*Managing the information that drives the enterprise*

# STORAGE

**DON'T GET BURNED; TAKE CONTROL OF FILE SYNC AND SHARE**

**LTFS MAKES TAPE MORE USEFUL THAN EVER**
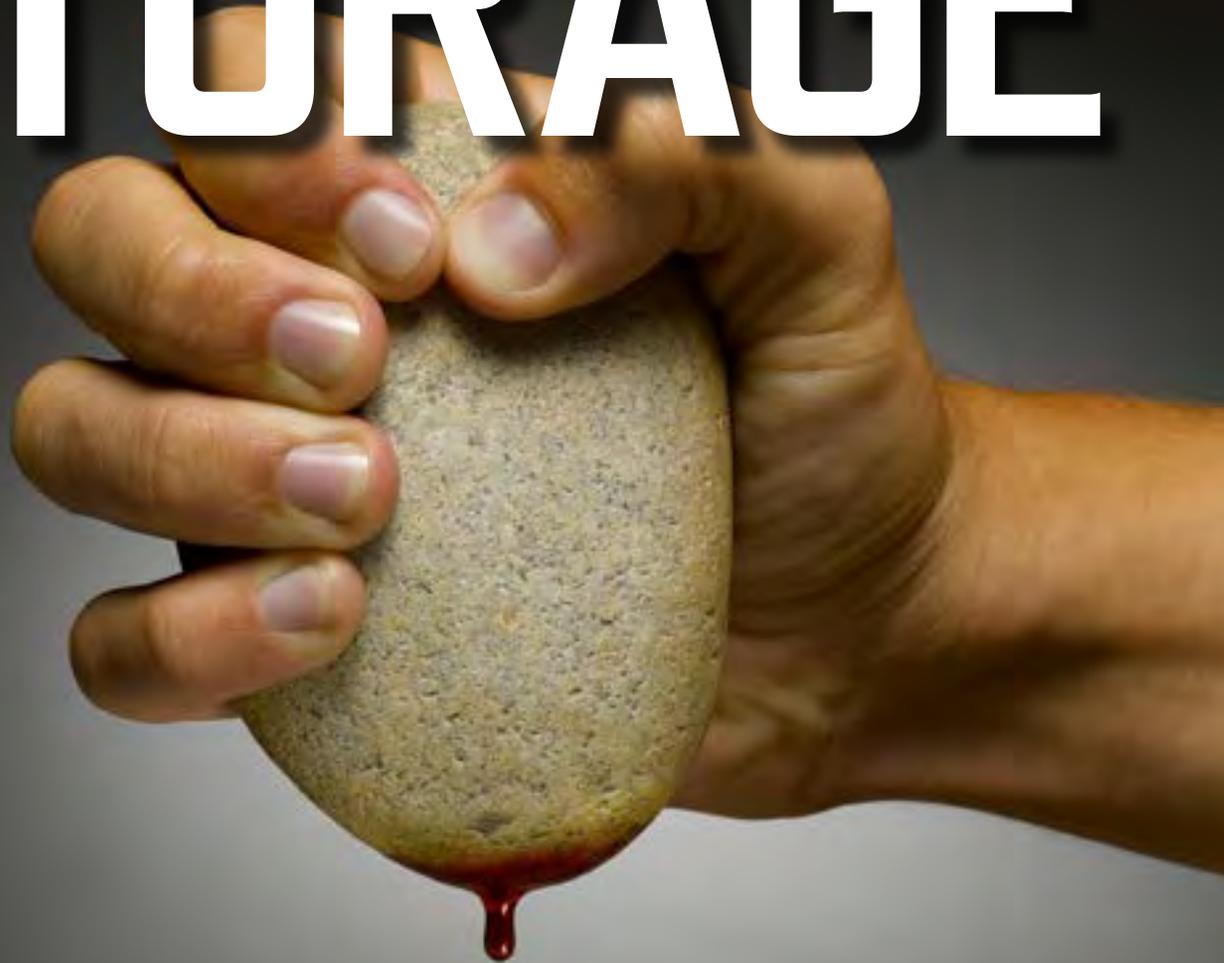
# How to Get the Most Out of Solid-State Storage

There are a variety of ways to implement solid-state storage, each with its own benefits and drawbacks. Learn which options will be best for your company.

**TechTarget**

# The day you get back, challenges will be solutions

**Storage Decisions** CONFERENCES AND SEMINARS

TechTarget

Storage Decisions events are guaranteed to deliver practical advice and technical guidance that you can put to use as soon as you return to the office. Featuring the world's most renowned storage experts, our events offer tips and vendor-neutral advice for tackling current issues. Our innovative vendor showcases help you avoid falling behind by providing exclusive insight into market trends and emerging technologies. Plus, only you and your true peers in IT can qualify for FREE admission, ensuring an unmatched experience at an unbeatable value.

**Visit www.StorageDecisions.com to register for an event near you.**

EDITORIAL | RICH CASTAGNA

# Back to the future: Fleeting fashions of storage

*Are all those new storage techs actually new or is it déjà vu all over again, but with cool new names?*

A COUPLE OF SUITS in my closet have very wide lapels. I also have a few with skinny lapels. Given the limitations of men's fashion—especially for "dress-up" clothes—one of those styles is bound to come back, right? However, I did deep-six the Nehru jacket a long time ago, as I'm pretty sure *that* one will never come back.

Style is a cyclical thing that goes beyond the world of fashion, so far beyond that it even affects data storage technologies. (You saw that coming, right?) If people lost interest in something four or five years ago, maybe if you just dust it off, make a tweak here or there, and give it a new name, it might just come back into vogue.

Here's a classic old fashion/new fashion storage story. Hierarchical storage management (HSM) was a mainstay

for mainframes but a bust for open storage systems implementations. It was largely manual, didn't solve a well-defined problem and was even kind of hard to say. But the idea of putting data on the most appropriate storage based on its value made sense, so a few intrepid storage vendors gave it a nip and a tuck, redubbed it information life-cycle management (ILM) and trotted it out for a second try. The products were a little better than their HSM ancestors, had a mostly new acronym and the word "life-cycle" had a healthy, whole-grain feel. But ILM was a bust, too. I remember trying to put together a story on ILM and not only couldn't we find anyone doing it, it was hard to find anyone who knew what ILM meant.

Fast forward to 2009 or so, when solid-state storage started its incursion into storage systems in corporate data centers. The stuff was so expensive back then

that companies looked for ways to make just a little bit of it do miraculous things to speed up storage for critical applications. Once again, enterprising vendors resurrected the old HSM/ILM idea of moving data based on its requirements. In the case of solid-state storage, the requirement was pure unadulterated performance. HSM/ILM principles were a perfect fit: just move data into solid-state when it needed to be there, and get it out and back on cheap hard disks when it no longer needed lickety-split processing. It was a great idea built on "classic" HSM/ILM principles, but what it really needed was a new name: auto-tiering. I'm not sure if auto-tiering is the wide lapel or narrow lapel version of those older technologies, but it seems a lot more stylish.

