Managing the information that drives the enterprise

STORAGE

Vol. 8 No. 3 May 2009

Put dedupe to the test

Our resident backup expert, W. Curtis Preston, focuses on testing deduplication products to get real-world results and realistic expectations. ***page 9***

ALSO INSIDE

4 Tight budgets can drive creativity

18 Results of our exclusive Purchasing Intentions survey

28 Inside a storage tiering project

37 Dedupe choices abound

40 CDP picks up some steam

TechTarget
The Technology Media ROI Experts

The Glass Is Half Full

4 EDITORIAL In stressful times there's often an inclination to resort to stopgap measures: just patch it and pray, and deal with it later. But what could be a higher priority than using your diminished budget dollars to get more out of what you already have? by RICH CASTAGNA

Put Data Dedupe to the Test

9 Deduplication can drastically reduce the disk space required for backup, but different dedupe apps can yield very different results. Rather than accepting the deduplication ratios and performance vendors say they'll deliver, you should test these systems yourself using your own data.
by W. CURTIS PRESTON

The Time is Right for Efficient Storage

18 Storage budgets are lower this year, so storage managers will be even more careful with their purchases. With limited funds, they'll likely be looking for products that can help them manage their existing capacity more efficiently.
by RICH CASTAGNA

A Dynamic Four-Tier Storage Design

28 A comprehensive revamping of the storage environment at a major government agency shows how a tiered storage design can help meet operational expectations and plan for the future without busting the budget. by HERB FERGUSON

A Wealth of Deduplication Options

37 HOT SPOTS Hardware-based products propelled deduplication into the mainstream, so they've received most of the attention to date. But now many backup apps have integrated dedupe functionality, so you have more deduplication options.
by LAUREN WHITEHOUSE

CDP Picks Up A Little Steam

40 SNAPSHOT Continuous data protection (CDP) application adoption is hardly a juggernaut, but our most recent survey shows a modest increase in the number of CDP users. More respondents plan implementations, but some still aren't sold on CDP's benefits. by RICH CASTAGNA

Vendor Resources

41 Useful links from our advertisers.

SYMANTEC IS

THE MOST STORAGE FOR YOUR MONEY.

DISCOVER YOUR COMPANY'S UNTAPPED RESOURCES.

For more information on how you can save on storage, visit **go.symantec.com/sbs**

Confidence in a connected world.

symantec™

© 2009 Symantec Corporation. All rights reserved. Symantec and the Symantec logo are registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

The glass is half full

You can wring your hands over a shrinking storage budget or you can do something creative about it.

WHAT, ME WORRY?" says *MAD* magazine's Alfred E. Neuman, with his goofy, glazed-over grin. Move over, Alfred, I'm with you. No, I'm not subscribing to his haven't-a-clue outlook on life. I'm just doing my best to find the silver lining in an ever-so-gloomy economy, and I think it's time to try optimism and opportunities amid all the bad news.

And there's no shortage of bad news. Forrester Research recently published a report that predicts IT spending will drop by 3.1% this year in contrast to the 1.6% hike they had predicted earlier. (It's hard to figure out how they ever projected an increase.) Still, I'd argue that 3.1% doesn't equal a doomsday drop.

IDC also revisited some of its earlier prognostications for storage in 2009 and—surprise—rejiggered their numbers downward, too. The IT think tank says worldwide disk system spending will drop by 6.7% or, as they stated in their press release, "total disk storage systems spending will experience -6.7% year-over-year worldwide growth in 2009." Wait a second … a negative gain in growth? Talk about looking for a silver lining. However, IDC also predicts worldwide external disk system capacity will grow by 44% this year. True, it's lower than the traditional 50% to 60% annual growth, but it's not too shabby an estimate.

> But storage managers are smart. They think strategically. For the most part, they know they'll have to live with the decisions they make for a long time.

Even *Storage* magazine's own Purchasing Intentions survey (see "The time is right for efficient storage," p. 18) shows that on average your storage budgets are headed south by about 1.9% in 2009. That's the first time we've seen the budget number drift into negative territory, but it seems like a pretty shallow dip all things considered.

Clearly, things aren't as good as they could be or used to be. It's a serious situation, and I'm not making light of how the economic tailspin has affected lots and lots of people. I just don't think it will have that great of a negative impact on storage shops; indeed, I believe it could have a profound and positive effect.

Anyone managing storage operations this year isn't likely to be having

Copyright 2009, TechTarget. No part of this publication may be transmitted or reproduced in any form, or by any means, without permission in writing from the publisher. For permissions or reprint information, please contact Mike Kelly, VP and Group Publisher (mkelly@techtarget.com).

fun. You have new capacity demands and less money to spend than last year. What's new? For storage, and IT in general, budgets have to be twisted, turned and stretched every year. Do you ever remember saying anything like, "Whoa, there's way too much money in my budget"?

But storage managers are smart. They think strategically. For the most part, they know they'll have to live with the decisions they make for a long time. And some of the less strategic maneuvers made in the past, like continually throwing money at new arrays to stem the tide of capacity growth, just won't wash anymore. And that's a good thing, as Martha Stewart might say. That's why our survey results in this issue appear under such a positive headline. Being strapped for dough will make us all more resourceful, and keep us looking for more creative ways to deal with today's storage demands—which aren't so different from yesterday's and probably tomorrow's.

Being strapped for dough will make us all more resourceful, and keep us looking for more creative ways to deal with today's storage demands.

In stressful times there's often an inclination to resort to stopgap measures: just patch it and pray, and we'll deal with it later. It's about as normal a response as you can expect, even if it's only kicking the can down the street.

Some of you, however, will see all of this in a different light and recognize that tough times can also offer opportunity. Maybe this is the perfect time to do something about sprawling capacity, more than just buying additional disk, which requires more space, power and administrative attention. You could, for example, spend far less than what you would on a new array and buy a compression product that would squeeze your nearline storage down to more manageable proportions and free up some valuable space at the same time. Setting up an archive or an automated migration process to better tier your storage could yield equal (and impressive) results.

You can probably think of a dozen other things along those lines. And maybe you've been putting off a lot of them because there were higher priorities. But what could be a higher priority than using those diminished budget dollars to get more out of what you already have?

As perverse as it may seem, something as sobering as a recession can often reveal new opportunities, a chance to be creative and push the limits. Take it from a guy who just managed to quote Alfred E. Neuman and Martha Stewart in the same column. ⊙

Rich Castagna (rcastagna@storagemagazine.com) is Editorial Director of the Storage Media Group.

** Click here for a sneak peek at what's coming up in the June issue.*

SIMPLIFY YOUR STORAGE

TO IMPROVE OVERALL IT EFFICIENCY

INTRODUCING THE EQUALLOGIC™ PS6000 SERIES!

THE NEWEST ENTERPRISE-CLASS STORAGE SOLUTION FROM DELL.

Your storage solution should be simple, from initial purchase to scaling out your SAN. With a single, intuitive GUI for multiple group management, plus all inclusive features like Smart Copy for Microsoft® Hyper-V™* and support for VMware® vStorage, the EqualLogic PS6000 Series iSCSI SAN streamlines and simplifies management while protecting your data. What's more, its patented "peer" architecture allows for on-demand scaling, enabling a true pay-as-you-grow iSCSI SAN solution to help improve overall IT efficiency.

SIMPLIFY YOUR STORAGE AT DELL.COM/PSseries/PS6000

* Availability expected by Q2, 2009

COMING IN JUNE

Storage, Heal Thyself

Self-healing storage emerged in 2008 and, understandably, has generated considerable interest among storage managers. The first question is "Do these systems really work?" while the next is "How do they do it?". We look at disk systems touting their self-healing capabilities and describe the methods employed by each one.

Dueling Directors

Handling storage growth without creating storage networking islands is a challenge, but high-capacity director-class switches can help cap that kind of sprawl. The two heavyweights of storage networking, Brocade and Cisco Systems Inc., want to control your SAN traffic. We look at how their products compare and differ, as well as each company's plans for a unified data center network.

State of Storage Tiering

From HSM to ILM, storage tiering has been promoted as the best way to ensure that you don't overspend on storage. Putting the right data in the right place makes a lot of sense, but it's not as easy as it may sound. We surveyed *Storage* readers to gauge the progress they've made with tiering, and to see what roadblocks may be slowing their progress.

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

Associate Editor
Christine Cignoli

Creative Director
Maureen Joyce

Contributing Editors
James Damoulakis
Steve Duplessie
Stephen Foskett
Jacob Gsoedl

