INFORMATION
AT RISK?**



**LTO Cartridge Memory
Analyzer knows!**

Click For Free White Paper:
Detect Faulty Storage Media Before It's Too Late.

RamSan-620 is capable of sustaining 250,000 IOPS for both reads and writes, according to the firm's Hutsell. Similar to the RamSan-620 and Pliant Technology's Enterprise Flash Drives (EFDs), SandForce's SF-1000 family of SSD processors, which interface to both MLC and SLC drives, forego a DRAM write buffer. All three vendors emphasize the role of parallelization as key to overcoming the write performance gap and an overall increase in the number of supported IOPS.

"It's our custom parallel-processor architected ASIC that enables us to perform many of the housekeeping tasks, such as pre-erasing of unused blocks, concurrently, and it enables us to get write performance

GLOSSARY OF SSD terms

➔ **Solid-state drives (SSDs):** SSDs use memory chips, mostly non-volatile NAND flash, instead of rotating platters for data storage. The benefits of low latency, low power consumption and higher resilience compared to disk drives are a result of not having any mechanical parts.

➔ **Flash memory:** Flash is non-volatile, rewritable memory. Unlike DRAM, it requires erasing blocks of data before they can be written to, resulting in a lower write than read performance. Depending on the technology, flash memory supports only a finite number of writes. Although flash memory is available as NOR or NAND flash, SSD products use NAND flash because it's more durable, less expensive, its cells are denser, and writing and erasing are quicker compared to NOR flash.

➔ **Single-Level Cell (SLC):** SLC NAND flash stores one bit per cell. Because of its high endurance (approximately 100,000 writes per cell) and cost, SLC is predominantly used in enterprise-grade SSD offerings.

➔ **Multi-Level Cell (MLC):** MLC NAND flash uses two bits per cell. With about one-tenth of the endurance of SLC NAND flash and a fraction of the cost of SLC flash, MLC is mostly used in consumer products. Newer 3-bit per cell (1,000 to 5,000 supported writes) and 4-bit per cell (a few hundred supported writes) NAND flash are targeted for applications with a very limited number of writes.

in line with read performance,” explained Greg Goelz, vice president of marketing at Pliant Technology.

INADEQUATE SOFTWARE SUPPORT FOR SSD

While significant progress has been made to overcome or at least mitigate the issues related to NAND flash, software support to manage and efficiently take advantage of solid-state storage has evolved at a much slower pace, becoming one of the primary obstacles to more rapid enterprise adoption of SSDs. To counteract the prohibitive effect of the high price of solid-state drives, storage systems need to be able to maximize the use of SSDs by automatically and transparently shuffling data between the fast SSD tier and slower disk tiers. While most storage vendors acknowledge the need and relevance of policy-based data migration between the fast but expensive SSD tier and disk tiers, keeping frequently accessed data in solid-state storage and more static data on disks, only a few can offer an automated solution today.

Leading the pack is Compellent Technologies Inc.’s Storage Center storage-area network (SAN). Its Dynamic Block Architecture tracks the characteristics and usage of every data block; this metadata information is leveraged by the product’s Data Progression feature, which automatically moves data from SSDs to disk tiers and vice versa based on how often blocks are accessed.

“Our Data Progression is the killer app for SSD because users can add drives to existing systems and then let automation take over,” said Bob Fine, Compellent’s director of product marketing. Contrary to Compellent, the majority of enterprise storage vendors depend on a manual two-way process for migrating data between solid-state drives and disk tiers, first analyzing I/O activity and, in a second step, migrating data to the appropriate tier. Depending on a manual process for now, EMC Corp. has announced Fully Automated Storage Tiering (FAST), which will be available for EMC’s Symmetrix V-Max systems later this year. FAST will automate the movement of data across multiple storage tiers based on business policies, predictive models and real-time access patterns. IBM supports automatic data migration to SSD via its Data Facility Storage Management Subsystem (DFSMS), but it’s only available on the mainframe z/OS platform with DS8000 storage, with a manual two-way process still required for other systems.

Both Sun Microsystems Inc.’s Sun Storage 7000 Unified Storage

"Our Data Progression is the killer app for SSD because users can add drives to existing systems and then let automation take over."

—Bob Fine, director of product marketing, Compellent

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Systems and NetApp Inc. filers with Performance Acceleration Modules (PAM) circumvented and solved the software challenge at a storage architecture level by using NAND flash as cache rather than as disk replacements. As a result, SSD is closely woven into their storage architectures and firmware, with the advantage that all data and apps benefit from solid-state drives, eliminating the requirement to shuffle data between tiers. “We want our Storage 7000 customers to have all of their working data in flash,” said Michael Cornwell, Sun’s lead technologist for flash memory.

SSD ARCHITECTURES

Contemporary storage systems have been designed to cope with the limitations of mechanical disk drives, in particular to reduce the impact of high latency and the low number of IOPS mechanical disks can support. With SSDs, this basic truth has changed and capacity limitations of storage controllers have become the limiting factor. Simply replacing disk drives with SSDs can overwhelm storage systems if too many solid-state drives are added. “Storage controllers are just starting to adjust to the new performance requirements of SSD, and today customers need to heed the recommendations and guidelines of storage vendors on how many SSDs they can add,” said Greg Schulz, founder and senior analyst at Stillwater, Minn.-based StorageIO Group.

