



You can't stop the massive growth in data, but you can do a better job of managing it

CONTROLLED EXPLOSION



Storing and managing a large—and ever-increasing—amount of data presents a formidable business challenge. “Data volume is growing faster than IT budgets and faster than technology is evolving to solve the problem,” says Sanjeet Singh, a product marketing manager at Dell. Midsized companies are particularly challenged, because they typically lack the resources (both money and manpower) to handle the deluge.



COMPUTERWORLD
Custom Solutions Group

InfoWorld
Custom Solutions Group



“IT departments are beginning to shift their focus from adding more storage to improving the efficiency of the storage systems already in place.”

These companies often deal with the problem by adding storage capacity on an ad-hoc basis, installing systems from different vendors with different management software. While this may fill a short-term need, in the long term it creates an expensive IT management headache and only adds to data center inefficiencies. And it raises the question: is adding another server or a bigger disk the best way to deal with this explosion of data, or is there a better way?

Three Pillars of Data Management Efficiency

With infrastructure budgets under constant pressure, IT departments are beginning to shift their focus from adding more storage to improving the efficiency of the storage systems already in place. Today, the name of the game is improving productivity and efficiency while eliminating waste.

Indeed, many analysts predict strong growth for technologies that help mid-sized companies increase efficiency and get greater business value from their data storage investments. Some technology providers have zeroed in on three pillars of efficiency that bring almost immediate relief from the cost, complexity and administration of data management:

I. CONSOLIDATION

Server-attached storage has limited space for disk drives, which means when IT departments need more storage, they're forced to purchase another server. This yields isolated storage “islands” that are difficult and costly to manage because

Managing mountains of data: The big picture

TO CRAFT A COMPLETE STORAGE STRATEGY, YOU NEED TO DRILL INTO ALL LEVELS OF THE INFRASTRUCTURE -- BUT DON'T FORGET THE BIG PICTURE

BY MATT PRIGGE
INFOWORLD

When I think of enterprise storage, I reflexively think of a big stack of highly redundant, high-performance disk.

That disk generally gets a lot of attention because it's flashy, fast, and expensive, and it plays a central role in your most critical workloads. But the enterprise data explosion goes beyond your core primary storage assets; it also touches everything that protects, monitors, and secures that data. As you weigh storage investments, remember to consider the secondary effects of throwing all that snazzy disk onto your network. Primary storage. Storage in the form of spinning disk has been around for more than 50 years now. All businesses, from large enterprises to SMBs, need a coherent primary storage strategy to deal with the continuous and rapid expansion of data. The process begins with a thorough assessment of business needs, current data throughput, and the state of the existing storage infrastructure. A host of technologies, including storage virtualization,

the systems must be updated and maintained individually. “Companies throw servers into their environment not because they need more horsepower but because they need more storage,” says Steve Arrington, product marketing strategist at Dell. “Then they have to hire more people to manage all this stuff.”

Consolidating resources in a virtualized storage area network, or SAN, allows companies to break free from the constraints of server-attached storage. It separates data management from servers, which allows IT to speed up operations like backup, retrieval and archiving because the tasks don’t compete for system resources. A SAN can be used to mirror information at remote locations, ensuring quick recovery in the event of an outage. And the prices of SAN-based technologies such as iSCSI have dropped, making them more viable options for financially constrained midsize companies.

Storage resources can also be consolidated in a network-attached storage (NAS) environment. IT departments should consider cost, distance between offices, and number of users when choosing between a SAN or NAS.

When storage resources are consolidated into a single pool, IT managers can store more data on fewer drives, which use less power and require less physical space. The benefits over direct-attached storage are clear:

- + lower cost per megabyte of stored data
- + no single point of failure
- + centralized data means less administration
- + greater disk utilization
- + network-wide storage management
- + scalability; can add capacity without taking system offline
- + supports many operating systems

II. DEDUPLICATION

Data duplication (unnecessarily storing multiple copies of the same data) is a big drain on storage resources. It wastes capacity, creates longer backup windows, slows down retrieval and generates more management overhead. According to one data management consultant, duplication rates of up to 30 percent are not uncommon among companies with no formal master data management plan. (And what midsize company has one of those?)

Deduplication—identifying and removing redundant data—is becoming a more common approach to backup, retrieval and archive (BURA), and the market is expected to reach the multibillions by the end of the decade. According to a leading research firm, using virtualization to eliminate duplicate data was one of the top strategic technologies of 2009.