SearchStorage.com

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
Christine Cignoli

SearchStorageChannel.com

Site Editor
Susan Troy

SearchDataBackup.com
SearchSMBStorage.com
SearchDisasterRecovery.com

Site Editor
Andrew Burton

Associate Site Editor
Heather Darcy

Assistant Site Editor
Chris Griffin

Features Writer
Todd Erickson

Executive Editor and Independent Backup Expert
W. Curtis Preston

Storage Decisions

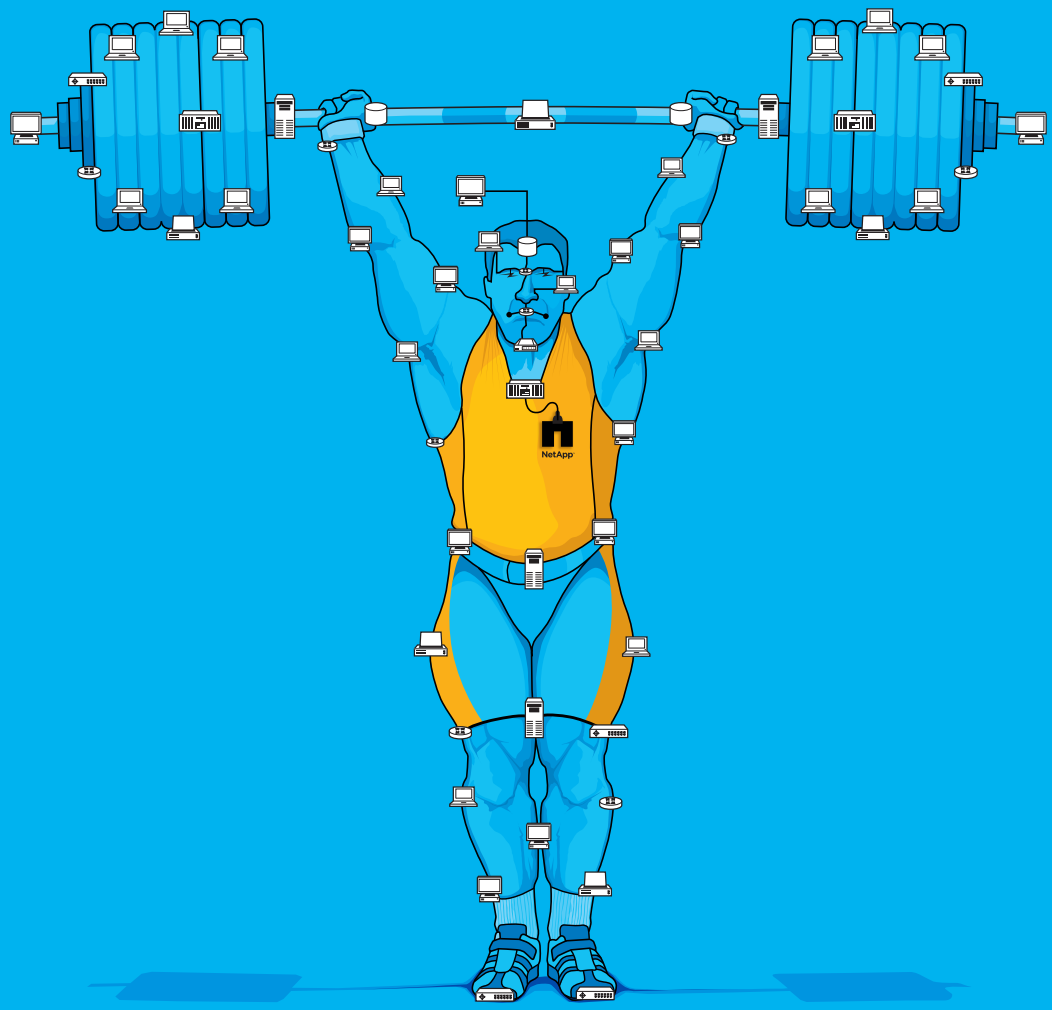
TechTarget Conferences

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



With NetApp® at the heart of your business, you can

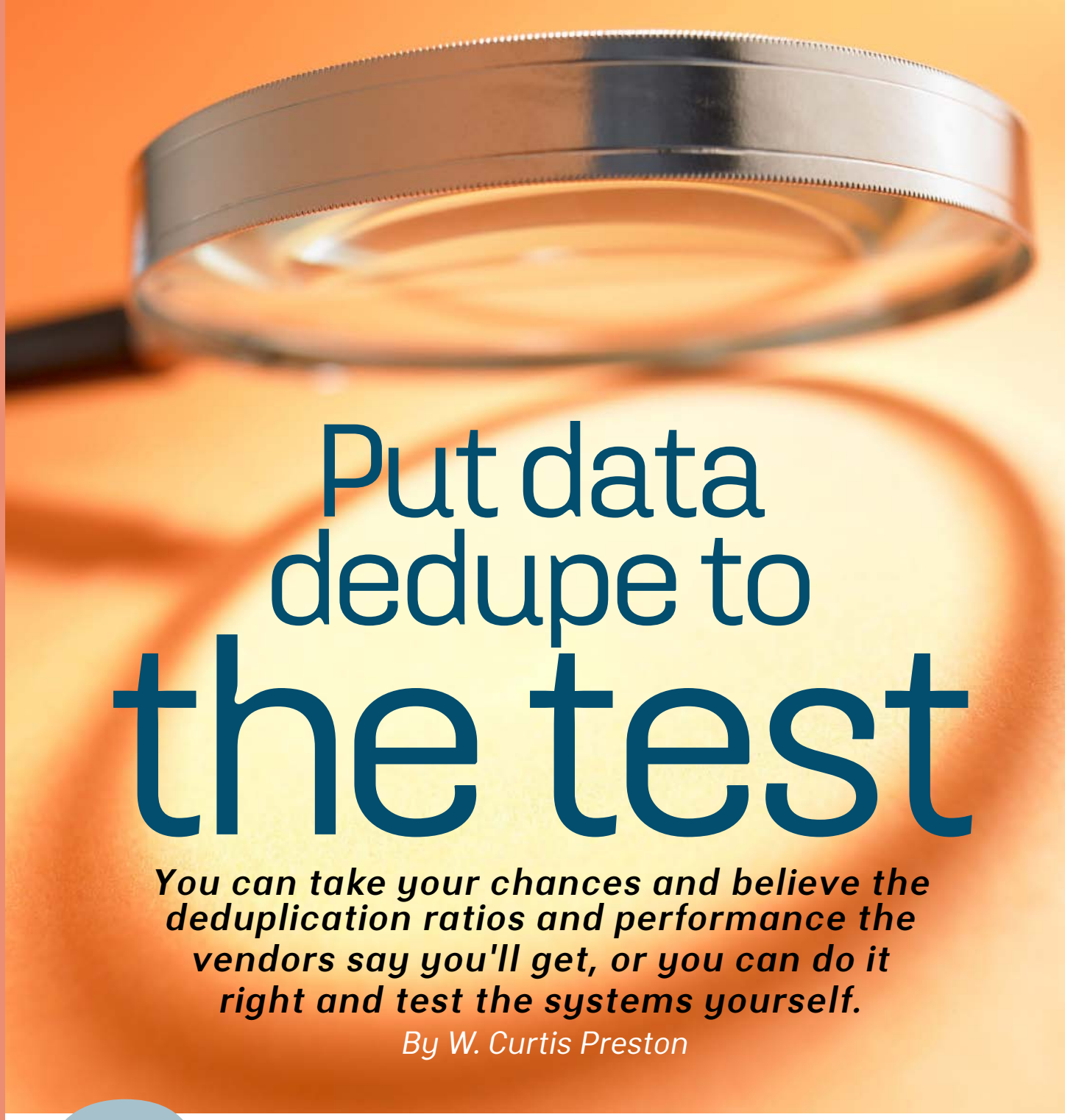
EFFORTLESSLY HOLD YOUR DATA

at a fraction of the cost and footprint.

Imagine storage and data management solutions smart enough to support the data you need, and not a lot of dead weight. It's possible when you partner with NetApp. Our industry-leading solutions use deduplication and other space-saving technologies to help you store data efficiently and reduce your footprint by 50% or more. So you can manage exponential growth while minimizing your storage investment—all with the support of a team that will exceed your expectations. See how we can help your business go further, faster. Find out how you can use 50% less storage, guaranteed,* at netapp.com/efficiency.

NetApp™
Go further, faster

© 2009 NetApp. All rights reserved. Specifications are subject to change without notice. NetApp, the NetApp logo, and Go further, faster are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. *This guarantee and related Program is limited to the terms set forth in the Program Guide and Acknowledgement For 50% Virtualization Guarantee Program document, applicable only to prospective orders placed after the Program effective date and is dependent upon your compliance with the terms and conditions set forth in this document and any of the instruction sets and specifications set forth in the referenced documents. NetApp's sole and exclusive liability and your sole and exclusive remedy associated with the terms of this guarantee and related Program is the provision by NetApp of the additional storage capacity as set forth in this guarantee and related Program.



Put data dedupe to the test

You can take your chances and believe the deduplication ratios and performance the vendors say you'll get, or you can do it right and test the systems yourself.

By W. Curtis Preston

SOME DATA DEDUPLICATION VENDORS are lying to you. Although I knew this, it became immediately apparent after the feedback I received to my BackupCentral.com blog post on deduplication performance. All I did in that article was compile and interpret publicly available information, but some people felt I was validating those vendors' claims despite a disclaimer to the contrary.

I received public and private comments from vendors and users alike that said, "How can you say vendor ABC does *xxx* MB/sec? We've

never seen more than half that!" or "Have you asked that vendor for a reference? I doubt a single customer is using the configuration you listed in your article!" Suffice it to say that some of those numbers, while openly published by the vendors, are complete and utter fiction. Which ones, you ask? Given the confusion about published statistics and the lack of independent testing of these products, the only way you're going to know is to test the product yourself. The following explains what you should test, and what I believe is the best way to conduct those tests.

Two types of dedupe

TARGET DEDUPLICATION
Data deduplication is done in an appliance that sits inline between the backup server and the backup target. The appliance receives the full backup stream and dedupes the data immediately.

SOURCE DEDUPLICATION
Backup software performs the deduplication on the backup client and the backup server before sending data to the backup target. This approach has less impact on the available bandwidth.

TARGET VS. SOURCE DEDUPE

There are two very different types of dedupe, and they require very different testing methodologies. With target dedupe, deduplication occurs inside an appliance that accepts "regular" backups, that is, backups from a traditional backup application. These appliances usually accept those backups via a virtual tape interface, NFS, CIFS or other proprietary API, such as the Open Storage (OST) API in Symantec Corp.'s Veritas NetBackup. Backups are received in their entirety and are deduped once they arrive. Target dedupe saves disk space on the target device, but does nothing to reduce the network load between backup client and server. This makes target dedupe more appropriate for environments where bandwidth utilization isn't the primary consideration, such as in a centralized data center.

Source deduplication products require custom backup software at the backup client and backup server. The client identifies a unique chunk of data it hasn't seen before, and then asks the server if it's ever seen the chunk of data before. If the server has backed up that same chunk of data from that (or another) client, it tells the client not to send the chunk over the network and simply indicates that the chunk was found in another location. If the chunk is determined to be truly unique, it sends the chunk across the network and records where it came from. This makes source dedupe most appropriate for environments where bandwidth is the primary consideration, such as remote offices and mobile users.

TESTING TARGET DEDUPE

There are three things to verify when considering a target dedupe solution: cost, capacity and throughput. When considering the cost of deduplication systems (or any system for that matter), remember

to include both capital expenditures (CAPEX) and operational expenditures (OPEX). Look at what hardware and software you'll need to acquire to use a particular appliance to match a given throughput and capacity model. Some dedupe vendors make it very easy to arrive at a CAPEX number: for example, you need to store 30 TB of data, and you back up 5 TB/day, so you need model *x*. It includes all the computing and storage capacity you need to meet your requirements. Other vendors just provide a gateway that you can connect to your own storage. Finally, some vendors provide just the software, leaving the purchase of all hardware up to you. Remember to include the cost of the server hardware in this configuration, making sure that you're specifying a server configuration that's approved by that vendor. In both the gateway- and software-only pricing models, make sure to include the cost of the disk in your comparison even if it's "free." The dedupe pricing world is so unique that there are scenarios where you can actually save money by not using disk you already have.

One final cost element: Remember to add in (if necessary) any "extra" disks, such as a "landing zone" (found in post-process systems), a "cache" where data is kept in its original format for faster restores or any disks not used to store deduplicated data. All of those disks should be considered in the total cost of purchasing the system.

You then need to consider OPEX. As you're evaluating each vendor, make note of how you'll need to maintain their systems and how the systems will work with your backup software vendor. Is there a custom interface between the two (e.g., Veritas NetBackup's OST API), or will your system just pretend to be a tape library or a file system? How will that affect your OPEX? What's it like to replace disk drives, disk arrays or systems that are part of this system? How will global dedupe, or the

DEDUPE TESTING TIPS

USE REAL DATA: To get accurate results, all testing should be done using copies of the data you actually back up.

TRY RESTORES, TOO: It's not enough to just test backup performance, you should also test dedupe products by doing typical restores.

RECREATE REPLICATION: It's likely you'll also replicate backup data to a disaster recovery site or vaulting facility, so you should test how well—and how quickly—a dedupe product handles replication.

TALLY COSTS CORRECTLY: The actual price of the dedupe product might not reveal the total bill for the solution. Be sure to include the cost of any new disks or disk systems, as well as software upgrades that may be required to implement the product correctly.

Agentless backup is not a myth.