There are currently four methods to complement storage systems with solid-state storage:

1. Adding SSD drives in lieu of disk drives
2. The use of NAND flash as cache in storage controllers
3. The use of NAND flash on servers rather than storage controllers
4. Standalone SSD arrays

Adding SSD drives in lieu of disk drives. Adding SSD drives via Fibre Channel, SATA or SAS interfaces to replace disk drives is the easiest and most popular way of adding solid-state drive support to existing arrays. Notwithstanding rigorous testing and qualification procedures, this approach requires few if any changes to storage systems because vendors can leverage what’s in place. The lack of automated data migration between SSD and disk tiers, and performance limitations of contemporary storage controllers, are the two biggest drawbacks of this approach. Nevertheless, it’s the method adopted by most storage vendors. EMC has been joined by Compellent, Fujitsu, HP, Hitachi Data Systems, IBM, LSI Corp., Pillar Data Systems, Sun and many smaller array vendors, offering SSD drives in addition to hard disks for some of their arrays. The overwhelming majority of these vendors have been using STEC drives as their first generation SSDs, largely because STEC was the first vendor capable of meeting enterprise storage requirements. With disk

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

drive vendors like Seagate Technology LLC, promising startups like Pliant Technology and SandForce, and Intel Corp. targeting the enterprise storage space, STEC's predominance will be challenged.

The use of NAND flash as cache in storage controllers. NetApp and Sun are leveraging NAND flash as cache. By doing so, both vendors have overcome the software issue of automated data migration between SSD and disk tiers, but they have changed their storage architectures to embrace NAND flash, eliminating the possibility of overwhelming their arrays if too much solid-state drive storage is added. By front-ending disk drives with NAND flash instead of replacing disk drives, all data and apps benefit from SSD, not only data that resides within the SSD tier.

NetApp offers the Performance Acceleration Module (PAM), which can be added to any NetApp filer with available PCI Express slots. Depending on the controller, up to five modules can be installed for a unified cache as large as 80 GB today and up to 512 GB later in the year when a higher density PAM card will become available. PAM is used to cache metadata only. "By storing a copy of the metadata in flash memory on the storage controller, we're seeing a 30% to 50% performance gain for typical workloads," said Patrick Rogers, vice president, solutions marketing at NetApp. "Filers with PAM and SATA drives have become a viable alternative, replacing filers with FC drives, because of comparable performance at a significantly lower cost," he said.

Unlike NetApp, Sun uses flash memory in its Sun Storage 7000 Unified Storage Systems to cache all reads and writes—not only metadata—and therefore has one of the most advanced architectures to support flash memory.

The Sun Storage 7000 Unified Storage Systems run Solaris on an x86 platform with an optimized storage stack and the Zettabyte File System (ZFS) that supports a Hybrid Storage Pool of DRAM cache, SSD and mechanical disks. The solid-state drive is situated between the DRAM-based Adaptive Replacement Cache (ARC) and SATA drives. The ZFS Intent Log (ZIL), which holds the write journal to allow the file system to recover from system failures, is written to a write-optimized SSD. The L2ARC cache comprises read-optimized SSDs to extend the DRAM-based ARC cache for read operations; L2ARC can be hundreds of gigabytes in size, and its purpose is to keep working data in memory to minimize disk access. This Hybrid Storage Pool enables the Sun Storage

"By storing a copy of the metadata in flash memory on the storage controller, we're seeing a 30% to 50% performance gain for typical workloads."

—Patrick Rogers, vice president, solutions marketing, NetApp

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

7000 Unified Storage Systems to support more than 800,000 IOPS, according to Sun.

The use of NAND flash on servers rather than storage controllers. Although the Sun Storage 7000 Unified Storage Systems is a standalone storage system, it makes the point for those who argue that flash memory belongs in the server rather than the storage controller. “Just like L2 cache extends memory on the CPU and DRAM extends L2 cache, flash memory is intended to extend DRAM,” explained David Flynn, chief technology officer at Fusion-io. The Fusion-io ioDrive and ioDrive Duo NAND flash PCI Express cards provide direct-attached storage (DAS) for servers. Being a server company that also sells storage, Sun concurs that servers are the right place for flash memory. “Flash memory is a game-changer for server architectures, and next-generation servers will extend DRAM caches with flash memory,” Sun’s Cornwell said.

Standalone SSD arrays. Complementing disk arrays with SSD-based storage systems that run parallel to traditional storage arrays is the least-disruptive method of adding solid-state storage to a storage environment. The leading vendor of standalone SSD arrays is Texas Memory Systems. Offering both DRAM and NAND flash-based SSD arrays, the company sells its RamSan family of products directly and through OEM relationships with BlueArc Corp., NetApp and others. On the downside, standalone solid-state systems aren’t able to leverage existing array components and are therefore likely to be more expensive. Moreover, they’re less integrated with the disk tier than other architectural approaches, making it even more difficult to overcome the data migration challenge between the solid-state drive and disk tiers.

SOLID-STATE OUTLOOK

Solid-state storage has just begun to play a role in enterprise-level systems, but it’s apparent that its rise is unstoppable. Enterprise storage systems are moving toward two-tier architectures, namely, a solid-state drive tier for transactional and changing data, and a large capacity SATA disk tier for more static data. With the continuous innovation that has overcome some of the limitations of NAND flash, as well as newer memory technologies like magnetoresistive random access memory (MRAM) on the horizon to eventually replace NAND flash, the real challenge to rapid adoption of SSDs is the lack of storage architectures that are capable of seamlessly integrating and efficiently taking advantage of solid-state drives. ☉

Jacob Gsoedl is a freelance writer and a corporate director for business systems. He can be reached at jgsoedl@yahoo.com.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

**Compliment or Replacement
to Tape Backups**

Exchange Recovery

Rapid Failover Server

**Bare Metal Restore
to Dissimilar Hardware**

15 Minute Snapshots

Offsite Replication



It's time for IT departments to have an affordable solution that provides a simple, reliable and flexible way to keep their business running at full speed. That's why Zenith Infotech has created the ARCA.