We've seen a similar swing in styles with cloud storage. It's been around for ages (nearly 20 years), but the services had mundane names like managed storage provider (MSP). MSP fit into that era's *x*SP naming conventions, alongside Internet service provider (ISP) and application service provider (ASP). There were probably other *x*SPs I'm forgetting, but whatever they were, they all seemed to blur together.

But that's in the past. Today, we have cloud storage, which is fun, fluffy, ephemeral and always depicted with a cute little picture. No more of that *x*SP nonsense. You may still not actually know where your data is stored, but hey, it's the cloud and isn't that pretty cool?

Want another back-to-the-future example? How about software-defined storage (SDS)? Besides being a phrase and a concept that stretches credulity (what's been defining storage all along?), it seems to have about a zillion different definitions depending on which SDS vendor you're talking to. The best part of the SDS concept is that it doesn't eliminate hardware; it just kind of adds another

> **The best part of the SDS concept is that it doesn't eliminate hardware; it just kind of adds another layer between you, your apps and your storage.**

layer between you, your apps and your storage. When EMC recently rolled out ViPR—its version of SDS—the firm was quick to point out it probably wouldn't mean that you wouldn't need a VMAX or VNX or two (or three) to run under (alongside?) the ViPR software. So it doesn't sound like the ViPR software is doing a lot of storage defining, does it?

That's because SDS is, for the most part, storage virtualization warmed over. Well, maybe not even warmed over all that much, as some of the new stuff lacks the sophistication of storage virtualization products that have been on the market for years. Marketers want to make SDS sound new and exciting, and suggest that hardware

will somehow just disappear. But SDS is just like storage virtualization technology that adds a new layer. Granted, it's a layer that should make managing diverse storage systems easier, but it's something extra that ultimately doesn't eliminate anything.

Storage virtualization never caught on big time either. Our Purchasing Intentions surveys show that approximately 29% of companies have virtualized at least some of their storage, with the emphasis on "some," as only 22% have virtualized all their block storage and 23% all their file storage. But those numbers aren't real shockers, even given the obvious benefits of virtualizing storage, especially in mixed-vendor shops. Storage virtualization was pretty expensive, took a lot of time to implement, and was tough to backtrack on and reverse if the results weren't satisfactory. So, with a few exceptions, the storage virtualization market sputtered and stalled for most vendors. Until it was reborn as software-defined storage. But let's not rush to judgment: SDS is still a brand-new marketing pitch and we don't know if it'll have the kind of success cloud and auto-tiering seem to be having.

I'm not knocking those storage vendors who pull these

old storage technologies out of mothballs, spiff 'em up, give 'em new names, and trot them out as the latest and greatest. It's a good thing, as it shows that good ideas may hang around long enough to get a chance at a second or even third life. And no matter how much the marketing

> **Our Purchasing Intentions surveys show that approximately 29% of companies have virtualized at least some of their storage, with the emphasis on "some."**

wizards buff them up and re-label them, they'll succeed only if the ideas and technologies lurking underneath the snappy new looks will bring them back into style. Hmmm, maybe I should have held onto that Nehru jacket after all. ■

**RICH CASTAGNA** *is editorial director of TechTarget's Storage Media Group.*

STORAGE REVOLUTION | JON TOIGO

# How to speed up cloud storage: Tachyons

*But if you don't have any of those hypothetical particles that move faster than light, there might be another solution for WAN bottlenecks.*

**L**ATELY, I'VE BEEN giving a lot of thought to a technically nontrivial hurdle that cloud storage and WAN-based data replication aficionados don't want us to talk about: Our WANs are subluminal (slower than the speed of light). Outside of the Syfy channel, we're hard-pressed to collect enough tachyons on which to piggyback our data signals so they'll move faster than the speed of light. Neither are there any traversable wormholes in the public interexchange carrier (IXC) network, nor a sufficient quantity of the spice mélange to enable space time distortion (even for *Dune* fans) so our data can travel great distances without really moving. Basically, we're stuck with subluminal velocities in our data transfers. Latency and jitter are facts of life.

There are some vendors that don't want us to think about this too much. I recently pulled a paper I was writing for a client because they wanted to "soften" references to WAN limitations that might stymie the appeal of their cloud value proposition. I saw their edits as an attempt to dumb down the discussion of what are substantively important gating factors on cloud service viability: the vicissitudes of data transfer over long distances. Frankly, because clouds depend on WAN links to deliver access to services, I fail to see how they can possibly guarantee—with a straight face, that is—any sort of meaningful service levels.

Try as they might to deliver quality infrastructure, applications or capacity services in conformance with a service-level agreement, cloud- or network-based service providers are at the mercy of the miles and miles of copper and glass that connect customers to their facilities.

And all of them are controlled by a gaggle of incumbent local exchange carriers (ILECs), competitive local exchange carriers (CLECs) and IXCs that provide connections between local access and transport areas (LATAs). That voice, video and data is able to move at all through this hodgepodge of turf captains is a miracle. Moving data through these component parts efficiently and reliably would be an even greater miracle.

The political and economic stresses and strains in the relationships between these carriers can impact transfer rates through their cobbled networks. Each day, one of my clients wonders whether their transfer of less than a hundred gigabytes of data between Sacramento and the Silicon Valley, on a path "owned" by nine carriers, will take a few seconds or the entire day. There's simply no way to predict the transfer rates.

Add to the bureaucratic nonsense the sobering realities of switch protocols that deliver not the shortest physical path between point A and point B, but rather the least number of router hops. Moving data through a public WAN is like taking a multisegment airplane trip from Tampa to Dallas; the carrier gets you to your destination not by flying across the Gulf of Mexico, but rather using a route that moves you through Boston, Charlotte and Detroit before landing in Dallas.

Then toss in routing problems within the switch fabric itself. Queuing delays, buffering delays and buffer bloat, dropped packets, resend requests and other processing delays can gum up the works and slow transfers even further. Pass-through technologies are limited in their ability to streamline or expedite traffic through heterogeneous infrastructure where the proprietary quality-of-service protocols provided by one switch vendor are dropped as soon as traffic hits a switch from a different vendor.

> **Add to the bureaucratic nonsense the sobering realities of switch protocols that deliver not the shortest physical path between point A and point B, but rather the least number of router hops.**

Adding more bandwidth doesn't fix the problem, nor does compressing or deduplicating data payloads. Think of a traffic jam on your favorite highway: that little Fiat 500 ahead of you isn't moving any faster than the big Peterbilt 18-wheeler in the next lane, is it?

It's irritating when vendors obfuscate these matters. I almost brought my ire to bear on one vendor last month when I received a press release proclaiming that Bridgeworks Limited had beaten the whole latency thing. Despite my anger, I sat for a couple of Web meetings with Bridgeworks (and with folks from Agilesys, which

partners with them in the U.S.) to learn more about their SANSlide product.