Asigra.

www.RecoverYourCool.com

lack of it, affect your ability to scale the product to meet your needs?

There are two ways to test capacity. The first is to send a significant amount of backups to the device and compare the size of those backups with the amount of storage they take up on the target system. This will show your dedupe ratio. Multiply that ratio times the disk capacity used to store deduped data and you'll get your effective capacity. The second method is to send backups to the device until it fills up and then record how many backups were sent. The latter method takes longer, but it's the only way to know how the system will perform long term. (The performance of some systems decreases as they near capacity.)

Finally, there are several things you should test for performance.

Ingest/Write. The first measure of a disk system (dedupe or not) is its ability to ingest (i.e., write) backups. (While restore performance is technically more important, you can't restore what you didn't back up.) Remember to test both aggregate and single-stream backup performance.

Restore/Copy/Read speed. The second measure of a disk system (dedupe or not) is its ability to restore or copy (i.e., read) backups. I like to point out that the whole reason we started doing disk-to-disk-to-tape (D2D2T) backups was to use disk as a buffer to tape; therefore, if a disk system (dedupe or not) isn't able to stream a modern tape drive when copying backups to tape, then it misses the point. Remember to test the tape copy where you plan to do the tape copy; for example, if you plan to replicate to another system and make the tape there, test that. Finally, don't assume that restore speeds will be fine, and remember to test both single-stream and aggregate restore performance.

Deduplication. Once the data arrives in its native format to the device, it must be deduped. Inline boxes dedupe the data the second it arrives. The original data never hits disk; therefore, an inline vendor's dedupe speed is the same as its ingestion speed. Post-process vendors can take from seconds to hours to dedupe data. You'll have to investigate how long the dedupe process actually takes.

Replication. Your dedupe ratio comes into effect with replication as well. The better your dedupe ratio is, the fewer blocks will have to be replicated. But the only way to know for sure how replication will work is to actually do the replication. Observe how many blocks of data are replicated and note when the replication starts and stops. You may be able to capture this data from the dedupe vendor, but to test it yourself you may need to use a network tool to get this information. Remember that not all vendors start replicating at the same time. Of course, nothing can be replicated until it's deduped, but don't assume that an inline vendor will replicate backups immediately after they're deduped; many vendors will wait until a given tape is no longer being used or a file is closed (in the case of NAS).

TEST WITH PRODUCTION DATA, BUT...

You must test target dedupe systems by storing your actual production backups on them. However, don't test your dedupe system by backing

up production systems directly to it. Vendors would love for you to test that way, as it's hard to give the system back when you're using it to store real backups needed for real restores. It's always a bad idea to use a test system in production.

So how do you test dedupe systems with production data without backing up production systems to them? It's simple. Copy your production backups from tape or disk to the dedupe system. When testing restore/copy speed, copy backups from the deduped device to disk or tape because the "reconstitution" process the dedupe system has to go through for a copy is exactly the same as what it does for a restore.

You must test target dedupe systems by storing your actual production backups on them.

Determine how long you plan to store your backups in the dedupe system. In my opinion, if you plan to store 90 days of backups in your dedupe system, that's how many days of backups you should store in your test system. (It won't take 90 days to store 90 days' worth of backups.)

If you plan on testing 90 days of backups, pick a period of 90 days that starts with a full backup (or an IBM Corp. Tivoli Storage Manager backup set) and continues for 90 days. If you're testing multiple dedupe systems, make sure to use the same set of backups with each test (*ceteris paribus*—with all other factors or things remaining the same). Copy the first full backup (or backup set), followed by backups that are 89 days old, then 88 and so forth. Do that until you've worked your way up to 90 days.

Each simulated "backup day" should include a single backup (i.e., one backup set copied until it's complete), simultaneous backups (as many simultaneous copies as you have tape drives), deduplication and replication. If possible, the simultaneous backups should supply enough throughput to reach that system's maximum throughput. Once all of those activities have completed, the next day's "backups" can continue by copying the next set of backup tapes into the system.

The beginning of each simulated "backup week," including the first one, should include a number of simulated restore tests. The best way to test restore speed is to actually copy a predictably sized backup set from the dedupe system to tape. You should do two single restores by themselves (i.e., one backup set copied from the dedupe system to tape until it's complete), followed by two sets of simultaneous restores (as many simultaneous copies from the dedupe system to tape as you have tape drives). The reason you should copy two sets in each test is that you want to copy from the oldest and newest backups in each test cycle. What you're looking for with these tests is a difference in restore time from older backups in relation to newer backups, and from backups when the system is relatively empty to when the system is relatively full.

One key to doing this right is automation. This will allow you to do

testing around the clock and will provide the best way of documenting the timing of all activities. Automating things is also the key to *ceteris paribus*, which is absolutely essential when testing multiple systems. If possible, another approach is to use a completely separate backup server and tape library. That will isolate the test from the backup traffic, both for the sake of production backups and to ensure that production backups don't impact the test.

TESTING SOURCE DEDUPE

In most cases, source dedupe is being considered as a replacement for some backup software that's already "doing the job" via backups to an inexpensive tape device. While that configuration comes with a lot of drawbacks that source dedupe intends to fix, the fact that you're replacing an existing product creates a greater burden of proof on the source dedupe product.

Basic backup functionality. You'll be using this product to perform all backups and restores for supported clients. Make sure you try everything with this product that you currently do with your backup system. Schedule automated backups and see how it reports their success. If any of them fail (and you should force some of them to fail), what happens next? What's it like to rerun failed backups? What are restores like? Can the administrator do them or can users do them? Use the same workflows you're accustomed to using and see if they can be adapted to this new product.

Advanced backup functionality. Do you plan to replicate these backups to a second location? Once you've replicated all backups to a centralized location, do you plan to copy some or all of them to tape? How does that work?

Performance. What kind of backup performance do you get? How fast is the replication? How much data is sent across the wire? (Don't assume that two different deduplication products will send the same amount of data over the wire.) If you're planning on replicating across long distances, how do latency and an unreliable connection affect the overall performance and stability of the product?

As with target dedupe, there's no substitute for real data during testing. But unlike target dedupe, it can be very difficult to reliably test these products doing anything short of backing up the types of systems you plan on actually backing up. You can back up test systems, but the test is only valid if you can simulate user activity, such as emails to the Exchange database, and new and updated files in the file system. Without those changes, your source dedupe system will perform very well, but will offer no insight into how it's going to perform in the real world.

Most people can't simulate real user activity on a large test environment, so their only alternative is to back up real systems. Once you've verified in a test environment that the software in question can run on the types of systems you'll be testing, you need to begin a proof of

concept on "real" systems that will represent the types of systems you'll be backing up. To minimize risk, it's best to start with systems that aren't currently being backed up and don't have a mission-critical uptime requirement, such as laptops. Select a few users to pilot the software, make sure they're aware it's a pilot and ask them to report their experiences to you. Once you've logged a little time with those types of systems, you can expand to file servers at a remote site, followed by application servers (such as Exchange). Just remember that each time you start backing up a new type of system, you risk negatively impacting the stability or performance of that system—so you must watch for any instabilities during each test.

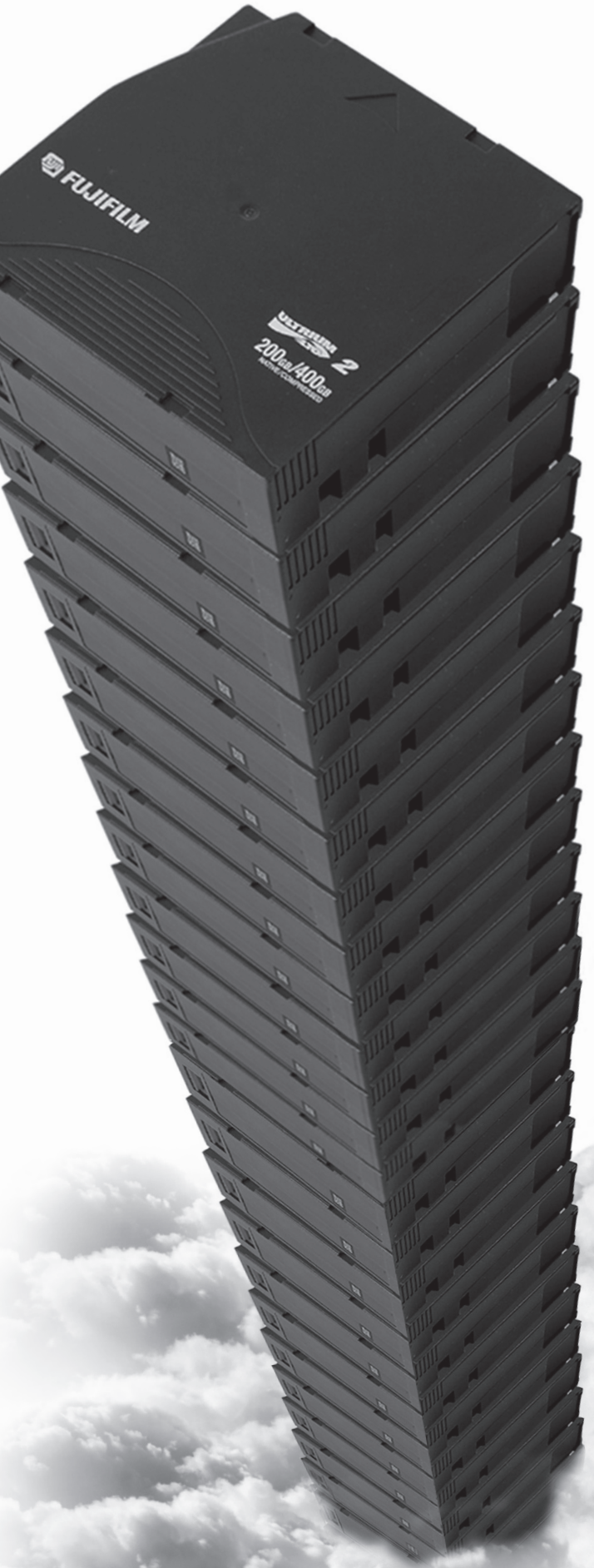
Any systems already being backed up in the proof-of-concept test should continue being backed up via the previous method until you're in production with the new system. If they're Windows systems, you must verify that the two programs won't interfere with each other by resetting and/or using the Windows archive bit. The worst-case scenario would be if they're both using it, as new or modified files would get backed up by the next backup product to run and wouldn't get backed up by the following product. You must verify how the archive bit will affect two products running in parallel.

Most people can't simulate real user activity on a large test environment, so their only alternative is to back up real systems.

Make sure to get answers for all of these questions. Also simulate all of the things that are likely to happen, such as a laptop user suspending their laptop in the middle of a backup, an Ethernet cable being unplugged or an Internet connection timing out. The hardest question to answer may be how many bytes are actually sent across the network, so you may need third-party network monitoring software to get a verifiable number.

Nothing proposed here is easy. However, the potential risks of buying dedupe systems without proper testing are simply too great to consider skipping testing. With some dedupe vendors possibly exaggerating their products' prowess, testing is the only way to separate truth from fiction—and probably save some money in the process. ⊙

W. Curtis Preston is an executive editor at SearchStorage.com and an independent backup expert.



Stop spending extra money on tapes and offsite storage for your IT data.