For IT departments of all sizes, the Zenith ARCA can be a lifesaver if there is a catastrophic server failure. But what about the day-to-day issues that have you scouring disk and tape backups for a deleted file? With the Zenith ARCA, time spent goes from days and hours to minutes and seconds.

Backup • Recover • Restore

See how at www.ZenithARCA.com/OnDemand

Learn more at www.ZenithARCA.com/Learn

Use storage more efficiently

Getting the most out of what you already have isn't just smart, it might be the only way to keep your storage shop alive and well.

By Robert L. Scheier

DATA STORAGE EFFICIENCY is often an elusive target for storage managers, but with declining budgets and unchecked capacity demands, getting the most out of your storage resources is more important than ever. Storage vendors say they feel your pain and can help you store more data for less cost than ever before.



Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

But storage managers should press those vendors on the actual savings they'll see with their own data and environments. It's also important to understand that some space-saving technologies may require new monitoring tools or additional storage controllers to ensure they don't crash important applications. In addition, a new technology might not always be the answer; non-technical methods, such as pushing back on users' demands, may do as much to help keep storage budgets under control (see "[Just say 'No,'](#)" [this page](#)).

The amount of storage under management is growing approximately 40% per year, according to a March 2009 survey of 400 U.S. IT professionals commissioned by Symantec Corp. However, the survey indicated that storage budgets will grow only 15% to 20% in the next year, and 20% to 24% over the next two years. If Symantec's survey paints a not-so-pretty picture of the coming years, *Storage* magazine's own Purchasing Intentions survey indicates that the pressure is on right now. In that survey, also conducted in March, respondents said they'll add an average of 43 TB of new disk capacity this year, despite seeing their storage budgets dip by 1.9% vs. last year. This crunch is leading many organizations to consider using newer technologies, many made possible by the growing use of storage virtualization, to keep more data available while spending less money.

JUST SAY 'NO'

Need to cut your storage requirements but don't have the time or money for fancy new hardware, software or consultants? Then try pressing your users on whether they still need all the disk space, copies or performance they insisted on when times were good, said Roger Cox, research vice president at Stamford, Conn.-based Gartner Inc.

"When money was no problem, people overprovisioned their storage," he said. If a user said an application needed half a terabyte, the storage administrator would provision a full terabyte so they wouldn't have to go through the painful process of expanding the volume if the app actually needed more. "Go back, revisit your provisioning policies" and cut back on some of that overprovisioning, Cox said.

Consider asking users to reduce the number of copies they require, or see if they can live with somewhat slower performance by crowding more data onto each drive. And if users insist they can't give anything up? Try charging them for excess copies, overprovisioning or gold-plated performance. With their own budgets strapped, that should quickly help impose some self-discipline, Cox suggested.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Here are some tips for using these new tools most effectively (see “[Quick list: Efficiency options](#),” p. 29).

Data archiving

Moving less important data to slower, and less expensive, storage media is a mature and well-established cost-cutting technique, with more than 75% of respondents to Symantec’s survey archiving email and files or planning to do so. What’s new is that many are combining archiving with data deduplication and other techniques for greater efficiency.

At Health Alliance Plan of Michigan in Detroit, Dan Trim’s storage budget has been falling for the last eight years, but data archiving has cut storage growth rates from as high as 48% down to a much more manageable 14%. Trim, the health insurer’s director of infrastructure technology, was pleased with how Symantec’s Veritas Command-Central Storage, among other tools, gave him “a deeper look” into how he was using his storage and “how to back up files over a certain age, get them off to tape, offsite and off my disk,” he said.

Perry Fritz, enterprise operations manager at Rockline Industries Inc. in Sheboygan, Wis., estimates he has saved approximately \$20,000 in disk arrays since he started archiving the company’s email files with managed services provider Mimecast in June 2008. The paper goods manufacturer moved into archiving not to reduce its storage needs, but to speed email retrieval for legal or other purposes. It chose a hosted service because it was less expensive than purchasing archiving hardware for each of its three sites, and provided a central archive point for its 200 GB to 300 GB of email data. The savings on disk purchases, along with the disaster recovery (DR) capabilities, has been one of the “added benefits” of going with a managed storage service.

Compression and data deduplication

Compression is one of the oldest methods for saving space and data deduplication is one of the newest, but they’re related and each has a role to play in holding down storage spending. Understanding how the technologies differ is the key to using each one most effectively.

Compression uses mathematical algorithms to simplify large or repetitive parts of a file, with different compression products aimed at different use cases and various types of files. Some storage shops use the compression capabilities built into popular operating systems such as Unix, or even low-cost utilities such as WinZip on Windows platforms. Later this year, NetApp will release compression features “covering all the platforms we now cover, including primary storage,” said Chris Cummings, NetApp’s senior director of data protection solutions.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Data deduplication eliminates duplicate patterns within a data store, and in ideal cases—such as repeated backups of almost identical files—vendors claim they can reduce data sets by 15:1 to 20:1. It's no wonder that 95% of respondents to the Symantec survey are at least discussing data deduplication, with 52% either implementing or having implemented it.