The first thing I discovered was that the "no more latency" claim was a bit of hyperbole introduced by PR flacks. SANSlide doesn't fold space time; it just optimizes the way connections are used and how data is placed onto the link by using "patented artificial intelligence" to fill the pipe. Bridgeworks claims—backed by many customer accounts and endorsements—it was improving the performance of data transfers across WANs by as much as 50x without requiring "warm" data, compressed data or deduplicated data, and without the use of hard disk caching, UDP or modified TCP/IP protocols.

My takeaway thus far, and I plan to test their appliance soon, is that Bridgeworks isn't moving data more quickly through networks, but packing data more efficiently onto pipes and balancing the load across multiple virtual and physical interconnects.

In short, SANSlide optimizes the network capacity you're paying for by packing more luggage and passengers onto the plane, so to speak, so the payload gets to where it's needed as efficiently as possible. But you're still at the mercy of the CLECs, ILECs and intra-LATA IXCs when it comes to traffic routing, so we haven't really beaten the subluminal limits that are cooling enthusiasm for clouds

and WAN-based data replication.

An important caveat: Not all data transfers are latency sensitive. Depending on how you intend to use your data—how synchronized replicated data must be with source data, for example—a certain amount of latency may be tolerable. In cases where latency isn't a significant issue, SANSlide technology may be a great choice for WAN optimization. It's possible to use a pair of SANSlide appliances to "black box" the WAN in much the same way channel extenders spoofed connections between the mainframe backplane and remote peripheral devices 20 years ago. To the mainframe, the channel extender appeared to be the peripheral, only it buffered the commands and data, and placed them on the WAN link for delivery at top speed. At the remote location, the peripheral device talked to the extender as though it was communicating with the mainframe backplane. This is very similar to the operational model of Bridgeworks' SANSlide.

It isn't a wormhole or a Tesseract, but it looks like pretty interesting technology. ∎

JON WILLIAM TOIGO *is a 30-year IT veteran, CEO and managing principal of Toigo Partners International, and chairman of the Data Management Institute.*

# SOLID-STATE STORAGE BEST PRACTICES

No longer just a luxury item for well-heeled data centers, solid-state storage is more affordable than ever and available in a variety of form factors with a choice of deployment options.

**SOLID-STATE STORAGE** technology has completely revolutionized consumer electronics products, replacing spinning disk drives in virtually all types of compact consumer devices such as mobile telephones and tablet PCs, and becoming the standard for ultra-thin laptop computers. The same benefits and enthusiasm for solid-state storage are now being felt in data centers.

While the number of units shipped for some types of hard disk drives (HDDs) is declining, solid-state drive (SSD) shipments continue to soar, and are quickly gaining acceptance for enterprise applications. Solid-state storage is a good fit for database applications, virtualized servers, virtual desktop environments and other workloads that need higher performance based on any of the following criteria:

- Transactional performance, measured in I/Os per second (IOPS)

By Dennis Martin

- Total throughput or bandwidth, measured in megabytes per second (MBps)

- Lower round-trip time for I/O operations, measured in latency (milliseconds or microseconds)

Just as with HDDs, SSDs can be grouped into two basic categories: enterprise and client. Enterprise SSDs (and HDDs, for that matter) are rated for 24x7 operation, usually have better performance, typically have longer warranties and expected life, and cost more per gigabyte (GB). Client SSDs (and HDDs) are generally not rated for 24x7 operation, usually have higher capacities and cost less per GB.

### NAND FLASH SSDs AND WRITES
Due to the physics of the NAND flash used in SSD technologies, there are a finite number of writes that can be performed on each bit of NAND flash media.

Industry and SSD manufacturers use the concept of the number of full capacity overwrites than can be performed in a day. An enterprise SSD will be able to sustain writing several times its capacity every day for at least five years. A client SSD will usually be able to sustain writing less than its full capacity every day for many months. The term terabytes written (TBW) is often used to indicate this characteristic. Because of this, enterprise SSDs are

suitable for applications that perform a large number of writes every day, such as database applications.

### FORM FACTOR OVERVIEW
SSD technology is available in several form factors, either by itself or combined in a hybrid fashion with other technologies such as HDDs. By itself, SSD technology can be implemented as a cache in one of several locations in the IT infrastructure or as primary storage, either in discrete devices or in large, all-flash storage arrays. In a hybrid fashion, SSD technologies can be combined with HDD technology to form individual hybrid devices or large hybrid flash-optimized storage arrays.

- **Disk drive form factor.** The disk drive form factor is common packaging for SSD technology. SSDs are available in 1.8-inch, 2.5-inch and 3.5-inch "disk drive" form factors that use the same connectors and interfaces available for today's HDD technology, with the 2.5-inch size the most common physical size for SSDs. Capacities for these SSDs range up to 2 TB in a single drive today. SATA and SAS are common interfaces for SSDs in these form factors, just as they are for HDDs. The newer enterprise SSDs are beginning to use 12 Gbps SAS as the interface to deliver higher performance than was previously available. Drive form-factor SSDs are also available with older interfaces, but they're on the decline.

Because of the higher performance of SSDs, new interfaces are being developed such as SATA Express, SCSI Express and Non-Volatile Memory Express (NVMe) that can accommodate the drive form factor. Expect to see these new interfaces available on shipping products in late 2013 or early 2014.

● **PCI Express (PCIe) card form factor.** Another common form factor for SSD technology is the PCIe card form factor. These cards fit into a PCIe slot in a computer and provide fantastic performance with very low latency because the storage is directly accessible via the fast PCIe bus. Enterprise versions of these PCIe SSDs tend to be more expensive than the same capacity of a drive form-factor SSD, but the performance is usually better. The largest capacity PCIe SSD today is 10 TB, but the cost of this particular SSD is several times that of a typical server.

Many of these PCIe SSDs support PCIe 2.0 and require either an x4 or x8 slot. Some newer cards also support PCIe 3.0, which became available in new servers sold beginning in 2012.

One important aspect of this SSD form factor is the card's physical dimensions. Some are half-height and half-length, which means they'll fit into most servers, even the smaller form-factor servers. The full-height and/or full-length PCIe SSD cards have additional capacity but may not fit into all servers.

## Five biggest solid-state implementation mistakes

- Not understanding application workloads; for example, using SSD caching with a cache "unfriendly" workload.

- Not expecting that bottlenecks may move after deploying SSD technology.

- Not buying enough SSD to anticipate the demand.

- Not understanding the differences in use cases between server-side SSD and external SSD.

- Not understanding the difference between "fresh out of the box" performance and the "steady state" performance of SSDs.

In addition, some of the larger-capacity PCIe SSDs require additional power beyond the standard 25 watts that can be drawn directly from the PCIe slot. Some servers are equipped with additional power connectors.

● **SSD-specific form factors.** Because of their small size and low power requirements, SSDs are available in other form factors, many of which are designed for smaller applications. These include mini-SATA (mSATA), which is about the size of a business card (or smaller); it has capacities up to 256 GB, uses the SATA command interface and fits into a PCIe slot. Another is the μSSD (micro SSD), which is a silicon chip mounted directly onto a motherboard that appears to the operating system as a SATA interface storage device and is designed for very low-power applications such as mobile devices.