There are many deduplication options. Single-instance storage (SIS) is a deduplication technique that identifies, transfers and stores a single instance of a file. More comprehensive deduplication approaches locate more redundancies and therefore result in greater data reduction. For example, advanced data compression further reduces capacity requirements by squeezing data down to a smaller footprint. The best technologies use multiple techniques to achieve maximum data reduction, which ultimately results in a significant decrease in hard and soft storage costs.

III. BACKUP, RETRIEVAL AND ARCHIVE

According to a November 2009 report from the Enterprise Strategy Group, midsize organizations’ data protection strategies are in transition: only 6 percent of companies back up to disk exclusively, 46 percent to tape exclusively and 48 percent to a combination of both. Of those backing up to tape only, 45

new SSD products, enhanced caching schemes, and disk deduplication provide new tools to craft a rational storage architecture that provides performance where it’s needed and scalability as a matter of course.

Backup infrastructure. For as long as we’ve had primary storage assets, we’ve needed to protect them from loss. Backup infrastructure is just as critical to your organization’s survival as primary storage. Natural disaster, fire, viruses, data corruption, user error, and administrator error are only a few of a nearly limitless number of bad things that can happen to data. Recently, backup options have expanded dramatically; what used to be a commodity space filled with varying types of tape drives and and write-once media has opened up to all manner of disk-to-disk and offsite backup solutions. Automated backup deduplication, compression, and site-to-site replication are rapidly evolving as the technology matures.

Data deduplication.

Deduplication has become almost a prerequisite for a backup solution. A deduplication system identifies and eliminates

percent plan to move to disk backup over the next 12-24 months. The transition is being driven by a need to improve backup and recovery performance and reliability, manage costs, keep pace with capacity requirements and deal with unwieldy management of tape-based solutions.

Most midsized companies lack the staff to manage their tape infrastructures. Disk-based BURA solutions can be viable options, but only if companies address disk capacity costs, scalability, management overhead, and efficient offsite data protection. iSCSI SAN arrays and software like ArchiveIQ eliminate the high cost typically associated with disk-based BURA solutions. While disk capacity may cost ten times as much per gigabyte as tape, software solutions like ArchiveIQ are designed to keep more backup data on disk than tape, enabling a lower cost per gigabyte.

The software simplifies and accelerates file recovery via file name indexes that support searches across several months of deduplicated data. That means administrators can dig into each recovery point like a normal file share. And since the software does not chop data into pieces, recovery can be carried out at full disk speeds.

Conclusion

It's no secret that many companies are struggling to minimize waste while increasing productivity and efficiency. The three pillars of efficiency described above help midsized companies cope with a growing volume of data, yielding tangible results. But just as important, these technologies improve an organization's ability to generate revenue and boost profitability by streamlining business processes.

Companies will continue to add storage capacity, of course, but there is a definite shift underway to investing in storage efficiency technologies. According to IDC, "in revenue terms, 2010 will essentially be flat, but the fundamentals by which storage is acquired, consumed, provisioned and managed will change dramatically."

IT spending among small and midsized companies is expected to grow 3.7 percent this year to nearly \$510 billion, predicts IDC. Companies will spend on data storage technologies—consolidation, deduplication, unified NAS/SAN and disk-based BURA—that improve efficiencies. These are real alternatives to adding capacity; alternatives that help midsized businesses streamline their data centers and cut costs.

redundant files, reducing the amount of space necessary to store backups from 10:1 to 50:1 and beyond, depending on the level of redundancy in the data. Almost all major backup software either includes or has announced native deduplication capabilities. The real promise of deduplication won't be realized completely until it becomes commonplace in primary storage, not just in the backup tier. If implemented correctly, deduplication can also increase disk cache performance dramatically, particularly in virtualized server infrastructures with a high degree of duplication.

Log analysis and reporting. As primary storage and backup environments continue to grow in size and complexity, collecting and reporting on how well those environments are working becomes imperative. No CFO I know likes to get an unexpected request for a pile of money to upgrade a SAN that's reaching capacity. Monitoring and reporting back to your organization about the state of your storage assets is critical. Whether it's to give accurate budget guidance to the corner



Storage Management

CENTRALIZING STORAGE

Centralizing storage sounds like a job for Captain Kirk, not for a CEO steering a much smaller enterprise. Yet putting all your company's data into one location is an easy process that can positively affect your bottom line.