However, deduplication works best on data to which only minor changes are made over time (such as backups of lengthy business documents or engineering plans) rather than data of which only one copy exists, such as a CAT scan stored on a medical system. By contrast, most compression delivers at least 2:1 compression ratios “on almost any data set,” sometimes with little or no performance hit, said George Crump, lead analyst at storage consulting firm Storage Switzerland.

In fact, Crump said, deduplication “loses value the closer it gets to primary storage” where there are usually fewer multiple copies. To prevent dedupe from slowing down disk access on primary storage, the deduplication would have to be done after the data arrives on disk, added Andrew Reichman, a senior analyst at Cambridge, Mass.-based Forrester Research Inc. “This will require swap space to write data un-deduplicated and then deduplicate it to a separate set of disk,” he said. This “could eliminate the capacity reduction,” he added, which is the whole point of deduplication.

Health Alliance Plan's Trim said he's seeing approximately a 50% savings in storage capacity with Symantec's Veritas NetBackup PureDisk.

Different vendors squabble over just where and how to use dedupe. Symantec, for one, is pushing a “dedupe everywhere” strategy, while NetApp's Cummings said he doesn't recommend it “for your tier 1, highly transactional, high IOPS database environment. But we do see it as being safe and having little or no performance impact” for storing virtual servers, tier 2 databases, file services and archiving.

For Chris Watkis, IT director at Grey Healthcare Group Inc. in New York City, data deduplication was an unexpected benefit from his purchase of a FalconStor Software Inc. Virtual Tape Library in 2007. His main goal was to speed backups and restores as the medical marketing company moved into more markets, created bulkier content such as video and held onto that content for longer periods.

But the deduplication reports from the FalconStor appliance showed how much redundant data was being eliminated before backup. Armed with that information and some off-the-shelf storage management tools, Watkis now regularly scans his servers for redundant files and has recovered 40% of the space on his 16 TB storage-area network (SAN). Those savings are critical since he had to cut his storage budget by 20% in fiscal 2009.

Joseph Stedler is senior engineer and Dallas data center manager at External IT USA Inc., a managed service provider based in Richardson,

Managing the information that drives the enterprise

STORAGE

The Essential Guide to Data Deduplication

data deduplication

Data dedupe can reduce the amount of disk required for backups by removing redundant data, but there are a few things you need to know before implementing this technology.

Exclusive Extra for Storage Subscribers

The Essential Guide to Data Deduplication is a brand new expert resource brought to you by the Storage magazine team. This guide outlines the fundamental aspects you need to know about data dedupe and provides a number of tips and best practices from some of the leading experts in the field. Get a handle on data deduplication with 24/7 access to this new Essential Guide in an easy-to-access, easy-to-read, easy-to-print online format.

Get Your FREE Copy Today: www.SearchDataBackup.com/Dedupe_Guide

Sponsored by:



Texas. For data deduplication, he considered both a Data Domain Inc. appliance and Veeam Software's Veeam Backup & Replication software, which combines deduplication and backup for VMware ESX virtual servers. Despite its much higher cost, Stedler said, he went with the hardware-based Data Domain appliance because of its higher performance and the ability to replicate data among his various data centers.

The host-based deduplication provided by EMC Corp.'s Avamar has received "a lot of favorable feedback" from Gartner customers, said Roger Cox, research vice president at Stamford, Conn.-based Gartner Inc., as have target-side deduplication products from Data Domain, Diligent Technologies Inc. (which was purchased by IBM Corp. last year) and Quantum Corp. (which is OEMed by EMC).

David Floyer, chief technology officer and co-founder of the Wikibon project, an online IT support community, said NetApp's data dedupe is aimed at primary, rather than backup, storage as is the case with other vendors such as Data Domain, and can save 30% in the cost of disk. In calculating total ROI, however, Floyer warns that because the data must still be "rehydrated" to its original state before being used, customers still need enough storage controllers to assure the proper IO and bandwidth for critical applications. This lowers the overall cost savings, he said, from 30% to 15%.

Users need to look out not only for overall reduction ratios, but for how long it takes a product to compress or deduplicate data, and then return it to its original, readable state, said Greg Schulz, founder and senior analyst at StorageIO Group in Stillwater, Minn. Saving huge amounts of space isn't much good if you can no longer work within your backup or restore windows, he noted.

Thin provisioning

To use an airline analogy, thin provisioning lets storage admins "overbook" disk space by provisioning more disk space than an application is likely to use. The disk space is only actually occupied when the application writes data, leaving the unused capacity available to other applications, rather than sitting allocated but unused.

Just as the gate agent must keep a close eye on how many passengers actually show up, storage administrators need real-time monitoring tools so they know when to add more physical disk or expand the size of logical volumes if too much data shows up for "seats" on the array.

Using thin provisioning, External IT USA's Stedler has allocated 215 TB of space for his VMware servers, backed by only 5.5 TB worth of actual disk. He said he's very pleased with how DataCore Software Corp.'s SANsymphony provides real-time reports about "how much storage has been allocated ... vs. how much has been claimed by a virtual machine," and how he can set thresholds for alerts when drives reach a certain utilization level.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Forrester Research's Reichman said NetApp has "decent capacity visibility with their Operations Manager tool that's designed to report on NetApp-specific storage," but faulted the company's SANscreen for lacking visibility to the file-system level. He also faulted EMC's Ionix ControlCenter and IBM's Tivoli Storage Productivity Center Suite as lacking the detailed reporting needed to support thin provisioning, and said "none of the big SRM [storage resource management] tools seem to have hit the nail on the head ... and I think it's a big reason why users are slow to adopt thin provisioning."