Implement data deduplication solutions for your environment.

Our data de-duplication assessment services will design the best solution for your environment based on data files trend results, current backup solutions, and remote site solution needs

CDI's Implementation Services Include:

- EMC's DL3D services with Legato Networker for file system and databases
- EMC's Avamar services for VMWare VCB, File Systems, and databases (Exchange and SQL)
- Symantec's services using Puredisk solution for file system and databases (Exchange & SQL)

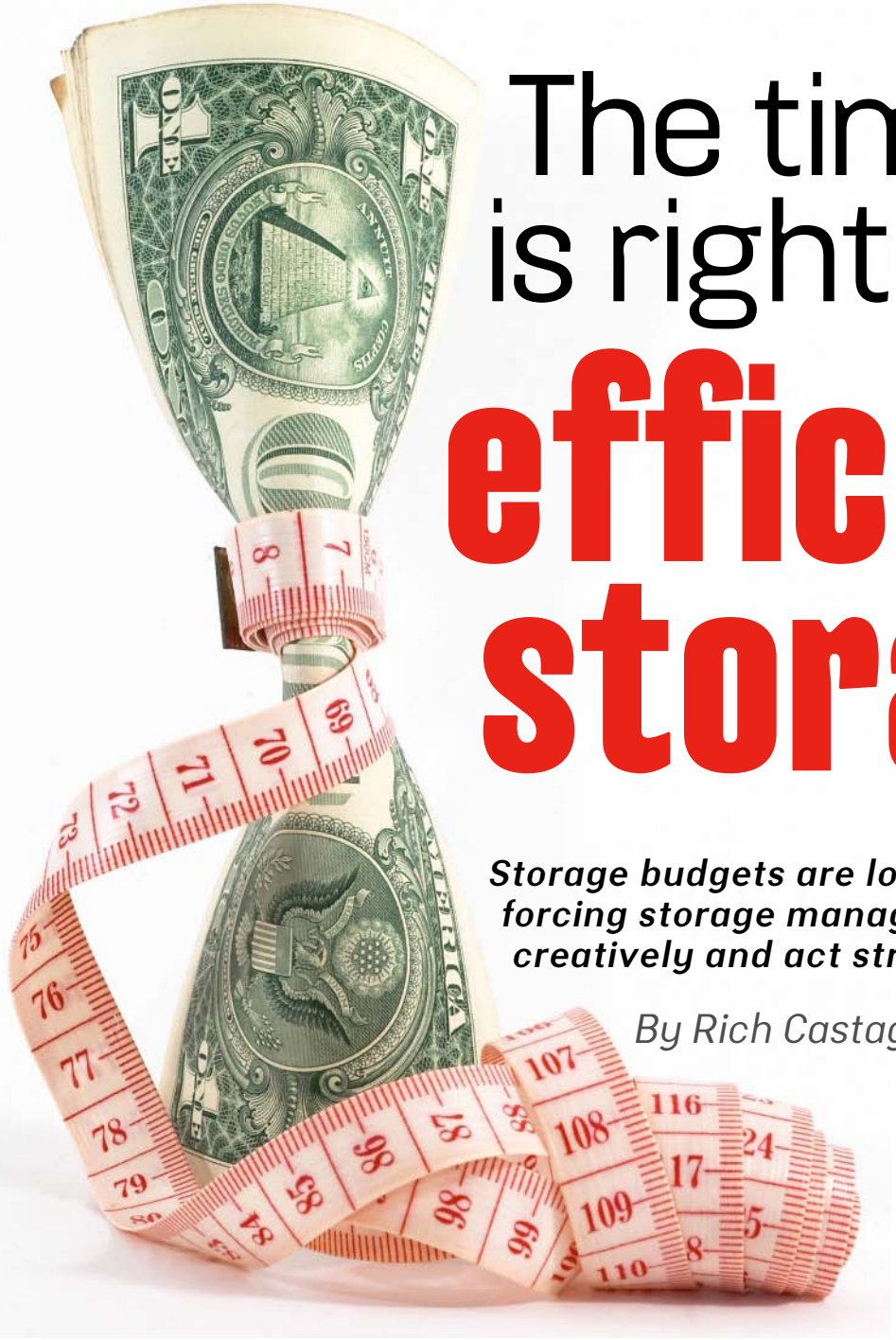
CDI is the partner who was involved in the single, largest purchase of deduplication hardware in the world – to date.

COMPUTER DESIGN & INTEGRATION LLC

EMC² where information lives®

www.cdillc.com
New York 500 Fifth Avenue, Suite 2720, New York, NY 10110 212.575.4860
New Jersey 696 Route 46 West, Teterboro, NJ 07608 201.931.1420

© 2009 Computer Design & Integration LLC



The time is right for efficient storage

Storage budgets are lower this year, forcing storage managers to think creatively and act strategically.

By Rich Castagna

WHAT DO ARCHIVING, thin provisioning, data deduplication and network-attached storage (NAS) gateways have in common? They help you shoehorn a lot of data into a little space without breaking the bank. And in 2009, that bank appears to be a little less flush than in recent years, so more creative approaches to storage management will be essential.

It's been said that there's always a budget crisis in IT, but 2009 might prove to be a particularly memorable year. For the first time in the seven years we've been conducting our Storage Purchasing Intentions survey, respondents indicate they'll have less money to spend on storage than they did last year. On average, budgets will dip nearly 2%, more than a five point swing—the biggest change we've ever seen—from the modest but positive increases we saw

last spring and fall (see "Storage budgets dip in 2009," p. 23). Over the years, we've seen lower budget indications in our spring surveys, with budgets bouncing back a bit when we field the fall version of the survey.

While 29% of the respondents surveyed say their company's storage budget will actually rise this year, the number falls far short of last spring's 46%. At the other end of the budget spectrum, 37% say their storage budgets will be lower than last year's, a big jump from the 16% and 17% reported on the 2008 surveys. And you're bound to feel the pinch, whether your shop serves a small, midsized or enterprise-scale company.

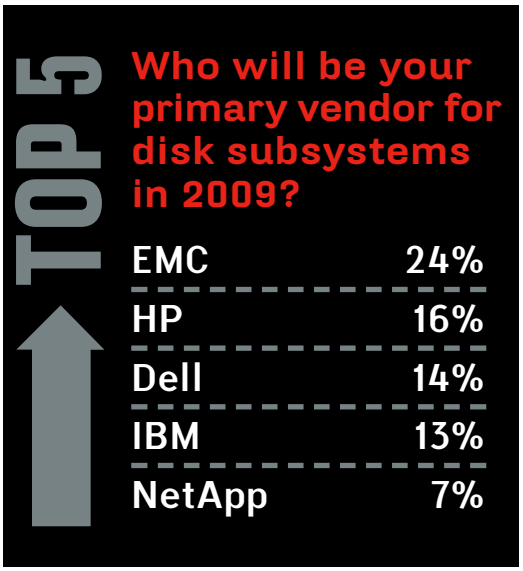
> "For this year, the budget is still OK, but they don't want to commit to next year or the year after that."
>
> —Jason Jed, network manager, Office of the President, University of California (UC) in Oakland

Hilary Tullier, business systems manager at Houston-based ACT Pipe and Supply Inc., hopes to see an increase in his storage budget this year. "Last year it was pretty much static," he said. "It was just break and repair." He's evaluating arrays to replace more than 10 TB of direct-attached storage (DAS) and is favoring iSCSI over Fibre Channel (FC) because "it doesn't really break the budget."

Similarly, Jason Jed, network manager in the Office of the President at the University of California (UC) in Oakland, expects to push ahead with a project to add approximately 12 TB of storage and move his systems to a local collocation facility while mirroring to another UC site in San Diego. "For this year, the budget is still OK," he said, "but they don't want to commit to next year or the year after that."

But for some, plans may have to be deferred for now. Lance Wyatt, information systems manager at United Steelworkers in Pittsburgh, said he was evaluating a move from DAS to networked storage "before the economy tanked." While DAS is serving the company well for now, any significant changes will have to wait. "We're probably looking at maybe third quarter [of] 2010, probably 2011," he said.

In an ironic twist, the average storage budget of $2.9 million is approximately 12% higher than last fall, but about 9% lower than spring 2008. For a lot of companies, a storage budget of that size would be the stuff of dreams; however, the majority of our survey participants (56%) make do with one-third of that amount or less.



That's why archiving, thin provisioning, data dedupe and NAS gateways drew particular interest from this year's respondents, as they all address using existing storage resources more efficiently while perhaps delaying or scaling back new purchases.

ABOUT THE SURVEY

The *Storage* magazine/SearchStorage.com Purchasing Intentions survey is fielded twice a year; this is the seventh year the surveys have been conducted. *Storage* magazine subscribers are invited to participate in the survey, which gathers information related to storage managers' purchasing plans for a variety of storage product categories. This edition had 777 qualified respondents across a broad spectrum of industries, with the average company size measured as having revenue of $1.6 billion.

NOT EVERYTHING IS DOWN

Budgets may be tough to predict year over year, but storage managers can make one prediction with 100% certainty: Their need for disk capacity will continue to grow. This year, respondents will add an average of 43 TB to their existing disk environments, which is approximately 10% more than what they planned on last fall, but lower than the 47 TB they predicted in spring 2008 (see "Expected new disk capacity," p. 23). Either way, it's a lot of capacity to add and the second highest number we've seen in the surveys.

Larger companies (those with revenue of $1 billion or more) expect to add a whopping 75 TB in 2009. But smaller companies, with their more modest capacity requirements, expect to add an average of just 23 TB.

With more than 300 TB of disk capacity already installed, Troy Downing, systems analyst II at Rain and Hail L.L.C., a farm property insurance firm in Johnston, Iowa, will be looking to add more disk. "Right now we've been figuring about 10% to 15% growth per year," he said. "My guess is that probably toward the end of the year we may end up adding another array." Some of that expansion is due to a virtual desktop infrastructure project slated to roll out this year.

At Horace Mann Educators Corp., an insurance company for the education community in Springfield, Ill., Thomas Janssen, director of IT, expects the company to add some capacity to a few of its installed EMC Corp. arrays. "The disk storage that we would be purchasing in 2009 is primarily going to be for tier 2 storage for the Clariion," said Janssen. "We're probably looking at another 30 TB this year."

Regardless of how much money is available to spend and ever-rising capacity needs, over the years we've seen little movement in how storage managers break down their budgets. This time, it's more of the same, with the biggest budget chunk (39%) earmarked for disks and disk systems. With diminishing budgets and rising demand for capacity, one would expect that percentage to grow year to year, but those conditions are largely offset by the dramatic price cuts we've seen for disks and disk subsystems over the past few years.