Decentralization is an inevitable by-product of growth. First there's just the CEO's desktop; then the first few sales representatives are added, followed by a marketing professional and an accountant. Each person gets a computer, and each starts building a silo of critical company data. This structure results in the occasional annoyance of having to email files to one another — which can create a version control problem. And when someone's sick, access to certain information is restricted until the person who's sick gives up the password or comes back to work.

Security and backups are also a growing concern, but the team takes some comfort in the fact that if one computer fails, the rest of the data is safe, or if a hacker breaks in to one PC, he can't see the information in the others. Of course this security comes with a wink and a nod: when that lone unit fails, the data will likely be gone forever.

The decision to centralize often comes when the frustration boils over: a critical file is needed, but can't be accessed; an employee quits and their customer list leaves with them; or a hard drive dies and the data is mourned rather

office, to maintain regulatory compliance, or to communicate your backup timeframes to your application stakeholders, it's vital to know what's going on under the hood.

Enterprise data

protection. Here we have a broad set of practices and technologies designed to protect private and proprietary data both "at rest" and "in motion" throughout the enterprise and its business partners.

Full-disk encryption for laptops and PCs, database encryption, encryption of backed up and archived data, secure file sharing solutions, database and file server monitoring, multifactor authentication, and data leak prevention solutions are a number of the point solutions that companies are using to enforce policies (and meet compliance requirements) for handling sensitive data, including financial, health, customer, and private information.

Cloud solutions. Cloud infrastructure services (combined with virtualization) are poised to serve as the foundation of many companies' disaster recovery plans. In

than resurrected. Whatever the situation, it's a turning point that spurs the purchase of a server or a call to the local IT vendor to start pressing the existing server's most powerful buttons.

Benefits of Storage Centralization

Cost Savings

Growing companies approach storage the same way their printers do toner: they use fantastic amounts and always ask for more. As the emails and PowerPoint® files pile up, those seemingly huge 50 GB hard drives start to look tiny. Centralization won't limit the amount of information your company produces, but it can store what you make more efficiently and at a lower cost.

Let's say you email a 10 Mb PowerPoint file to five employees. Two respond with edits. With no centralization, the total memory cost of that process would be 80 Mb (your original version, five copies sent to teammates, plus two new ones). With centralization (and a process called Single Instance Storage) the



Most businesses decide that they can achieve a higher level of overall security with one set of high, strong walls, rather than five sets of semi-high, sometimes strong walls around five baskets.

server's software recognizes that six of the files are identical, and will delete all but one. The total cost: 30 Mb. Repeat this process with thousands of files over several years, and the costs savings become very clear.

When you do have to purchase more memory, having a central location makes it easier to install in your network. Instead of adding external hard drives to each individual PC, you simply add a box that holds several hard drives to your server. Some of these boxes hold up to 15 drives, and you only need to purchase them one at a time as your needs grow.

Security

When it comes to security benefits, storage centralization is a double-edged sword. It's true that putting all your eggs in one basket (or all your data in one box) can increase security risks. But consider this: are you ready, willing and able to protect each "basket" holding your company's data? If your company

addition, many of the first practical cloud-based applications have been built to store, manage, and process massive data sets, leveraging large clusters of commodity hardware and using programming frameworks such as MapReduce for reliable and scalable distributed computing.

Today most enterprises deal with rampant data growth simply by throwing more terabytes of capacity at the problem. But managing and securing huge volumes of data becomes more and more painful over time, and the cost of buying and maintaining new hardware without increased efficiency can't be sustained forever. Without a big-picture view of the challenge and an interlocking strategy for each of the six major segments, you'll never succeed in effectively managing the explosive growth of data.

This article, "Managing mountains of data: The big picture," was originally published at InfoWorld.com. Follow the latest developments in storage at InfoWorld.com.

has five PC's, that means installing and maintaining five separate firewalls, anti-virus programs, intrusion detection tools, and five sets of backup files. Centralization means installing and maintaining just one set of these programs. Faced with this choice, most businesses decide that they can achieve a higher level of overall security with one set of high, strong walls, rather than five sets of semi-high, sometimes strong walls around five baskets.

Internal security levels can also be established and managed easier. From the server, certain kinds of data or specific files can be blocked from specific employees. Access logs can be read to see who has been reading or changing what files, should the need for a review come up. Finally, if employees leave the company, centralization means their data stays with the team.