Reichman and other analysts singled out smaller companies such as Compellent Technologies Inc. and 3PAR Inc. as providing such detail on their own platforms, helping drive faster adoption for them.

On the higher end, Gartner's Cox said he's gotten "good feedback" from customers using Hitachi Data Systems' Universal Storage Platform V and Universal Storage Platform VM storage controllers. In general, customers are reluctant to risk the availability of their high-end systems with new technologies such as thin provisioning, Cox said. "And in the Oracle world, where a lot of the high-end systems are installed," he said, the popular database precludes the use of thin provisioning because it pre-formats the disk pools itself.

StorageIO Group's Schulz recommends factoring in the cost of monitoring tools when calculating the benefits of thin provisioning, and warns against creating performance bottlenecks by forcing too many servers to compete for read/write access to too few disks.

Tiered storage

Moving older or less valuable data to slower, less-expensive media is another time-honored way to reduce costs. Comparatively low-cost SATA drives can cost as little as one-tenth the price of high-performance Fibre Channel (FC) drives, while using far less power and offering much greater density than FC drives, Forrester Research's Reichman said. SATA drives can be particularly effective when used with other features such as thin provisioning and wide striping.

Storage Switzerland's Crump advises not getting bogged down in a tiered storage strategy. "Take the oldest data you have and just move it," he said. "If nobody has accessed it in two years, the chances of anyone accessing it again are between slim to none." By avoiding the cost of designing a more formal policy, he said, "you can handle the [rare] time when someone" actually needs an older file.

Within the next 12 months, Gartner's Cox said, vendors will start allowing customers to move not just volumes but individual pages within a data set to slower, less-expensive storage as those pages become less important or go longer without being accessed. Automated tiered storage as this "sub-volume" level, he said, "is going to have a big impact on storage efficiency."

Wide striping

Wide striping is a variation on RAID in which data is distributed among multiple disks, using only a relatively small amount of the capacity of each disk to maximize its performance. Wide striping is particularly effective as a cost-saver when used with relatively low-cost SATA drives, Forrester Research's Reichman said, compared to using higher priced Fibre Channel drives to deliver the needed performance.

"3PAR, Compellent and NetApp have all been doing wide striping for some time, and all of them are claiming the ability to derive high performance from SATA drives," he said. EMC and Hitachi Data Systems also provided disk pooling and wide striping when they released thin provisioning capabilities several years back, he said, but don't stress the use of SATA disks as much as other vendors.

In some "wide striping" solutions, the most frequently accessed data is stored automatically on the outer tracks of the disk so it's accessed most quickly, with other less-frequently used data stored

QUICK LIST: EFFICIENCY OPTIONS

These technologies can help you run a more efficient storage shop by making better use of your already installed capacity.

Efficiency technology	What it does	How you save
Archiving	Moves older data out of production data stores	<ul style="list-style-type: none"> • Frees expensive primary disk space • Back up less data • Speeds applications
Compression and data deduplication	Reduces amount of data stored on disk	<ul style="list-style-type: none"> • Redundant files eliminated • Individual files take up less space • Only unique data is stored • Backups require much less disk space
Thin provisioning	Let's you "oversubscribe" your installed disk capacity	<ul style="list-style-type: none"> • Avoids wasteful unused space in allocated volumes • Pools available installed capacity so that it's used more efficiently with higher utilization rates • Delays new disk purchases
Tiered storage	Moves less valuable data to cheaper, lower performance disk	<ul style="list-style-type: none"> • Takes less-demanding apps' data off primary storage to avoid/delay buying more tier 1 disk • Preserves high-performance storage for apps that require it • Reduces backups by not backing up static data as frequently
Wide striping	Spreads data out across more disks	<ul style="list-style-type: none"> • Gets better performance out of inexpensive disks • Avoids or delays buying expensive higher performance disk storage

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

elsewhere to make the most use of all the available capacity.

Rob DiStefano, IT systems manager at Earth Rangers Foundation, a Woodbridge, Ontario, non-profit organization, used this capability in the company's Pillar Data Systems' Pillar Axiom 600 to boost disk utilization to 80% vs. only 40% on older network-attached storage (NAS). Using Pillar's drag-and-drop interface, DiStefano said, he was also able to reduce his administration costs by a factor of 10.

YOUR MILEAGE WILL VARY

As with all products, vendors will choose the ideal scenario for calculating how much storage they can save. In real life—in your shop—efficiency efforts may yield somewhat less-impressive results.

In his Dallas data center, External IT USA's Stedler has seen only a 7.5 times reduction in data vs. the 15 to 20 times reduction promised by Data Domain. But he said much of the difference is due to variations in the type of data he's deduping and that "overall, we're quite happy with its performance."

Rich April, director of network engineering at Boston-based health care provider Harvard Vanguard Medical Associates, cited a recent conversation with a storage vendor that claimed it could reduce his storage needs by 60% to 80%. Those numbers assumed the data being deduped were mostly files; however, his older primary database environment "doesn't play nicely with these newer technologies," so he wouldn't get anywhere near those savings. He has, however, seen a 70% reduction in his backup data by switching from tape-based to disk-based backup of remote-office file shares using EMC's Avamar.