An interesting form factor for SSD technologies that may have enterprise applications is the dual in-line

memory module (DIMM) slot form factor. These products are NAND flash and non-volatile DRAM devices that fit into a standard DDR3 DIMM socket but provide non-volatile storage capacity for a server. These could provide an interesting way to add enterprise storage to a server with many DIMM slots.

## HYBRID DRIVES AND ARRAYS

There are a number of hybrid SSD/HDD offerings that incorporate both SSD and HDD technology in the same offering. Individual hybrid drives are generally intended for consumer and desktop applications, and typically use the SSD as a cache in front of the HDD. Larger hybrid flash-optimized storage arrays are built with a combination of discrete solid-state devices (SSDs or PCIe cards) and separate HDDs, and are designed for enterprise storage applications. The SSDs in these hybrid arrays can be used as a cache or for tiering in a primary storage design.

## USE CASES AND WORKLOADS

Although we've seen improved performance using SSDs for almost every workload in the Demartek test lab, there are some workloads that perform especially well with SSD technology. Database workloads work well with SSDs not only because of the higher raw performance of SSDs, but because of the reduced latency. We've seen

## More solid-state storage information and performance reports

**DEMARTEK LABS** has additional free information available on the SSD Zone of its website. There are a number of performance reports for specific solid-state drive technology; a comprehensive SSD Deployment Guide is also available.

significant performance improvements when using SSDs as a cache and as primary storage. Placing the application data entirely on SSD media (primary storage) will yield enormous benefits immediately for that application. We've seen performance improvements of 8x to 16x or more by using SSDs as the primary storage location. Unfortunately, you may not be able to afford that kind of configuration for every application in your environment.

While an individual HDD can provide hundreds of IOPS, SSDs typically provide thousands or tens of thousands of IOPS with a single device. In some online transaction processing (OLTP) application environments, including database and Web server environments, short response times are often critical. A transaction may require many successive queries to a database, where each query depends on answers returned from the previous query. In those cases, user response times are entirely dependent on how quickly storage can return answers to the entire series of queries. In many cases, we've seen sub-millisecond latency using SSD technology, which may be more important than the raw IOPS or throughput performance.

## SSD CACHES AND CACHE-FRIENDLY WORKLOADS

Deploying SSDs as a cache has the advantage of potentially sharing the performance benefit among many applications because the cache will improve any I/O that is "hot," regardless of the application. SSD caches are also a good way to get started with a relatively small amount of solid-state storage. Caches need time to "warm-up" and become populated with the hot data, which can take minutes or hours, depending on your environment. Workloads that have "hot spots" with repeated accesses to a subset of the entire data set are considered cache-friendly and will benefit from an SSD cache. We've seen OLTP workloads increase in performance 2.5x to 8x by using an SSD cache, depending on the cache offering, speed of the back-end storage and other factors.

SSD caches can be deployed in three general locations within an IT infrastructure: server-side, in the network or in a storage system. Each of these has its advantages and disadvantages, so your choice will depend on your needs. In some environments, the ability to share the cache across different applications will be important; in others, not having to change the servers or back-end storage might be important. Server-side SSD caching has the advantage of offering much lower latency to the application, but moving an application to a different server (such as in a virtual machine environment) may require the cache on the new server to be re-warmed before the maximum benefits can be achieved. ∎

**DENNIS MARTIN** *has been working in the IT industry since 1980. He is the founder and president of Demartek, a computer industry analyst organization and testing lab.*

# ENTERPRISE FILE SYNC-AND-SHARE APPS

Cloud-based file sharing and syncing are becoming more popular as employees use their own smartphones, tablets and other mobile devices to access corporate data. Here's what you need to know to keep your company's data safe.

BRING-YOUR-OWN-DEVICE (BYOD) policies and an increasingly mobile workforce are putting new pressures on IT and changing the requirements for how workers want (and need) to access corporate data. IT is no longer in a command-and-control role where it can dictate the hardware and software that employees use. Workers want to access files anywhere, anytime and from a variety of endpoint devices.

Many business users have turned to consumer cloud-based file-sharing and collaboration services. These services are simple to use, but they're a dilemma for IT as it needs to balance business enablement with the ability to maintain the control and security of company data.

## THE DANGERS OF CONSUMER FILE SYNC AND SHARE

Consumer-focused cloud file sync-and-share services offer

By Kristine Kao and Terri McClure

free storage and file-sharing functionality, and are quick to install and easy to use. Business users can increase productivity by removing barriers to file access and sharing, but there is some risk. The danger lies in employees signing up for personal accounts, as data stored in these services falls outside the control of IT, potentially exposing organizations to data leakage and data breaches.

IT may not know an account even exists or what's stored in it, and has no way of applying the organization's security and access policies to protect company assets. This is especially disturbing because employees with such accounts often sync corporate data across multiple devices, such as a home computer, tablet or smartphone. If an employee with a personal online file-sharing account leaves the company, the former employee still owns the account and any corporate data stored there is still accessible by them using their personal access devices. Clearly, this can create an unacceptable security, legal and business risk for employers.

## IT SEEKS SOLUTIONS FOR
## SYNC-AND-SHARE RISKS

Most IT shops are aware of this risk and have data security top of mind. Enterprise Strategy Group research indicates that 70% of organizations know or suspect this type of rogue employee activity takes place within their companies, and they're sprinting to catch up and regain

control over company data.

First and foremost, IT needs to rein in personal account use among employees. Many companies have formal policies or discourage employees from having their own accounts, but while blacklisting consumer cloud-based applications may curtail the security risks short term, end users will ultimately find ways to get around

> **IT may not know an account even exists or what's stored in it, and has no way of applying the organization's security and access policies to protect company assets.**

company firewalls. It's only a matter of time before they realize they can hop into a Starbucks and get on a Wi-Fi network, or pull out a hotspot and gain access to their personal accounts. The best way for IT to handle this is to deploy a company-approved corporate online file-sharing account. This helps to ensure data stays within company-approved applications and that IT will control company data, yet grants employees the access and functionality they feel they need to be productive.

The good news is that the corporate online file-sharing market is a burgeoning one, with more than 50 vendors

offering some combination of "secure" file store, sync, share, send and collaboration. There seems to be an offering to suit nearly every business need—the challenge for IT is that there is no one-size-fits-all file-sharing solution because organizations have varying security and business requirements. IT therefore needs to think about a number of considerations so it can implement the best corporate online file-sharing solution.

## PUBLIC, PRIVATE OR HYBRID

The first, and most important, consideration is the type of deployment model. Online file-sharing services can be delivered via public, hybrid or on-premises offerings. In a pure service or hybrid offering, software is delivered as a service and some (or all) corporate file data primarily resides within the cloud (albeit for sync and share; it also lives on laptops, desktops and mobile devices). With an on-premises implementation, IT organizations deploy the application and supporting infrastructure in-house, and secure and maintain it like any other enterprise app with the focus likely on secure file sharing and mobility.