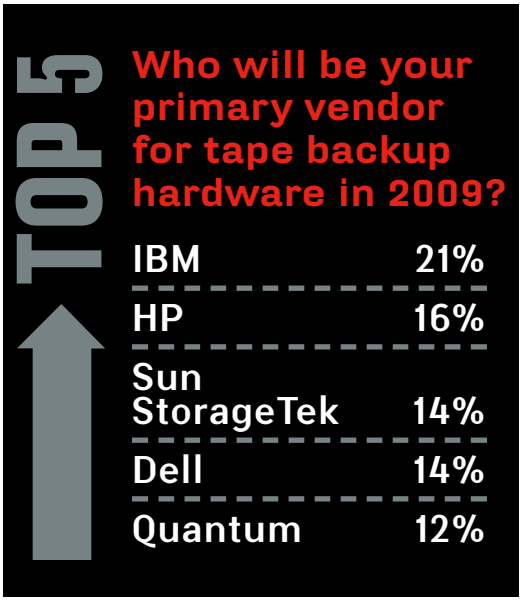
NO SKIMPING IN SOME AREAS

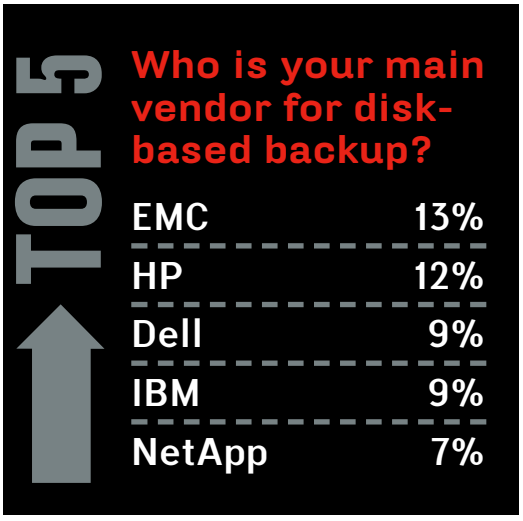
The types of disk systems our respondents will shop for hasn't changed a lot. Midrange systems have broad appeal for companies of all sizes and still grab the most attention, with 45% of those surveyed opting for this system class. Over the past several years, midrange systems have managed to increase their turf by going "down market" to offer smaller companies enterprise-class features while extending further into enterprises as cost-effective, high-density alternatives.

While it might be a sign of the times that one respondent said their primary disk storage vendor for 2009 would be eBay, it doesn't appear that many storage managers are ready to cut corners when it comes to arrays. Fibre Channel arrays and NAS systems top the currently installed list, with 65% and 64% of respondents, respectively, saying they have those types of storage. iSCSI, once touted as the rising star of storage, has indeed risen, with 32% of respondents saying they've deployed iSCSI arrays vs. 27% last spring. iSCSI doesn't appear to be riding the coattails of server virtualization as some predicted, as more than 50% of those surveyed are using FC for their virtual server environments vs. approximately 12% using iSCSI—numbers virtually the same as those we saw last fall.

But there's not much evidence that iSCSI storage will emerge as a low-cost alternative to Fibre Channel during these tough times. That said, one of iSCSI's strongest selling points is still its price, with 33% of those surveyed going for iSCSI because it's cheaper than FC, and another 15% just looking for cheaper storage in general. The bottom line for storage buyers is still the features and functions that disk systems offer; 33% of respondents say that was their key purchasing criteria, which is the highest number we've seen in two years. Surprisingly, price was the prime consideration for just 16% of those surveyed, which is 12 points lower than reported a year ago.

About 40% have already or will install iSCSI storage this year, no real change from the last two surveys. Whatever territory iSCSI storage is winning, it's picking up an inch or two at a time, with only modest increases in those saying they have plans for iSCSI compared to last spring. Mid-sized companies seem to show the most interest, with 44% saying they're going with iSCSI vs. 38% a year ago.





ACT Pipe and Supply's Tullier hopes to add two iSCSI arrays. "We have a DR site right now that's really a cold site, and I'd like to make it more of a warm or even hot site by putting in two iSCSI arrays and having them do continuous replication," he said.

There doesn't seem to be a knock against iSCSI for its performance or reliability: only 18% say performance deterred them (the same as last fall), while 11% state it was reliability (vs. 14% in the fall). In fact, 44% of respondents who have deployed iSCSI say they have mission-critical applications running on their systems—a jump of 13 percentage points vs. spring 2007, with much of that increase coming from large companies (see "How iSCSI storage will be used," p. 23).

MORE EFFICIENT STORAGE

If there's any silver lining to this economic cloud, it's that storage managers are redoubling their efforts to run their storage more efficiently. While proponents of storage virtualization have long extolled its greater efficiencies, most shops have shied away from often difficult and costly implementations. Twenty-seven percent of the current survey respondents have virtualized at least some of their storage (up one point from last fall) and another 21% plan virtualization evaluations this year.

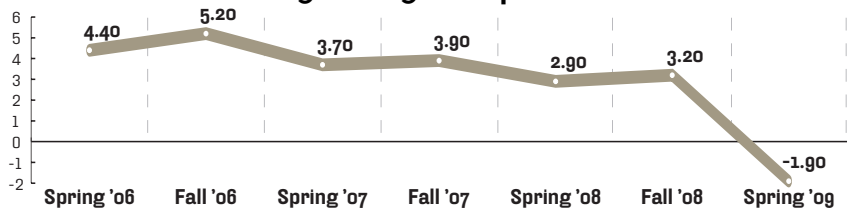
If there's any silver lining to this economic cloud, it's that storage managers are redoubling their efforts to run their storage more efficiently.

A one-point hike isn't statistically significant, but other indicators suggest storage virtualization is being considered a little more seriously these days. Compared to last fall, the number of respondents with planned 2009 purchases was up in four of the five storage virtualization technology categories. Software solutions were favored, with 14% of those surveyed planning to acquire software to run virtualization in their array and 10% opting for host-based storage virtualization. Plans to buy storage virtualization appliances dipped to 11%, a mere point lower than last fall.

Thin provisioning—another very effective way to conserve disk dollars—is also getting more interest. On another survey (Storage Priorities for 2009, fielded late October 2008), thin provisioning was the top pick

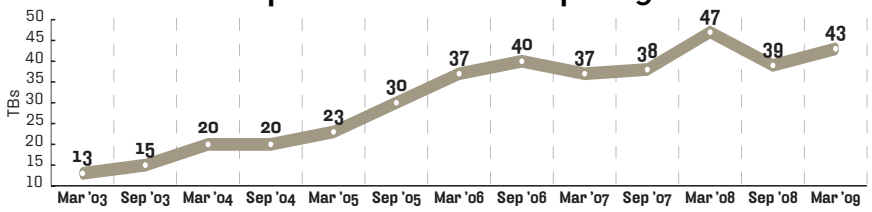
BY THE NUMBERS

Storage Budgets Dip in 2009

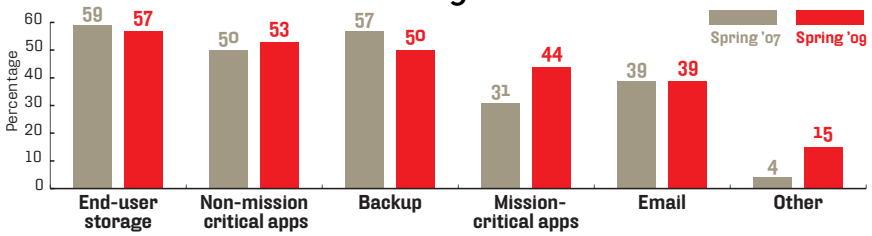


Percent change in storage budget compared to previous year

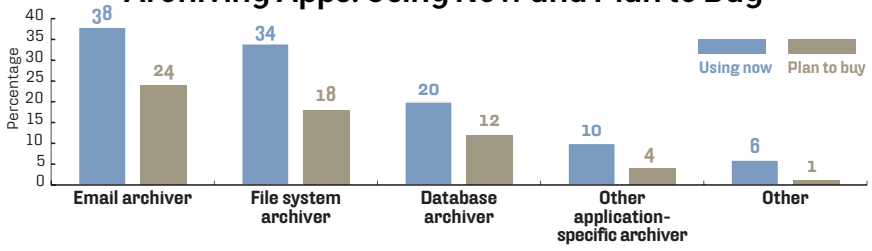
Expected New Disk Capacity



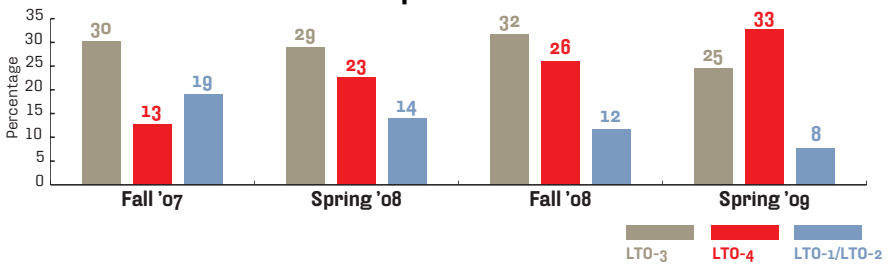
How iSCSI Storage Will Be Used



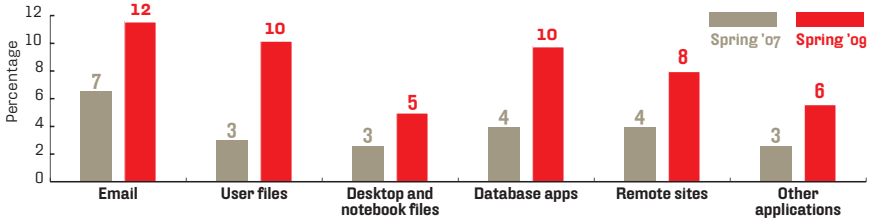
Archiving Apps: Using Now and Plan to Buy



LTO -4 Adoption on the Rise



Increased Interest in Cloud Backup Services



among disk-related technologies, with 55% of respondents planning to deploy or evaluate it in 2009. Thin provisioning is a relatively low-cost array option considering its potential benefits; in some cases, thin provisioning capabilities are built in and just waiting to be turned on. Rain and Hail's Downing is particularly interested in Hewlett-Packard Co.'s new Enterprise Virtual Array 8400 (EVA 8400). "It's my understanding that they're going to have some thin provisioning," he said.

Just getting data on the right type of storage can reap significant savings and free up prime high-performance capacity. For Horace Mann's Janssen, that'll be a key project in 2009. "We'll be doing a lot of migrating off tier 1 into the appropriate tier this year," he said. This will require some additional storage shelves for the firm's Clariions, but no array acquisitions are planned. "Those all came at the very end of last year," said Janssen.

TOP 5

Who will be your primary vendor for backup software in 2009?

Symantec	38%
IBM Tivoli	17%
EMC	10%
CommVault	7%
HP	5%

Archiving can also free up expensive primary storage capacity by moving less frequently used or unused data to cheaper disk or tape. Like thin provisioning, archiving is more of a cost-reduction maneuver than a cost avoidance one, but with the prices of very high-capacity disks so much lower than those of primary storage-capable drives, it's no wonder that 63% of respondents are using some form of archiver (email, file system, database, etc.), a jump from the 56% who reported using them a year ago. Email archiving, in particular, has been picking up steam, with 38% using it now vs. 28% last spring. And while purchasing plans are being scaled back across all product categories, it looks like cuts in spending for archiving applications will be less drastic than in some other areas (see "Archiving apps: Using now and plan to buy," p. 23).