Among the factors to consider in deciding what savings you'll see are the amount and type of data in your environment; the capabilities of your existing storage network, controllers and arrays; and your requirements for application and backup performance. As StorageIO Group's Schulz said, when it comes to storage savings, "your mileage will vary" based on your specific environment. But with all the new and emerging ways to save on storage, the trip is well worth it. ☉

Robert L. Scheier is a freelance technology writer based in Boylston, Mass. He can be reached at bob@scheierassociates.com.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update



Software vendors expand data deduplication offerings

By Beth Pariseau

BACKUP REDESIGN AND DATA DEDUPLICATION have remained hot in the storage market this year, even in the midst of our economic woes. But just offering data dedupe as an add-in feature isn't enough anymore; vendors are also coming to market with ever-finer differentiations as the technology becomes ubiquitous in backup operations.

In 2009, six companies—Acronis Inc., Barracuda Networks Inc., CA, CommVault Systems Inc., IBM Corp. and Symantec Corp.—have either added or expanded data deduplication in their backup software. CommVault threw down the gauntlet first with the announcement of data deduplication for Version 8 of its Simpana backup software in January. Simpana 8 also introduced a unique ability to dedupe data to tape.

You can add data deduplication to your backup operations with an appliance, VTL or target array—but backup application vendors offer software-only alternatives.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

CommVault's approach looks to cut down on the number of tapes it will take to restore an individual file by having users specify a predefined retention window. If that window is 30 days, the index with reference data for that 30-day period is stored on each tape so data can be re-stored from just one tape. At the end of the retention window, a new index is created.

CA's ARCserve Backup Version 12.5 offers users the option of "re-inflating" data before it's stored on tape or copying deduplicated full backups directly to tape. Incremental deduped backups like CommVault's aren't supported by ARCserve—only a full backup, which must first be restored to the media server, can be copied to tape in its deduped form.

"We don't think that during a [tape] recovery of an incremental data set is the time to be rehydrating data," said Don Kleinschnitz, senior vice president of engineering at CA.

Zahid Ilkal, CommVault's senior product manager, countered that most users don't use tape for operational restores. "In the extremely rare situation that we need to restore an individual file from tape, a longer recovery time is expected and tolerated by end users and is a minor tradeoff to make," he said. Users can also use a preview function in Simpana to look at files on tape without having to fully restore them. If a user is doing a full disaster recovery (DR) restore of all data, Ilkal said, "Our recovery technique is optimized in this case to restore all data from the front to back in the tape set without jumping across tapes for DR."

Acronis, Barracuda, IBM and Symantec don't currently offer dedupe on writes to tape.

Another emerging trend is a movement toward deduplicating data at the source, or the client server that hosts the application. CA ARCserve, CommVault Simpana and Symantec Veritas NetBackup PureDisk currently offer data deduplication at the backup server level, reducing network traffic between the backup server and the backup target, but not between the client and the backup server. NetBackup PureDisk has had the ability to deduplicate data at the source since before Symantec picked up the deduping IP

FOUR ROUTES TO DATA DEDUPE

Data deduplication can generally be implemented in a backup system in four different ways:

1. In-line appliance (e.g., Data Domain Inc.'s DDX product line, IBM Corp.'s Diligent ProtecTier)
2. Target backup disk array (e.g., ExaGrid Systems Inc.'s EX series, NEC Corp. of America's Hydrastor)
3. Virtual tape library (e.g., Sepaton Inc.'s S2100 series with DeltaStor, Quantum Corp.'s DXi-Series)
4. Backup application software (see "[Backup apps with dedupe](#)," p. 33)

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

when it acquired Data Center Technologies in 2005, but Symantec has only announced it will deduplicate data from the source when PureDisk melds with its NetBackup and Backup Exec apps over the next six months.

Acronis users have the option of deduplicating data from either the source or the backup server level. Barracuda's recent integration of BitLeap data dedupe IP (acquired in November 2008) with Yosemite's backup software app agents adds application-

aware dedupe at the source level. IBM remains the outlier in this regard—its dedupe is offered post-process, at the backup target.

Whether and how much vendors are charging for adding dedupe to existing backup software is another differentiator for some offerings; Acronis, CA, CommVault and Symantec charge for the feature. IBM customers get dedupe free of charge, while Barracuda claims Yosemite's \$1,500 unlimited server backup license keeps its offering competitive.

Lauren Whitehouse, an analyst at Milford, Mass.-based Enterprise Strategy Group (ESG), said global deduplication is the next frontier for software and target device dedupe vendors alike. "Everybody's low on the maturity curve here," Whitehouse said. For CommVault, dedupe is global only within the same policy group (though CommVault argues this is sufficient for most customers), while dedupe is global only among system files for CA; application data is deduped separately within each backup server.

Arun Taneja, founder and consulting analyst at Hopkinton, Mass.-based Taneja Group, said these differentiators among backup products will ultimately be temporary, as data deduplication moves out of backup toward primary storage and application hosts. In five years, he predicted, "the benefits of deduplication on the primary storage side will flow right through to the back end—there won't be this microscopic focus on the backup and archiving world." ☉

Beth Pariseau is a senior news writer at SearchStorage.com.

BACKUP APPS WITH DEDUPE

To date, six backup application vendors offer data deduplication options for their software suites:

- Acronis Inc. Backup & Recovery 10
- Barracuda Networks Inc. BarracudaWare Yosemite Server Backup
- CA ARCserve 12.5
- CommVault Systems Inc. Simpana 8
- IBM Corp. Tivoli Storage Manager V.6
- Symantec Corp. Veritas NetBackup PureDisk



Is cloud-based backup right for you?