Speed

Centralization can also build your top line by helping employee's access data faster. Many servers use something called RAID (redundant array of independent disks) to speed their systems and provide a backup in case of hard drive failure. This is a combination of hardware and/or software that saves two or more copies of all the company's files across multiple hard drives (drive A, B, C, D, E and so on) in a single server. Let's say that two employees are accessing the server at the same time. Employee number one is trying to open a Word document, while employee number two is trying to save an Excel file.

A hard drive can only read (open) or write (save) files one at a time; they can't perform both actions simultaneously. But with RAID, the server can open the file for employee number one from drive A, while writing the file for employee number two onto drive B. The result would be faster file access for everyone.

Speed can also be boosted with network attached storage (NAS), a fancy term for a server that's designed to do a certain task, like file sharing. The hardware and software of this server are tweaked to provide maximum speed for calling up files or sharing them with teammates. These devices keep pace with employees who run at top speed, and reduce morale-busting frustrations with an otherwise sluggish IT infrastructure.

Disaster Recovery

Hard drives are the hardest working parts of any computer or server. Constantly in use and consisting of many moving parts, they're prone to breaking down over time. With the perspective that a failure is a matter of when, and not if, what can storage centralization do to make the problems easier to solve?

First, backing up one set of files from the server is a lot easier than doing it on for three, five, 10 or as many computers as the company is running. RAID can also come to the rescue. With multiple copies of the company's files saved over several hard drives, the failure of one of those drives means zero loss of data. Second, consolidating data on a server means you have more robust options for backing up and creating an offsite copy of your data, including removable disk, tape and even other hard drives that can be stored on a different device in another location.

There are a few drawbacks to storage centralization. The first is the upfront cost of the server, then ongoing maintenance. But compared to the cost of managing the same data in many different locations, it can be worth the investment. You're inviting major problems if you put all your data in one basket then fail to protect, properly manage access to and back up that data.

Power savings on a platter: How to save on storage

BY ROBERT L. MITCHELL
COMPUTERWORLD

After servers, the biggest area of focus for energy efficiency in your data center should be storage. Storage is increasing in power density. As you go from gigabytes to terabytes it gets hotter. The increase in power density for data center storage didn't spike as servers did a few years ago, but it's still significant. A fully loaded storage frame from EMC can consume as much as 700 watts per square foot, or about 2.1 KW per rack, according to EMC.

While storage density per drive has increased, the actual number of drive spindles per bay and power consumption per spindle hasn't gone up much. "We've had a gradual increase in the number of spindles that can fit into a storage frame but it's nowhere near the change you saw for servers," says Dick Sullivan, director of enterprise marketing at EMC.

Storage spending focused on data consolidation, fast ROI

THEINFOPRO STUDY FINDS 45% OF FORTUNE 1000 FIRMS PLAN TO BOOST STORAGE SPENDING THIS YEAR

BY LUCAS MEARIAN, COMPUTER WORLD

A study released today by research firm TheInfoPro showed that those in charge of IT dollars at corporations around the U.S. and Europe are focused on buying technology that can optimize systems they are already using.

For example, the study concluded that companies are spending significant dollars on thin provisioning, which allocates storage capacity on an as-needed basis for apps, and data de-duplication technology, which eliminates duplicate files or blocks of data either at the primary storage layer or during backups.

On the other hand, TheInfoPro's survey found that few companies are deploying new business applications, which would, in-turn, drive new data storage technology purchases.

"Major spending increases won't resume until new business application installs once again create massive demands on storage needs," said Rob Stevenson, managing director of storage research at TheInfoPro. "In the interim, storage shops will focus on productivity improvement and hardware inventory adjustments to prepare for virtualization and cloud support needs."

TheInfoPro's study, which is based on interviews with storage decision makers in North America and Europe, revealed that while 45% of Fortune 1000 respondents plan to increase storage spending in the coming months, 29% still expect major budget cuts. In contrast, 41% of midsized enterprises plan to increase storage spending this year, while 25% expect further budget reductions.

Among the top technologies in TheInfoPro's so-called "Heat Indexes", which identify trends in early technology adoption for both Fortune 1000 and mid-sized enterprises, were: E-mail archiving, information lifecycle management and storage resource management systems, said Ken Male, TheInfoPro's executive vice chairman.

The average Fortune 1000 company has about 1.2 petabytes of disk-based data storage capacity today, so anything that will help manage the data tsunami or trim excess data is hot, Male added.

"Those technologies are top of mind because budgets are so tight right now," Male said. "It's a little bit of back to the future. These were technologies folks were interested in a few years ago and are becoming popular again because they help to forestall additional IT purchases. We're seeing smaller spends on things that will bring a fast ROI."