"We're starting to do some archiving with our email system," said Downing. "Right now we're in the beginning phases of implementing that." The farm property insurance firm uses Symantec Corp.'s Enterprise Vault and expects to expand its scope. "The next step is to go to our file stores and start doing that," he said.

DATA STILL NEEDS TO BE PROTECTED

Data protection is one storage management area that can't be derailed because of belt tightening, and it looks like storage managers may have to resort to robbing Peter to pay Paul to ensure their data is appropriately safeguarded. Peter, in this case, is tape. Spending plans for tape libraries,

drives and media have been dwindling, as indicated in the last few surveys. A few years ago, 47% of respondents planned to increase tape spending while another 34% expected to maintain previous spending levels. Today, only 19% will increase their spending (vs. 32% last fall), while 29% will decrease it. Half the group anticipates spending at 2008 levels, which were already relatively low.

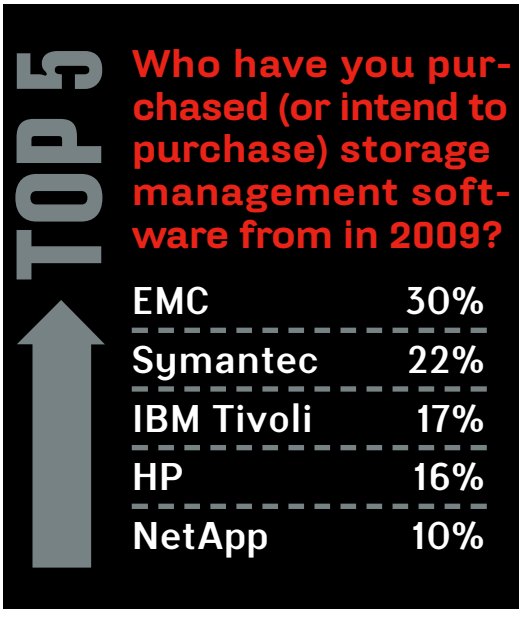
Sixty percent won't buy any tape libraries in 2009, which is the highest number we've ever seen. And those who are making library purchases will opt for smaller units with an average of 101 slots; last fall, the average number of slots was 117 and in spring 2006 it was 159. Clearly, less reliance on tape is anticipated, but higher-capacity tape drives like LTO-4 also contribute to the downsizing of tape library purchases (see "LTO-4 adoption on the rise," p. 23). Still, 80% spin off some or all data to tape, which is down somewhat from 86% one year ago.

But declining tape usage is an old storage story as the focus has shifted to finding more efficient uses of disk in backup environments. And any discussion of efficiency in backup has to start with data deduplication. Dedupe is arguably the hottest technology in storage these days, but that doesn't make it immune from reduced spending. Still, dedupe's numbers are up, with 19% already deploying it vs. 14% in spring 2008. New deployments might slow a bit in 2009, as 10% will decrease their dedupe spending. While not a huge portion of respondents, it's a considerably larger segment than the 1% reported last spring. But 55% will either increase dedupe spending or maintain it at 2008 levels; that's about 12 points lower than both of last year's surveys.

A dedupe project is underway at Hail and Rain, according to the firm's Downing. "We're in the middle of the research for that and hope to have something chosen by the third or fourth quarter of this year," he said.

Horace Mann's Janssen has spoken with a number of vendors about their dedupe products. "We will probably be considering that, but at this point, we aren't doing anything along the lines of deduplication," he said. Meanwhile, he's begun evaluating dedupe product alternatives. "But next year, I would say [it's] 90% certain we'll be doing something with data deduplication," he said.

Virtual tape library (VTL) and continuous data protection (CDP) products offer operational efficiencies that may also translate into bottom-line savings. We've seen some renewed interest in VTLs,



with 29% (vs. 23% last fall) of respondents using them now and 23% planning to deploy them this year. One of the reasons for the minor resurgence in VTLs may be that many of these products now incorporate deduplication technology.

CDP might benefit from more convenient deployment, too. Thirteen percent of respondents currently use CDP vs. 8% in last fall's survey. CDP adoption received a boost when many backup application vendors integrated the capability into their suites over the past year or so. Fifteen percent of the current survey's respondents say they'll add CDP this year.

One area where our survey numbers are trending upward is outsourced backup. The "cloud" has been grabbing headlines with a seemingly endless stream of service announcements and rollouts. We first asked about outsourcing backup in our spring 2007 survey and, until now, respondents have been mostly cool to the offerings. Back in 2007, 89% said they weren't using any outsourced backup services; this time, that number has been whittled to 79%. Since last fall, the number of outsourcing users has risen from 14% to 21%. Email backup services are used the most (12%), with backup services for user files not far behind (10%); the number of users of these services has doubled or nearly doubled across the board (see "Increased interest in cloud backup services," p. 23). Cloud storage services can be very cost-effective, especially considering their utility nature that allows users to pay for what they need when they need it without making any capital expenditures.

DEMANDING TIMES

All in all, 2009 is shaping up as a challenging year for storage professionals who will have to find creative ways to stretch their budgets, add needed capacity and ensure the safety of their data. But there are some positive signs that initiatives begun in the past aren't falling by the wayside in a withering economy.

For the first time, we received more "ayes" than "nays" when we asked if respondents had deployed storage security: 51% said they are, with most methods (encryption with appliances, at the drive or in backup apps) all gaining two or three points. In most of our surveys and interviews, encryption is consistently near the top of storage managers' to-do lists.

We conduct purchasing surveys twice a year and have found a pattern of adjustments that occur between the spring and fall, which is, no doubt, storage managers resetting priorities to the realities of doing storage in the real world. So we look forward to this fall's edition, and hope the results are just a bit more encouraging. In the meantime, some hunkering down is probably in order. "The economy has pretty much tanked a lot of what I call our 'sexy initiatives,'" said United Steelworkers' Wyatt. ⊙

Rich Castagna (rcastagna@storagemagazine.com) is Editorial Director of the Storage Media Group.

Special Report: Online Backup

Have you looked at Online Backup lately?

Visit the SearchDataBackup.com "Online Backup Special Report"
to find out how online backup has evolved:
www.SearchDataBackup.com/online_backup

Sponsored by:

The Web's best storage-specific information resource for enterprise IT professionals

TechTarget Storage Media · SearchStorage.com · SearchStorage.co.UK · STORAGE · Storage Decisions · SearchDataBackup.com · SearchDisasterRecovery.com · SearchSMBStorage.com

A dynamic four-tier storage design

By Herb Ferguson

A comprehensive revamping of the storage environment at a major government agency shows how a tiered storage design can help meet operational expectations without busting the budget.

VENDORS WORK VERY HARD to make the choice of a multitier, enterprise-class storage system an easy one for you. But in the real world, it's not so easy. A multitier, enterprise-class system needs a high level of scalability, and its different tiers need to serve the needs of various applications and databases. It's a substantial long-term investment, and it takes exhaustive planning and research to choose the right one. Perhaps a more fundamental question than which vendor's prod-

ucts to buy is whether to take the integrated, single-vendor approach or to build a system around the components that are most critical to your environment.

In March 2007, InfoPro Corp. was asked to guide a large government agency through such a storage system purchasing decision—upgrading from entry- and workgroup-level storage to an enterprise-class storage subsystem with a much higher capacity and the ability to scale beyond 1 petabyte (1 PB). The task was a tall one, given customer requirements and budgetary constraints.

The agency's existing environment was quite complex. There were numerous networks and approximately 75 servers (90% Sun Microsystems Inc. hardware) running Solaris, Linux and Windows operating systems with a wide range of business applications and databases, from product lifecycle management to document management apps. The environment was separated into loosely coupled sections that corresponded to the customer's business functions, for instance, production, staging and development. Each section had its own set of server and storage constraints; the production section required the highest uptime availability, whereas other sections had less stringent requirements. As for existing storage equipment, there were seven direct-attached SCSI- and Fibre Channel (FC)-based storage arrays from different vendors, each 1 TB to 3 TB in capacity, for a total capacity of 8 TB to 14 TB; there were also two Sun StorageTek L20 tape backup units.

In the year before InfoPro was called in on the project, the agency's user accounts and storage utilization rates were increasing at an alarming rate—utilization went from 3.9 TB in January 2006 to 12 TB in May 2007.

In the year before InfoPro was called in on the project, the agency's user accounts and storage utilization rates were increasing at an alarming rate—utilization went from 3.9 TB in January 2006 to 12 TB in May 2007, threatening to exceed available workgroup storage by the end of the year. There was an urgent need to get the upgrade completed as soon as possible.

There was also considerable pressure to select the right system for the agency's particular needs. Choosing the wrong one can make storage management a living hell and lead to project failures. To make the right choice, an in-depth investigation of the agency's requirements was needed, along with an analysis of available products, features and costs.

PROJECT REQUIREMENTS

In setting the requirements of the new storage system, InfoPro determined that the agency would need at least 110 TB initially, as well as the ability to expand beyond 1 PB to handle future needs. The 110 TB would consist of 10 TB of high-speed storage primarily for the database and 100 TB of medium-speed storage for other functions. In addition to the basic sizing requirements, there was a long list of other "wants":

- Server/storage certification paths (hardware/software wherever available)
- The ability to support immediate storage needs and to expand beyond 1 PB for future project growth
- A flexible and manageable base building block for internal and external storage multitiering
- Three-year, 24/7, on-site training, licensing, installation, maintenance and transitioning support for personnel/software/hardware with a four-hour response
- Hot site, geographic disaster recovery (DR) through replication and failover capabilities
- Increased throughput to 4 Gbps while sustaining existing 1 Gbps and 2 Gbps FC host bus adapters (HBAs)
- Increased storage stability and reliability with virtualization, imaging and snapshot technologies
- Migration/transition/reconfiguration of existing workgroup storage data, including database clusters
- Non-impacting application and database backups
- A 99.9% vendor guarantee of no data loss
- Storage accessible by all project resources
- 10 TB usable initial high-speed tier 1 internal storage expandable to 100 TB
- 100 TB usable tier 2 medium-speed external storage expandable to beyond 1 PB
- 48x 1 Gbps, 2 Gbps and 4 Gbps FC storage-area network (SAN) connections expandable to 180 connections
- Network-attached storage (NAS) cluster failover capability support for NFS, CIFS and iSCSI protocols
- Backup tape library with minimum 4x LTO-3 drives and 100 tapes, expandable to 16 LTO-4 drives and 500 tapes
- Non-disruptive hardware/firmware/software upgrades
- Non-disruptive RAID/logical unit number (LUN) initialization and management (Web based)
- Symantec Corp. Veritas backup/restore software and licenses (library management)
- Hot database backups and restores

TIER DEFINITION

Then it came time to identify storage tiers, requirements and features.