Before deciding if cloud-based backup is a fit with your company, you need to understand the two basic flavors of cloud backup — SaaS and hybrid.

CLOUD-BASED BACKUP is garnering a lot of attention from small and large companies alike. There are many reasons why you might consider outsourcing all or part of that function, ranging from the need to improve backup processes, a desire to reduce costs or an interest in taking advantage of the elasticity of the cloud. But how do you know if it's right for your company?

To answer that question, you need to understand the two basic flavors of cloud backup: backup software as a service (SaaS) and “hybrid” cloud backup.

With backup SaaS from the likes of Carbonite Inc., Decho Corp., IBM Corp., Iron Mountain and Symantec Corp., IT accesses an application hosted and operated at a central location via a Web interface and takes advantage of a shared, scalable infrastructure. Disk-to-cloud transfer of data occurs at scheduled intervals.

The hybrid scenario calls for the use of existing on-premises licensed backup software and hardware with data backed up to a cloud storage provider. Vendors in this space include Axcient Inc., Barracuda Networks Inc., CA, Hewlett-Packard Co., i365, IBM, SunGard, Symantec, VaultLogix LLC, Venyu Corp. and Zmanda Inc. The hybrid approach allows you to maintain on-premises control of infrastructure while taking advantage of off-premises infrastructure and services staff. Often, the cloud provider's data center facility, infrastructure, staff and processes are a step (or two) above what you may have at your primary site. Typically, data is backed up first to on-premises disk, and a duplicate copy is maintained at a third-party cloud provider.

Cloud-based backup is increasingly becoming the basis for the entire backup strategy at smaller companies.

WHERE CLOUD-BASED BACKUP MAKES SENSE

Cloud-based backup is increasingly becoming the basis for the entire backup strategy at smaller companies. At bigger organizations, it can help plug gaps in the data protection strategy—gaps that can't otherwise be filled due to staffing or budget constraints. For instance, some IT shops

don't have a core competency in data protection, so outsourcing all or part of the function may make sense. One of the current contributors to backup angst is a lack of budget to support staffing levels or to acquire and maintain infrastructure. Without the people, processes and technology to manage backup and recovery, it's difficult to meet service levels.

It's also possible that you just don't have the ability to provide complete protection. For instance, at some firms, laptop and desktop PCs aren't incorporated into server backup processes, creating vulnerabilities. Similarly, edge data at remote offices and branch offices (ROBOs) is often underprotected or not protected at all. In those scenarios, outsourcing these pockets of the data protection function may be wise.

DETERMINING YOUR SUITABILITY

Assessing your organization's challenges, abilities and assets will help to determine if deploying a cloud-based backup strategy is a better alternative than relying on an on-premises one. Here's what to consider:

Budget. Do you know what your current costs are for data protection? Have you evaluated staff costs to determine if eliminating any on-premises infrastructure, introducing automation or adopting more-advanced technology, would alleviate issues? A comparison of all capital and operational expenses for on-premises technology over three years vs. the operational expenses for cloud-based backup over three years may yield surprises. For example, for a backup tape strategy, the maintenance fees for on-premises hardware and software, media purchases and storage fees, and operations overhead over three years could pay for three years of a hybrid cloud service.

Daily capacity of backup data. How much data needs protecting (based on the total capacity of data and the daily change rate)? What's the frequency of backups required to meet recovery objectives? Calculate how much backup data needs to be transferred on a daily basis and, given the available bandwidth, whether or not the transfer can be accomplished within the backup window.

Data protection gaps. Do you have new directives to improve protection at ROBOs or with endpoints? Do you have the staff and capital budget to invest in these new initiatives? Do you currently have a disaster recovery (DR) strategy? If yes, do your people, processes and technology allow you to meet recovery time objectives (RTOs) and recovery point objectives (RPOs)? Leveraging a cloud backup vendor to augment current on-premises data protection processes may be more cost-effective (no upfront capital investment in infrastructure and no additional headcount required) than extending on-premises capabilities.

Infrastructure. Is the current infrastructure limiting your ability to meet the needs of the organization? Have budget constraints limited your ability to keep pace with technology advancements? Were you going to do a technology refresh soon anyway? Organizations without the capital budget to extend or refresh the IT infrastructure to meet data protection needs, but with a sufficient operational budget can

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update

fund monthly service fees through their operational budget to improve data protection processes.

Service-level agreements (SLAs) and compliance mandates. Can you successfully complete your backup within the prescribed window of time? Can you recover data to meet agreed-upon timeframes? If not, is the inability related to outdated technology or lack of sufficient operational staff? Can you meet corporate and/or regulatory requirements with your current people, processes and technology? Does your process for maintaining offsite copies introduce any security risks? If you have recovery SLAs that are more aggressive than what can be delivered by streaming data over your WAN link or physically transported on portable disk (not unlike tape media from offsite storage), cloud-based backup may not be for you. If you don't have the discipline and capabilities to meet compliance objectives, then outsourcing data protection to a vendor that can support compliance efforts may pay off.

Staffing levels and expertise. Has the economic climate impacted your ability to appropriately staff the data protection function? Do your current backup/recovery infrastructure and processes rely too heavily on operational staff? Do you have in-house expertise to properly architect, build and maintain data protection infrastructure and processes to meet objectives now and in the future? Operations staff is often the most costly aspect of data protection, so adding data protection capabilities without necessitating additional staff could be more feasible.