Public cloud solutions are typically easy to configure and deploy, but companies are completely dependent on the service provider for file security. And if a data center goes down or is breached, IT doesn't have any control over how long an outage would last or what data may be compromised. On-premises solutions take more

work to deploy and can be more expensive since the company absorbs all the associated data center costs while IT maintains full control of the application. Businesses in regulated industries are often more comfortable with this setup since they can better manage policies to stay in compliance. Hybrid cloud solutions fall somewhere between the two: IT can control what data stays behind the company firewall, but the control plane lives in the cloud so there could still be some risk of data leakage at the provider level.

## KEY FEATURES AND CONSIDERATIONS

In addition to deployment models, customers should ask vendors numerous questions about the product's features and functions to ensure their business and security requirements can be met. Here are some key features to pay attention to:

● **Integration** with existing tools and processes like single sign-on (SSO) and Active Directory must be considered. SSO lets users sign in using one set of credentials across their work devices and apps, and gives IT greater flexibility in using existing policies to manage accounts. SSO also allows organizations to plug their online file-sharing solution into their existing identity management solutions (Active Directory, LDAP, Ping Identity, OneLogin, Centrify and so on) to inherit existing passwords and

security policies, and to allow for auto-provisioning/deprovisioning of accounts from one control point.

● **Granular administrator controls** are extremely important to IT when it comes to protecting company data. Most offerings allow IT to set some type of permissions around data sharing (within an organization versus external sharing) though not all provide controls that define which users can access what data or which devices can be used. Some offerings allow IT to set password requirements (length, complexity, expiration period and so on) and some offer digital rights management functionality to help protect content (file expiration, printing restrictions).

In a mobile world, one of the most important administrative features to protect content is the ability to remotely wipe a lost or stolen device. The degree of control varies from vendor to vendor, and companies should do sufficient research to ensure that admin controls meet their particular needs and compliance requirements.

● **Auditing/reporting capabilities** are other key features to consider. In addition to setting policies to prevent accidental or malicious data leakage, administrators need some degree of reporting or auditing. Again, available functionality ranges by vendor and product but at the very least, IT should be able to understand how much data is stored in the corporate online file-sharing account and what's stored in all the accounts. Some vendors offer detailed visibility into individual user activity (with whom users are sharing files, what files they access, when they access files, what devices they use to access files). Others allow IT to impersonate users or even search files globally for a text string. These capabilities can help administrators spot suspicious employee behavior and react accordingly.

## SECURITY NEEDS MAY VARY

Not every department needs tight security; some organizations may find they need very secure online file sharing for some teams, while others just need to access data easily on multiple devices. The user experience can vary significantly among offerings, with the more secure solutions often not as user friendly at the endpoint.

This presents a challenge: Many IT teams report that a key challenge to deploying a corporate file sync-and-share solution is that employees continue to use their personal accounts because it's just easier for them. Endpoint ease of use is a key adoption driver (and inhibitor) and shouldn't be overlooked. Organizations may find themselves deploying multiple file-sharing applications for multiple use cases, providing easy-to-use, consumer-like solutions for general access and locked-down, highly secure solutions for teams that work with sensitive data.

It's worth noting that aggregator services are cropping

up to help users and IT manage multiple individual accounts. These services allow users to access multiple storage accounts from one point of control. Though this helps consolidate access to offerings into one service, this application category is very young and its security functionality is still nascent; however, it's worth keeping an eye on as it matures.

## CLOUD CONVENIENCE CAN'T BE IGNORED

We live in a world where information equals power. With the influx of online file-sharing solutions, distributing information has become easier than ever. As a result, it's now easier for information to fall into the wrong hands intentionally or unintentionally.

IT is tasked with the increasingly difficult job of ensuring company data is secure. Fortunately, sync-and-share

vendors have heard IT's cries for help and are integrating more robust security functionality with every new release. But no matter how hard you try, there's no such thing as a 100% secure IT environment.

IT organizations can reduce their risk by ensuring the product they choose has a deployment model that works for their business, as well as integrations and controls to make their data as safe as possible. But IT must perform the appropriate due diligence and ask smart questions. If an answer doesn't measure up to needs, they should ask what can be done. Many service providers publish terms of service, but they're often negotiable and can be customized for specific business requirements. ∎

**TERRI McCLURE** *is a senior storage analyst at Milford, Mass.-based Enterprise Strategy Group (ESG);* **KRISTINE KAO** *is an associate analyst at ESG.*

# CAN LTFS SAVE TAPE?

The Linear Tape File System makes tape look like a file system, enabling drag-and-drop operations that closely resemble a NAS share. LTFS is having an impact on archiving now, and we'll see broader applications soon.

By George Crump

**THE LINEAR TAPE FILE SYSTEM** (LTFS) has been available for almost three years. It was released when Linear Tape-Open (LTO) drives reached their fifth generation (LTO-5). As a file system, LTFS has since matured and currently supports core operating systems such as Windows, Linux and Macintosh. Now, LTO-6 with LTFS seems ready to reestablish tape as a primary data center storage device and may be able to move tape beyond its old boundaries of backup and archive.

## LTFS DEFINED

The goal of LTFS was to simplify the way users interacted with the tape device itself. The designers of LTFS wanted to make tape as easy to use as a USB stick: just plug it in and start writing data. But due to tape's capacity, data would be measured in terabytes instead of the gigabytes a USB stick could store. In large part, LTFS has achieved

that goal. Users simply need to insert an LTFS-formatted tape into an LTO-5 or LTO-6 tape drive and in a minute the LTFS volume appears on the user desktop. From there it can be browsed just like an external hard disk or flash memory stick, just a little slower.

LTFS relies on two changes from previous LTO generations. The first is the creation of partitions on tape, one of which LTFS uses to store catalog information about the files written to that tape. The second is a series of drivers for each of the major operating systems listed earlier.

When the driver is installed and a tape is inserted into an attached tape drive, the driver reads the catalog partition and presents a folder structure similar to what a user sees when they browse a hard disk or USB drive. Files can be dragged to and from this tape volume just like a hard disk, and the driver updates the catalog and stores the data on tape.

### WHY EXTERNAL STORAGE IS SO IMPORTANT

In today's data center, backup and replication happen in near-real-time and repeatedly throughout the day. The relative ease at which near-zero data loss and app availability can be delivered is unprecedented. While valuable for data center managers, this ability to make nearly instant copies has some risks. As a result, there are specific needs for external storage that tape, especially with LTFS, is competing for with other external storage technologies.

First, thanks to repeated backups and replicated storage, data that has been improperly modified or corrupted can be instantly propagated through the infrastructure making the "last known good copy" almost impossible to find. There have even been cases where an accidental deletion is replicated and data is erased throughout the en-

> **Files can be dragged to and from this tape volume just like a hard disk, and the driver updates the catalog and stores the data on tape.**

vironment. Similarly, a virus infection can be replicated into the copy data set, driving the need for an unconnected, offline copy of data that would be immune from these situations.