InfoPro determined that four tiers were needed (see "Storage tier configurations," below). The first tier would include 10,000 rpm and 15,000 rpm FC drives, with support for virtualization of other storage tiers, image copies, remote replication and LUN/volume management. The second tier would include medium-speed SAS or SATA II drives and storage, with support for various RAID levels and LUN/volume management. The third tier would include slow-speed, IP-based NAS with support for a failover NAS cluster, as well as NFS, CIFS and iSCSI protocols. The fourth tier would include disk- and tape-based backup with support for disaster recovery and vaulting. All of the tiers would need to be accessible to all of the environment's servers and applications and have built-in redundancy.

At that point it was decided that a loosely coupled approach made more sense than buying an integrated system from a single vendor. With a loosely coupled system, the tiers can be upgraded individually; it's also cheaper, despite opinion to the contrary, and eliminates vendor lock-in. Taking this approach allowed InfoPro to concentrate on the more critical top two tiers, leaving the NAS and backup/restore tiers for a later step in the selection process.

STORAGE TIER CONFIGURATIONS

This was the storage wish list for defining the tiered configuration. It identifies tiers, speed, size, requirements and features. Initial storage sizes represent a minimum footprint. Project sections and tier utilization fluctuate based on customer demands.

Tier	Tier 1	Tier 2	Tier 3	Tier 4
Speed	Fast	Medium	Slow	Dependent
Size	10 TB	100 TB	Uses tier 1 and tier 2	120 TB compressed tape
Requirements, capabilities and features	Virtualization, image copies, replication, FC drives, LUN/volume management	SATA II/SAS drives, LUN/volume management	NAS, failover/cluster, NFS, CISF, iSCSI, LUN/volume management	Tape, backup/restore, DR/vaulting
Production	✔	✔	✔	✔
Staging	✔	✔	✔	✔
Integration	✔	✔	✔	✔
Development			✔	✔
Training			✔	✔
Demo			✔	✔

PRODUCT ANALYSIS

The next step was to survey the range of available midrange/enterprise-class storage products on the market. That survey produced a 30-page document detailing the options. But rather than spending a lot of energy on the wide range of products, InfoPro decided to stay with the storage marketplace's top four at the time: EMC Corp., Hewlett-Packard (HP) Co., IBM Corp. and Sun. But of those four, only storage hardware from EMC and, of course, Sun would be certified by Sun to be compatible with the agency's heavily installed base of Sun server hardware. By choosing a non-Sun-certified storage system, the agency could end up with voided warranties and be caught in the middle of vendor disputes. That certification concern became key as the project progressed.

With a loosely coupled system, the tiers can be upgraded individually; it's also cheaper, despite opinion to the contrary, and eliminates vendor lock-in.

The tier 2 products were considered first, before tier 1. Not all of the vendors' midrange storage products fit the bill for the project's tier 2 requirements. But EMC's Symmetrix DMX series products, HP's StorageWorks Enterprise Virtual Arrays (EVAs) and Sun's StorageTek 6540 arrays had excellent support for tier 2 requirements.

The enterprise-level products InfoPro looked at were "best of breed" in March 2007. Three products—the EMC DMX-3, HP StorageWorks XP12000 disk array and Sun StorageTek 9990V System—could support a 1 PB-plus storage subsystem. (The HP and Sun systems are rebranded Hitachi Data Systems Universal Storage Platform [USP] 1100 units.) The IBM DS series product didn't make the grade as it couldn't meet the 1 PB-plus native requirement, and the EMC DMX-3 supported expansion beyond 1 PB only with its own disk array units. Since the HP and Sun systems were essentially the same product (from Hitachi), it didn't make sense to buy HP and risk installing a non-Sun-certified system. But throwing HP out of the running didn't mean that a decision had been made. Because Sun certifies EMC equipment, InfoPro decided to lower the 1 PB requirement and compare the Sun system with the EMC system.

EMC'S DMX-3 VS. SUN'S STORAGETEK 9990V

EMC and Sun take different approaches to enterprise storage implementation. EMC's DMX-3 is more network-centric, while Sun's 9990V has a built-in controller and firmware. They also differ in I/O subsystem approach: The DMX has a point-to-point bus with a direct matrix architecture, while the 9990V has a crossbar switch with a built-in controller. (See "Comparison of Sun 9990V and EMC DMX-3," p. 33, for more differences.)

The 9990V received points for its flexibility and management advantages. Because it can use lower cost, tier 2 storage and manage it as tier 1 storage via a single Web interface, it offered the agency greater

COMPARISON OF SUN 9990V AND EMC DMX-3

Sun 9990V	EMC DMX-3
Addresses up to 32 PB external storage with virtualization; certified drives not required; I/O channels share 64 GB cache	Limited to 575 TB addressable storage internally only (with 300 GB drives); doesn't support external virtualization; drives must be EMC certified; I/O channels limited to 16 GB shared cache
Supports an internal capacity of up to 332 TB; supports RAID 6 for increased redundancy and reliability	Supports an internal capacity of 575 TB; doesn't support RAID 6
Uses a built-in controller that supports virtualization of internal and external multivendor storage (up to 32 PB); software relies on proprietary controller	Uses network-based Invista app; software relies on switch-based intelligence
Uses logical partitioning via built-in controller for internal and external storage; management software is simple yet lacks some functionality	Uses an in-array tiering approach that supports increased granularity of functions and autoprovisioning; no external tier capability; software is more complex

flexibility for purchasing and configuring disk. The system would also allow creation of internal and external (tier 1 and tier 2) RAIDs and LUNs of varying sizes and levels without configuration BIN files. Finally, because the agency's server hardware was already primarily Sun equipment, there would be a big advantage in using it for storage hardware as there would be no compatibility disputes or certification issues to be concerned about.

THE DECISION

Given these factors, Sun won the agency's business, not only for its tier 1 and tier 2 needs, but for tier 3 and tier 4 as well. Here's what the agency decided on (see "Final tiered storage system design," p. 34):

- Tier 1: Sun StorageTek 9990V System
- Tier 2: Sun StorageTek 6540 array
- Tier 3: Sun StorageTek 5320 NAS Appliance (model now discontinued)
- Tier 4: Sun StorageTek SL500 Modular Library System

The central component of the storage subsystem is the Sun 9990V, which performs all tier 1 functions via the Storage Navigator Web interface; virtualizes tier 2 storage for servers; and handles the NAS cluster, tier 4 backup storage, as well as LUN management for all storage tiers. The system—which currently has 48 300 GB, 10,000 rpm FC drives amounting to 14 TB of tier 1 storage—has performed exceptionally out of the box, with no tuning. The only anomaly occurred during a firmware upgrade to the 9990V, pointing to a configuration problem with the 6540; that problem was resolved without system interruption.

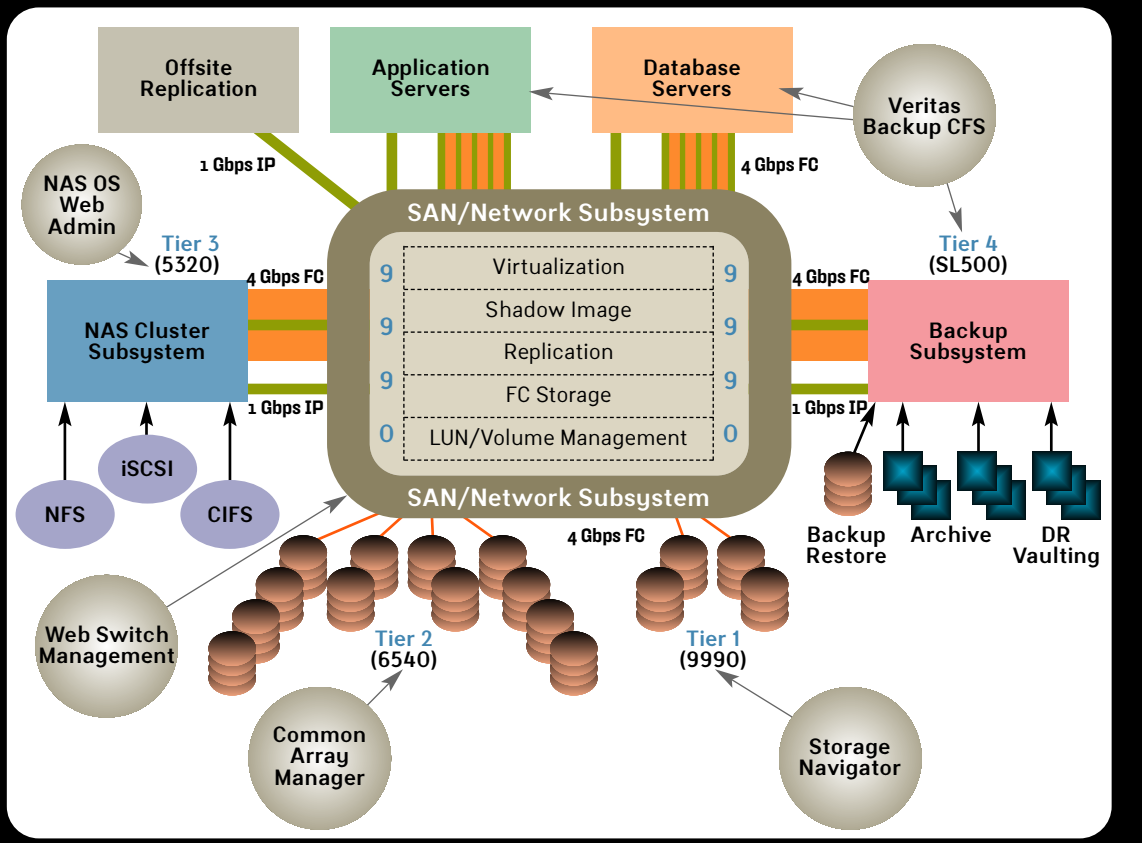
In its role as the tier 2 system, the Sun StorageTek 6540 provides more than 100 TB of 3 Gbps SATA II drives virtualized by the 9990V as external storage. There are eight 4 Gbps FC connections between the 6540 and the 9990V, balanced based on I/O loads between the two 6540 controllers. Array and LUN slicing and dicing is handled interactively without hindering performance on the other tiers and servers. The system has performed admirably since it was installed.

At tier 3, two Sun StorageTek 5320 NAS Gateway units are linked to the 9990V through eight 2 Gbps FC connections and to project servers via multiple 1 Gbps network connections managed by Ethernet switches. The system supports all three protocols the agency uses: NFS, iSCSI and CIFS. LUNs presented by the 9990V are sliced into local volumes that reside on a proprietary Sun StorageTek file system. The system has passed the tests InfoPro has put it through: failover testing and interactive expansion of the NAS cluster volumes shared via NFS to networked servers.

Finally, at tier 4, the Sun StorageTek SL500 tape library has four LTO-3 tape drives, 150 tapes, a T2000 backup server and Symantec Veritas software. This tier handles disk and tape backup, and vaulting and restoration

FINAL TIERED STORAGE SYSTEM DESIGN

This diagram shows the final enterprise storage system design. Physical separation of tiers provides dynamics and flexibility.



of application and database data. Each LTO-3 drive can read and write at a sustained 80 MBps individually, with an aggregate throughput of 320 MBps. The system is modular and can be expanded in increments of 150 tapes with an additional four drives. It's currently backing up 80 TB of data per month.