DECISION TIME

Once you've considered those questions, you're ready to decide whether or not cloud backup makes sense for your company. Because many organizations still rely on on-premises solutions and tape media storage for offsite copies, it's a good bet that improvements can be made to the existing data protection process. Local disk-based approaches can improve performance and reliability, which saves time in backup, recovery and management processes. And disk becomes even more cost-effective if data deduplication is part of the mix.

Alternately, leveraging a cloud-based service can save time and money. Management of the data protection environment is removed from IT (which is often the biggest expense in backup environments), cutting down on the amount of operational overhead needed. Because a cloud-based service is funded out of an operational budget, it's often more palatable for budget-constrained organizations. Finally, backup via a cloud service delivers built-in DR.

A popular approach is to combine these two strategies—local disk-based backup for operational recovery of data and remote cloud-based backup for disaster recovery. This model delivers the benefits of both and removes a huge burden from IT organizations. ☺

Lauren Whitehouse is an analyst focusing on backup and recovery software and replication solutions at Enterprise Strategy Group, Milford, Mass.

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

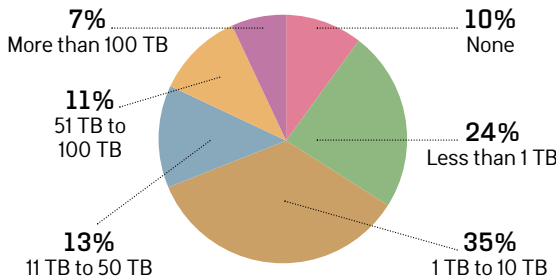
Solid-state storage update

Capacity still biggest backup bugaboo

BACKUP IS HARD and, despite the numerous products that are supposed to help, it's not getting a lot easier, at least according to our latest *Storage* magazine survey. Asked what their biggest backup issue was, 65% of respondents said just keeping up with growing capacity. But that was an improvement over last year when 73% cited capacity as the main sticking point. Backup windows are another source of pain, as 58% said the time it takes to complete a backup is their most vexing backup problem, just nudging out backing up redundant data (55%). Fifty-six percent said they added more disk to address their backup woes vs. last year's 61%. Data deduplication might have helped reduce that number, as 42% are now using or have definite plans to use dedupe to pare their backup data, a 5 point increase from the 2008 survey. And an even bigger segment (47%) will give implementing dedupe serious consideration to see if it can streamline their backups. —Rich Castagna

Capacity is biggest backup issue
Is cloud-based backup right for you?
Vendors expand data dedupe offerings
Use storage more efficiently
Solid-state storage update

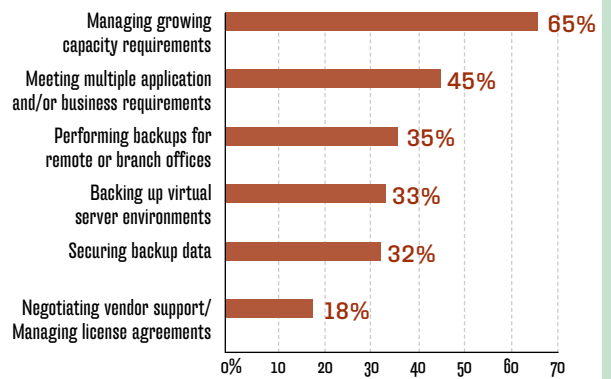
How much disk capacity will you add this year specifically for backup use?



Which of the following technologies and/or remedies are you using now (or have definite plans to implement) to help address your backup and recovery challenges? (Check all that apply.)

- 56% Additional disk capacity
- 42% Data deduplication
- 36% Additional tape library capacity (adding slots/drives or new libraries)
- 29% Additional bandwidth
- 28% Continuous data protection (CDP)
- 23% Some form of encryption
- 17% Wide-area file system or application acceleration/bandwidth optimization appliances to connect remote offices
- 8% Contracting out with a service provider for backup and/or recovery

What are the biggest backup and recovery issues you're facing today?*



* Note: Respondents could make multiple selections

25

Average TBs of disk capacity to be added this year just for backup.

“We need to migrate to disk-based backups, but even with deduplication, finding a solution that backs up 100-plus TB of data within our budget is a real challenge.”

—Survey respondent

Check out the following resources from our sponsors:

Fujifilm USA, page 7

[Fujifilm: Optimizing Performance and Maximizing Investments in Tape Storage Systems](#)

Maxell, page 15

[Help Identify Problem LTO Cartridges Before They Fail](#)

[Detecting Worn Tapes Before It's Too Late](#)

Quantum Corp., page 4

[City Where Fantasy Thrives Solves Reality of Backup Challenges with Quantum DXi7500 Express](#)

[Case Study: Cutting Backup Times from Eight Hours to Twenty Minutes at U.S. Army Command](#)

Spectra Logic Corp., page 9

[Backup and Archive Easier Through Tape Without Pain. Read the White Paper.](#)

[Blistering Fast Dedupe: Backup and Dedupe 20 TB in Less Than Seven Hours with Spectra nTier!](#)

Zenith Infotech, page 21

[On Demand ARCA Demos](#)

[Learn More about the Zenith ARCA](#)

Capacity is biggest backup issue

Is cloud-based backup right for you?

Vendors expand data dedupe offerings

Use storage more efficiently

Solid-state storage update