The second use case is caused by data growth outpacing available wide-area network (WAN) bandwidth. While bandwidth has improved and is now able to keep secondary sites synchronized to within seconds of primary sites thanks to incremental replication, there are times when a full data set still needs to be transferred to another location. Examples include seeding a cloud storage facility with a data baseline, sharing a large project with a collaborating firm or delivering the legal documents requested in a discovery motion. These are

situations where the bandwidth of a FedEx truck can deliver a larger overnight payload than the fastest WAN connections. There's a need for a storage device that can be easily transported in such a fashion.

Finally, there's the financial reality of storing so much copy data on disk for years. In addition to the physical cost of the disk capacity itself, the cost of data center floor space, electricity and cooling can be significant. With the move to big data real-time analytics, there's a greater need to keep more data online. But there are also sets of data for which the recovery point is known or where each iteration of a file doesn't need to be kept online. When these delineations are understood, a strategy of moving as much of the non-analytic data to an external, offline and compact device makes sense. For example, a database backup has limited online value after two or three more recent backups have been taken, eliminating the need to keep dozens of copies on disk and available for immediate restore.

## THE LTFS ADVANTAGE

Tape with LTFS has several advantages over the other external storage devices it would typically be compared to. First, tape has been designed from Day 1 to be an offline device and to sit on a shelf. External hard drives weren't designed to be powered off and stored for years on a shelf. With the appropriate LTFS driver installed, tapes can be inserted into any LTO-5 or LTO-6 drive and read. No special application is required, eliminating what has historically been a big shortcoming of tape: The need to have proprietary software running in all locations that would read the proprietary format of each application.

An LTFS-formatted LTO-6 tape can store 2.5 TB of uncompressed data and almost 6 TB with compression.

### Read more about LTFS

Check out these resources for additional information about LTFS technology and applications:

- Mixed perceptions over what LTFS can and can't do

- LTFS offers frugal cost profile of tape and access times of NAS

- Will SNIA's new working group inspire LTFS product development?

- Are there any products on the market utilizing LTFS for archiving?

- Impact of Linear Tape File Systems on the media and health care industries

That means many data centers could fit their entire data set into a small FedEx box. Also, tape is more rugged than other portable forms of storage and better designed for transport. Again, with LTFS the sending and receiving data centers no longer need to be running the same application to access the data on the tape.

While data deduplication can help make storage more efficient by eliminating redundant copies of data on disk, it's seldom implemented on primary storage. Usually, a single file exists on primary, secondary and backup storage, each having to run its own independent deduplication processes. Tape allows for a clean sweep of data that simply doesn't need to be on any form of disk but still needs to be kept. The cost and capacity of tape makes these "just in case" copies very affordable.

### ARCHIVERS LEVERAGE LTFS
The previously cited examples of LTFS apps can all be achieved without any additional software, using just the LTFS drivers that are freely available. Archive app vendors have been quick to adopt LTFS to make their products more appealing. Most of these offerings will integrate a hard disk system of your choosing with a tape library and automate the movement of data between disk and tape.

Users interact with the archive as if it was just another NFS or CIFS mount point on their networks. Data is then automatically copied to one or multiple tapes as a matter of policy, but also retained on disk for fast access. As the disk area fills up with data, again based on policy, data is removed from the disk and is then only available on tape.

The technology to merge disk and tape into a single mount point has been around for decades. While archive products pairing disk and tape were successful in a few niche markets, broad adoption has been hampered in large part because of the proprietary nature of the way the archive products wrote data to tape. It meant that applications had to remain running for decades, although few applications had that life expectancy.

LTFS removes that concern. If the archive application can output data in the LTFS format, it can be moved from application to application or sent to a site with no application. This capability allows users to select different applications based on changing needs. It also forces archive application vendors to remain competitive from a development standpoint and allows the movement of project data between companies without requiring them to standardize on the same application.

### THE FUTURE OF LTFS
● **LTFS backup.** The next step for LTFS is for backup application vendors to embrace the standard. This will allow data portability between backup applications in the same manner it has among archive applications. Any time

a data center decides to switch backup applications, it needs to factor in the cost of running a single copy of the old application even if it's backing up to disk, since most backup vendors write to disk and tape in proprietary formats. With LTFS support, a user could simply keep historical backups on tape that could be imported into the new application when needed.

● **LTFS-integrated NAS.** One of the biggest challenges facing storage managers is keeping up with the growth of unstructured data. Much of this data doesn't need high-performance hardware. An ideal solution would be a NAS with tape integrated with it, essentially a primary storage version of the above archive example, but with a high-speed and larger disk cache. Data could then be automatically protected and eventually migrated from primary storage. Imagine a high-speed but cost-effective set of solid-state drives acting as the primary storage tier and then older data being moved to and recalled from LTFS tape without IT intervention.

● **Direct execution.** The final step in LTFS evolution is direct execution of data from the tape device and even direct modification. This means data wouldn't need to be restored to a disk area prior to recovery. A simple example would be streaming a video file from tape instead of transferring it to disk first. This would be ideal in situations where something needed to be looked up from the archive but not recovered. A database application with direct access to LTFS would be able to do this by extending its database to tape and enabling the search for old records or documents directly on tape. Another example would be using Microsoft SharePoint's Remote BLOB Storage feature to move older documents or spurious revisions of documents to a tape-based storage area.

## THE LOWDOWN ON LTFS

LTFS has the potential to change the way data centers think of and use tape. In the past, LTO and other tape formats were clumsy, slow and difficult to deal with. Now, with LTFS, it's as easy to interact with tape as it is with any other storage device. It has immediate application when data transfer volumes are high and bandwidth is low. But it also has broader implications on backup and archive processes, and potentially on the way databases and file systems access data.

**GEORGE CRUMP** *is president of Storage Switzerland, an IT analyst firm focused on storage and virtualization.*

# When protecting virtual environments, seeing is believing

*Data protection for virtual and cloud environments is tough, but it could be a lot easier if you had a better view into what's going on.*

**A**CCORDING TO A recent Enterprise Strategy Group report on *Trends in Protecting Highly Virtualized Environments*, virtualization rates as one of the top (or top) challenges in data protection. But digging deeper, one finds that five of the top six challenges in protecting virtual environments are related to visibility. Specifically, the top six challenges are:

1. Recoverability of data (disappointing but not surprising)
2. Validating backup success
3. Validating recovery success
4. Identifying factors impacting performance of backup operations
5. Simplified views across virtual infrastructure
6. Response time in troubleshooting backup failures, policy violations and so on

Looking more closely at the root words—validating, identifying, views, troubleshooting—you can see that almost everyone is struggling with visibility. Consider this:

- When you're protecting a physical server, you can easily understand every component within the solution, from the installed backup agent on the physical server across an identifiable network path to a backup server, with defined software and storage.

- When you're protecting a virtual machine, abstraction layers are added at the hypervisor layer, as well as two tiers of "agent-like" functionality between the

guest-based components alongside Microsoft Volume Shadow Copy Service (VSS), but also the interaction of the host APIs of vStorage/vStorage APIs for Data Protections or VSS and two operating systems, each of which could affect the functionality of the backup process.