PUTTING THE PIECES TOGETHER

Once the components were chosen, networking details needed to be worked out, and the guiding principle was to provide connections in the right places so that client communication wouldn't be impacted by storage system operations. In the current configuration, the storage subsystems' four tiers are connected directly to one another, separate from the SAN and client FC connectivity. The tier 4 T2000 backup server has direct network connections to the tier 3 NAS Gateway cluster servers to enable backups to be done without slowing down the rest of the network. And storage, application and database expansion can be done without affecting client and storage subsystems.

It's worth noting that software and firmware upgrades can be performed with little to no impact on other components.

The separate storage and client networks have been implemented with low-cost switch and virtual local-area network (VLAN) technologies, eliminating network contention, isolating traffic and increasing security. The backup server has its own direct FC and Ethernet connectivity, so backup and restores can happen around the clock. Administration of all network devices can be done via either the backup server or remotely from the administrator's desktop. And all FC connections are auto-sensed to 1 Gbps, 2 Gbps or 4 Gbps, except for the LTO-3 tape drives, which are set at 2 Gbps.

It's worth noting that software and firmware upgrades can be performed with little to no impact on other components. Firmware upgrades to the 9990V, for example, are done interactively by switching I/O and/or LUNs and connections from front to back; upgrades to the 5320 NAS Appliance are handled in a similar way. Firmware upgrades to the 6540 are also done in place, though I/O performance degrades during the upgrade. Upgrades to the T2000 backup server are done during non-backup cycles to avoid impacting the backup schedule.

THE OUTCOME

With the completed storage system upgrade in production for almost two years, the agency is able to assess how closely the initial plans mapped to actual use. There are now approximately 15,000 user accounts in the system, with daily user access peaking at around 10,000. Storage utilization is at approximately 40 TB. The agency is considering

implementing thin provisioning techniques to cut down on future storage needs; even so, it's expected that the system will use about 500 TB in the next two to three years and more than 1 PB in five to seven years.

A loosely coupled system like the one the agency implemented brings with it a lot of flexibility and room for growth. But there are tradeoffs: A loosely coupled, multitiered system is inherently more complex than an integrated one from a single vendor, uses multiple management interfaces, and carries compatibility and certification path concerns. In the case of the agency that InfoPro worked with, the upfront research and engineering work made it clear that a loosely coupled system was the right choice. But each IT environment is different; a proper decision process should include not only a comparison of the available systems but discussions with vendor sales and technical reps to make sure that you fully understand their technology offerings. And it's important to dig beyond the sales pitch. Do your own research and, when needed, pull other trusted technical pros into the discussion. Taking the investigation and planning steps of the project very seriously can mean the difference between a system that's universally applauded and meets expectations—as was the case for the multitier system detailed here—and one that's quickly outdated or inappropriate for the project it was bought for.

Herb Ferguson is a senior systems scientist at InfoPro Corp. in Huntsville, Ala., with 25 years of IT architect, engineering, R&D, programming and network experience.

A wealth of deduplication options

Hardware-based products propelled deduplication into the mainstream, but now that most backup apps include dedupe, you'll have to carefully evaluate the options.

DATA GROWTH GRABS most of today's IT headlines and many IT organizations believe data protection is one of the key contributors to the staggering data capacities that need to be managed. Why? Lots of copies are made by data protection processes—at least once per day, but sometimes multiple times daily—and kept locally for operational recovery. Copies of copies are also sent offsite for disaster recovery (DR) purposes. Most backup and replication solutions perform these processes inefficiently, making multiple copies of the same file despite only a small amount of the data within the file having been changed. Maintaining daily, weekly, monthly and yearly backup copies means that dozens of copies of the same data may be stored, and often for extended periods of time. It's this propagation of data that makes data deduplication a compelling technology for secondary storage environments.

Deduplication may be mixed and matched, taking advantage of features of both software and hardware products.

While the deduplication spotlight has been focused to date on hardware products that optimize storage capacity, the addition of dedupe capabilities in several backup apps could shift the focus in 2009.

As more organizations implement disk in the backup process to overcome the performance and reliability shortcomings of tape-based protection, data deduplication has emerged as a force to improve the economic feasibility of retaining data longer on disk (possibly eliminating tape) or increasing the number of workloads using disk as an interim stop on the way to longer-term retention on tape. Deduplication technology conserves storage space by writing only unique (new or changed) data to disk and linking it via pointers to the previously stored unchanged data.

DEDUPE APPROACHES COMPARED

Hardware vendors spearheaded dedupe adoption with powerful, purpose-built deduplication appliances that process backup data before

or after it's written to disk. Benign to the existing backup environment, this hardware-based approach made deploying dedupe relatively easy. Research from the Enterprise Strategy Group has found that the ability to integrate with existing backup processes and overall ease of use are more important adoption factors to organizations than specific technical considerations, such as a deduplication ratio or the granularity of deduplication.

Seamless integration with existing data protection practices, as well as IT's historic resistance to change when it comes to backup software, meant that backup solution providers that could offer deduplication had a more difficult time getting mindshare in the data center. When EMC Corp.'s Avamar came to market touting a better, more efficient way to back up data, the company faced an obstacle that was hard to overcome: reluctance to walk away from existing backup applications. IT organizations could clearly understand the benefits, but weren't motivated to initiate a technology change that would have a ripple effect on the operational aspects—people and process—of the data protection environment. EMC Avamar has therefore had to take a more circuitous route to the data center, providing a bandwidth- and storage-optimized backup solution for remote and branch offices, as well as an efficient data protection alternative for server virtualization environments.

However, the integration of acquired deduplication products by EMC (Avamar) and Symantec Corp. (PureDisk) with NetWorker and Veritas NetBackup, respectively, as well as recent introductions of native dedupe by CA, CommVault and IBM Corp. have a lot of IT organizations wondering which is the best implementation of deduplication—hardware or software? Bottom line: It's not a one-size-fits-all scenario.

FACTORS TO CONSIDER

Cost, performance, scalability and the deduplication domain are just a few of the considerations when evaluating deduplication in the backup process to determine whether a backup application's built-in dedupe capability or a feature built into a backup storage system will best serve your environment.

Cost. Presumably, an investment made in technology that can reduce storage capacity requirements by a factor of 20 will be easily justified. Is there an added fee to enable the feature whether it's a backup app capability or an "add-on" feature in a hardware device? Is an upgrade to a higher version or model required? Even if deduplication is standard in the product (hardware or software), what other cost implications are there for implementing it (e.g., will it require additional network, server or storage resources)?

Performance. Deduplication comes in all shapes and sizes as backup workloads have different requirements. Deduplication may be mixed and matched, taking advantage of features of both software and hardware products. Source-side dedupe in backup software may make the

most sense for remote systems because it delivers greater network efficiency, while target-side approaches may make more sense for workloads with the most stringent backup windows.

Scalability. While deduplication should mitigate the need to expand storage capacity, the impact of growth on the dedupe environment should be thought through. You need to determine how easy or difficult it is to expand the deployed product, and if expansion will introduce silos of storage (and thereby limit deduplication) and increase management. And does scaling require a forklift upgrade or can it be achieved more seamlessly?

Deduplication domain. You also need to consider the scope of the deduplication effort. Will your dedupe effort be limited to the confines of a single container—whether it's logical or physical—or are your goals broader?

Such a wealth of deduplication options provides ample choices, but it can also lead to some confusion. Vendors have the opportunity to educate users about deduplication technology in general, and specifically how their own solutions approach the task. And you need to understand your backup environment and requirements before short-listing solutions. Vet the vendors and their products, check their references and, most importantly, test the products using your own data over several backup cycles. ⊙

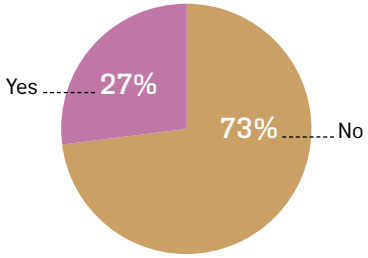
Lauren Whitehouse is an analyst focusing on backup and recovery software and replication solutions at Enterprise Strategy Group, Milford, Mass.

CDP picks up a little steam

AS SLICK AS IT SOUNDS, continuous data protection (CDP)—the ability to capture and back up new and changed data immediately—hasn't made a huge impact in most storage shops yet. On our most recent survey, only 27% of respondents currently use a CDP application, a modest increase of approximately four points vs. last year's survey. The rap against CDP has been that it's yet another app that adds complexity and administration to backup environments. Adoption was expected to increase as more backup application vendors integrated CDP into their products, but that doesn't seem to account for the slight rise in use, as only 23% of our CDP users tap their backup apps for that capability. CDP has been around for a while, but most of those users are relative newbies, with 59% saying they've been using it for less than a year. Don't expect a sudden surge in CDP deployments, as just 30% of non-users are considering deployments in the next two years.

—Rich Castagna

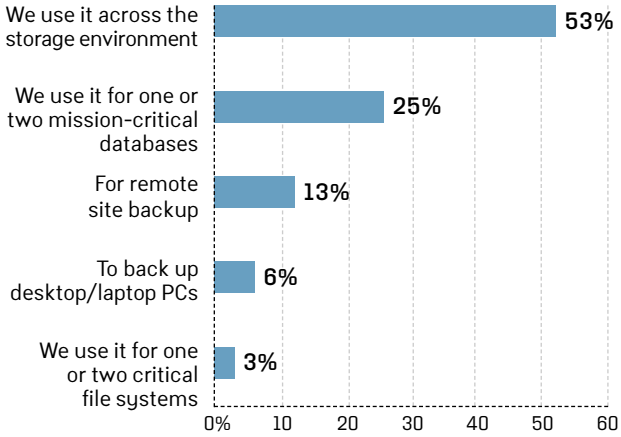
Do you use a continuous data protection (CDP) product?



If you're not planning to implement CDP, why not?

38%	We're meeting our recovery point objectives without it
23%	We don't have the resources for another data protection application
13%	Our current snapshot application is sufficient
8%	The cost of CDP outweighs the benefits
8%	The technology isn't mature enough

How do you use CDP?



41%

Respondents who have been using CDP for less than six months

> "The technology is not as mature as some solutions—we are taking a wait-and-see approach."
>
> *—Survey respondent*

Check out the following resources from our sponsors:

Asigra Inc., page 12

Software Overview

Agentless Architecture White Paper

Computer Design & Integration LLC, page 17

e-Guide: Backup Considerations -- Data Deduplication and Virtualization

An Infrastructure Services Firm

Dell Inc., page 6

IT Consolidation with VMware and Dell EqualLogic iSCSI SANs

Podcast -- Storage Virtualization Benefits in a Microsoft Windows Server 2008 Hyper-V Environment

NetApp, page 8

Use 50% Less Storage With NetApp. Guaranteed.

See How Much Space You Can Save With Our Calculator.

Symantec Corp., page 3

Symantec Is Storage Software

Stop Buying Storage Webcast Series -- On-Demand