- When you're protecting a private cloud, additional uncertainty is added because those virtual machines (VMs) are now dynamically created during elastic workloads based on overall load and usage; they can be provisioned through self-service portals and the VMs will move between hosts, clusters and datastores. Now, the backup software is truly trying to protect a moving target.

It shouldn't surprise anyone that the top virtualized infrastructure protection challenges are visibility related. Too often, when a backup application is perceived as inadequate for virtual data protection, the presumption is that the backup software needs to be changed. But if we ignore the obvious differences in backup products (e.g., manageability, deduplication and so on), the reality is that most backup products leverage a hypervisor vendor's approved methods for backing up VMs. Are there differences between products and methods? Absolutely. However, others would suggest you should discard your on-premises software and use a data protection service.

But if the challenges you're experiencing are visibility related (or would be resolved with better visibility), then perhaps the answer is not to change software or services but to increase the visibility into your virtualization protection solution. Large enterprises discovered a long time ago that they often have to sustain multiple backup products for a variety of reasons, so they use one of a very small handful of monitoring solutions designed to scrutinize diverse data protection products and provide a consistent dashboard and view. Some of those tools not only provide data protection monitoring (DPM) across backup tools, but are optimized specifically for virtualized environments. When users were asked about the benefits of DPM tools, compliance, improved troubleshooting and assessing success/failure of jobs were the top recognized benefits.

Said another way, IT organizations that choose to add enhanced monitoring tools may actually start to realize the ROI of what they've already paid for and have installed because they can get their existing data protection product(s) to work better. Interestingly, this also provides some business opportunities:

- **Service providers (SPs)** are often looking for new services to offer, and sometimes Backup as a Service (BaaS) is their choice. But SPs may not have the technical experience to deliver a solution that is more complicated than an all-off-premises Software as a Service or Infrastructure

as a Service offering. Instead, BaaS has cloud compo-
nents, but also requires on-premises agents, and re-
curring interaction between SP personnel and every
subscriber's IT team.

● **Local integrators (VARs)** are often looking to evolve
from transactional hardware-plus-software offerings
to services with recurring revenue and more strategic
value. But this may require urging the customer to rip
out the solution that very VAR installed. Moreover, not
every VAR is culturally equipped to stand up its own
infrastructure (even in a collocation facility) and offer
services.

Perhaps the answer for those SPs and VARs is to re-
consider jumping into the BaaS market, and instead to
enter the Monitoring (of Backups) as a Service market.
This is particularly ideal for VARs who have the techni-
cal staff to remotely manage/monitor the very backup

solution they installed earlier. The user doesn't have to
change their infrastructure, and their current investment
provides the capabilities and value they originally thought
they were purchasing.

Modern data protection is hard. Highly virtualized
and private cloud protection is much harder. The next
time you're dissatisfied with the results of your current
backup offering, first consider if you have the right visibil-
ity, including the correct processes and people, to ensure
the product can perform as advertised. Then think about
how much more strategic you could be if you didn't have
to babysit your backup solution because someone else,
who is legitimately more experienced, is doing it for
you. ∎

---

**JASON BUFFINGTON** *is a senior analyst at Enterprise Strategy
Group. He focuses primarily on data protection, as well as Windows
Server infrastructure, management and virtualization. He blogs at
CentralizedBackup.com and tweets as @Jbuff.*

READ/WRITE | ARUN TANEJA

# Flash is everywhere, but is all-flash best?

*The more flash the better, right? Maybe, but maybe not, depending on your applications' needs.*

O**VER THE PAST** three years the choices for <u>flash-based storage products</u> have skyrocketed. We now have PCI Express card and disk form-factor products for the server, caching devices that can be installed in the network, standard disk form-factor solid-state drives for traditional array architectures, caching in front of a traditional array and, of course, 100% flash-based arrays.

No storage array vendor wants to be left out of the "flash revolution." That's hardly a surprise. <u>Flash technology</u> is a godsend and its timing couldn't be better given the magnitude of the data to be stored, accessed, moved and analyzed in this era of Web 2.0 and log-data-spewing machines.

But with variety comes the difficult part of choosing

the right flash implementation for the job. This is further complicated by every vendor jockeying for leadership in this lucrative area and performance, endurance and other claims are often exaggerated, to say the least. So what's an overworked storage manager to do? It's a big topic that would require extensive discussion, so we'll <u>focus on all-flash arrays</u> and what kinds of application requirements they might be appropriate for. We'll look at other product alternatives and their implementation considerations in future columns.

## THE QoS ISSUE WITH ALL-FLASH ARRAYS

Recently, I've seen literature from some all-flash vendors claiming that the only way you'll be able to guarantee quality of service (QoS) for an application is if you use

an all-flash array. The thesis is that performance can be controlled independently of capacity level, and no adjacent application can ever cause a QoS issue for the application in question. When viewed strictly at the computer science level, the statement is hard to argue with. Since access to any part of flash is independent of whatever is

> **Well-designed hybrids can handle 80%-plus of the applications we normally deal with on a day-to-day basis, so they can be excellent and lower-cost alternatives to all-flash arrays.**

happening elsewhere, the IOPS (or throughput or latency) delivered to an application will be a constant, assuming correct sizing of the array for the application. Granted, there are some assumptions that latency and performance remain constant over time—that was indeed an issue in early flash systems—but most products have solved this problem. So the claim that QoS with no ifs, ands or buts can only be delivered by a 100% flash array is correct.

However, in most application environments, excellent QoS can be delivered by hybrid systems that use both flash and hard disk drives (HDDs), as long as the design

considered the QoS requirements from the outset. Mind you, this isn't true for all currently available hybrid systems. The best QoS control is realized with systems designed from scratch with flash *and* QoS in mind. Contrast this with systems where a flash drive is added simply as a replacement for a hard disk drive. Well-designed hybrids can handle 80%-plus of the applications we normally deal with on a day-to-day basis, so they can be excellent and lower-cost alternatives to all-flash arrays.

Of course, other system design principles still apply. In other words, you can't run 10 I/O-intensive applications as virtual machines and expect acceptable QoS for each one if the total IOPS delivered under the best circumstances fall short of that requirement. But that's also true for an all-flash array. Good design requires balancing the needs of the applications with the capabilities of the array, regardless of the physical configuration and media used. And if these principles are applied thoughtfully, you may not need an all-flash array.

### DOES DEDUPE LEVEL THE $/GB PLAYING FIELD?
Some all-flash array vendors claim their systems are so well designed—with built-in dedupe, for example—that their effective pricing per gigabyte of capacity is the same or lower than HDD-based systems. Their argument: Why bother with hybrids when you can get an all-flash system at the same price? However, based on what I've seen so

far, some all-flash systems do come close, but all-HDD or hybrid arrays are still less expensive than all-flash arrays, even with dedupe and compression assumptions figured

> **Typically, applications that are candidates for all-flash arrays are relentless in their need for IOPS without any periods of inactivity or reduction in system resource requirements.**

in. And be mindful of those all-flash arrays where performance is compromised when dedupe and/or compression is turned on; remember, you're considering an all-flash array for its performance and anything that gets in the way of that is heresy. The other factor to keep in mind is the type and sophistication of storage services available with the array. It's all those features that made the array useful in the first place.

The key point is to understand that most of your workloads will do just fine with well-designed hybrids. But there's no denying that when the application needs an all-flash array it needs it, so reserve them for those special purposes, at least for now. Typically, applications that are candidates for all-flash arrays are relentless in their need for IOPS without any periods of inactivity or reduction in system resource requirements.

It's easy to see a future where flash becomes pervasive. Some think HDDs will disappear completely. Even if that were to come to pass, it won't happen anytime soon. Pragmatically, we need to use flash intelligently alongside HDD-based systems for the foreseeable future. ∎

*ARUN TANEJA is founder and president at Taneja Group, an analyst and consulting group focused on storage and storage-centric server technologies.*

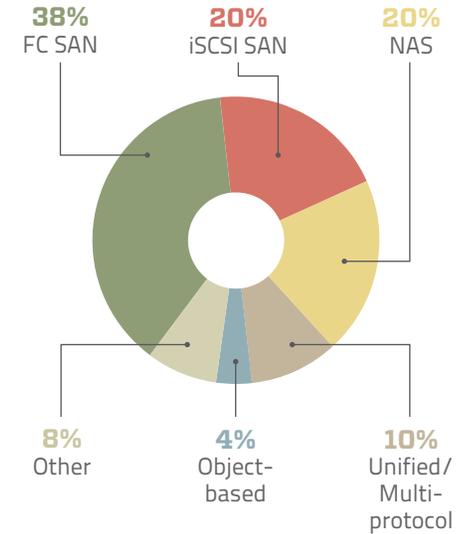# FC, NAS still main storage systems to handle growing data stores

With an average of 343 TB of active data, our survey respondents need a variety of data storage devices to cope with that volume. Sixty-nine percent park some data on NAS systems, while 65% use Fibre Channel (FC) SAN; iSCSI and unified systems round out the top four. FC SAN does most of the heavy lifting, with 38% of respondents' data residing on those storage devices. On average, respondents have 10 FC arrays and 9 NAS installed, as well as 6 iSCSI systems. That's a lot of hardware, but only 20% are tapping cloud storage for relief, stashing an average of 24 TB in the ether. While storage pros might be resigned to managing massive amounts of data, they're still looking for ways to pep things up, and 46% are doing it with solid-state storage. The vast majority (87%) of solid-state deployments are in hybrid arrays. Software-defined storage is in the news a lot these days: 27% have deployed it and they were DIY projects for 61%.  —*Rich Castagna*

**WHAT PERCENTAGE OF YOUR COMPANY'S TOTAL PRIMARY DATA CAPACITY (ACTIVE DATA) IS STORED ON THESE TYPES OF DEVICES?**

**38%** FC SAN

**20%** iSCSI SAN

**20%** NAS

**8%** Other

**4%** Object-based

**10%** Unified/Multi-protocol

**WHICH OF THESE TYPES OF STORAGE SYSTEMS DOES YOUR COMPANY HAVE INSTALLED? ***

| Storage system | Percentage |
|---|---|
| NAS | 69% |
| FC SAN | 65% |
| Disk backup systems | 64% |
| iSCSI SAN | 42% |
| Unified/Multiprotocol | 30% |
| Object-based | 16% |
| Other | 6% |
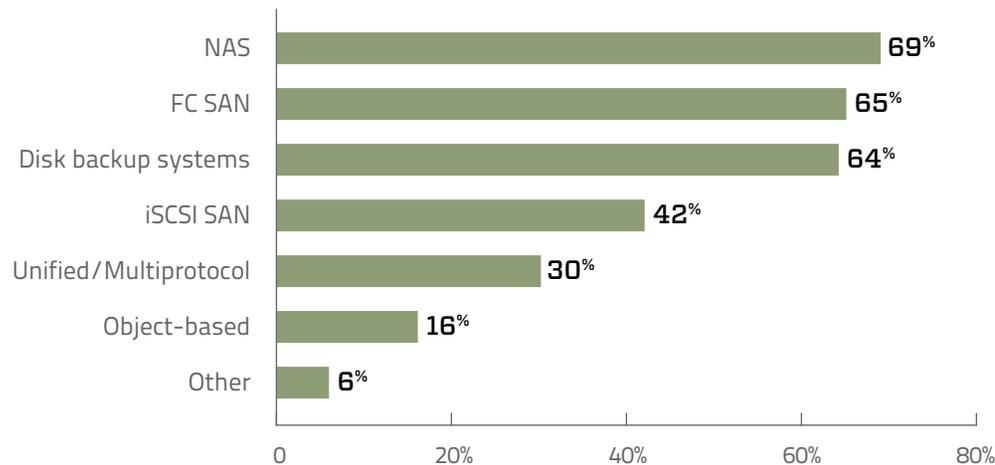
*MULTIPLE RESPONSES ALLOWED

# 343

**AVERAGE TBs OF ACTIVE STORAGE UNDER MANAGEMENT**

# TechTarget Storage Media Group

## STORAGE MAGAZINE

EDITORIAL DIRECTOR Rich Castagna

SENIOR MANAGING EDITOR Kim Hefner

EXECUTIVE EDITOR Ellen O'Brien

CONTRIBUTING EDITORS James Damoulakis, Steve Duplessie, Jacob Gsoedl

## SEARCHSTORAGE.COM

EXECUTIVE EDITOR Ellen O'Brien

SENIOR NEWS DIRECTOR Dave Raffo

SENIOR NEWS WRITER Sonia R. Lelii

SENIOR WRITER Carol Sliwa

SENIOR MANAGING EDITOR Kim Hefner

ASSISTANT SITE EDITOR Sarah Wilson

## SEARCHCLOUDSTORAGE.COM
## SEARCHVIRTUALSTORAGE.COM

EXECUTIVE EDITOR Ellen O'Brien

SENIOR MANAGING EDITOR Kim Hefner

ASSISTANT SITE EDITOR Sarah Wilson

## SEARCHDATABACKUP.COM
## SEARCHDISASTERRECOVERY.COM
## SEARCHSMBSTORAGE.COM
## SEARCHSOLIDSTATESTORAGE.COM

SENIOR SITE EDITOR Andrew Burton

MANAGING EDITOR Ed Hannan

ASSOCIATE SITE EDITOR John Hilliard

FEATURES WRITER Todd Erickson

## STORAGE DECISIONS TechTarget Conferences

EDITORIAL EVENTS MANAGER Jacquelyn Hinds

**About TechTarget:** TechTarget publishes media for information technology professionals. More than 100 focused websites enable quick access to a deep store of news, advice and analysis about the technologies, products and processes crucial to your job. Our live and virtual events give you direct access to independent expert commentary and advice. At IT Knowledge Exchange, our social community, you can get advice and share solutions with peers and experts.

COVER IMAGE AND PAGE 9: JEFFREY COOLIDGE/GETTY IMAGES