TheInfoPro's study found that IT managers are interested in solid state drives (SSD), but concluded that sales won't take off this year. According to Male, only about 50% of Fortune 1000 companies currently have SSD technology earmarked in their future technology spending plans. "There's interest, but it's not as top of mind as the market would like to say it is," he said.

1 | Refresh old storage.

Dramatic increases in drive storage densities mean IT can consolidate into fewer storage frames and save a considerable amount of electricity. A user who moves 1 terabyte of Fibre Channel storage from 73 gigabyte drives to a single 1 terabyte spindle will cut power use by about 94%, Sullivan says.

The disk drives themselves are also somewhat more efficient than they were just a few years ago. Automated power management features that can spin down disks, virtual LUNs, and variable speed cooling fans are some of the technologies that can help IT manage storage for greater efficiency, Sullivan says.

2 | Add serial ATA, Flash drives to the mix.

Fibre Channel disk drives are the most power hungry. Spindle for spindle, a SATA drive uses 28% less power than a similar FiberChannel unit, but they're slower and offer lower performance. Combining SATA with flash drives can compensate for that while saving power. Sullivan

TheInfoPro's survey did find a trend toward the adoption of automated data tiering and automated data provisioning, as well as backup virtualization management, or being able to manage the backup of an entire data center's virtual server environment from a single console.

To take advantage of that trend, EMC rolled out its fully automated storage tiering (FAST) technology across its Symmetrix, Clariion and Celerra line of storage arrays, while IBM introduced additions to its XIV grid storage system and its automated tiering capability.

"Hitachi Data Systems is working on automated tiering technology as well," Male said.

Another hot area for upcoming spending is 8Gbit/sec Fibre Channel and virtual server data management, which leaves Fibre Channel over Ethernet (FCoE) deployment lagging, TheInfoPro found.

Male said data transfer protocols that converge LAN and storage area network (SAN) traffic, such as the FCoE and iSCSI protocols, are gaining popularity, but that significant market adoption won't take place until late 2011. "We are seeing more of IP-SAN [such as iSCSI] and NAS [network-attached storage] getting more mind share as storage for virtual servers," he said.

Male said for the next 18 months or so, IT decision makers are choosing to continue using 4Gbit/sec Fibre Channel, or will be moving to 8Gbit/sec Fibre Channel.

"People are very comfortable right now with 4Gbit Fibre Channel," he said. "They'll go to 8Gbit when they begin doing refreshes."

Lucas Mearian covers storage, disaster recovery and business continuity, financial services infrastructure and health care IT for Computerworld. Follow Lucas on Twitter at @lucasmearian, send e-mail to lmearian@computerworld.com or subscribe to Lucas's RSS feed .

Read more about storage in Computerworld's Storage Knowledge Center.

Consolidated Storage, Improved Efficiency

As a business grows, so does its storage requirements. With an increasing number of files, e-mails, databases and application data, it needs more storage capacity to keep information.

To accommodate this surge, a company usually acquires storage units piecemeal, sometimes with each coming from a different vendor, using a unique management system. Operating these isolated islands of storage can be quite time-consuming, expensive and inefficient.

A popular and efficient answer to this storage problem is to consolidate their resources in a virtualized IT storage area network (SAN).

No longer just for enterprise-level data centers, SANs provide a wide range of benefits for businesses of all sizes. From streamlining infrastructure, expanding capacity and saving space to increasing performance and reducing overhead, SANs are efficient solutions to storage requirements.

SAN allows enterprises to move beyond the constraints of server-attached storage. Traditional storage architectures provide only limited space for disk drives, forcing IT departments to buy more servers as their storage needs grow. Networked storage, on the other hand, separates data management from servers, which enables companies to consolidate equipment and reduce maintenance, facilities, and software licensing costs.

expects storage frames to gradually migrate to a combination of Fibre Channel, SATA and flash drives. Eventually, he predicts, Fibre Channel will go away entirely.

Flash cuts energy costs while improving overall performance. Byte for byte, flash uses 38% less energy than Fibre Channel drives while increasing the input/output operations per second (IOPs) to power ratio by 98%.

So what does that mean? Consider a 65TB storage frame consisting of 528 146 GByte 15,000 rpm Fibre Channel disk drives. Replacing those with a storage frame containing eight 200GB flash disks, 104 400 Gbyte 10,000 rpm Fibre Channel drives and 32 1 terabyte SATA drives would reduce power by 32% and storage costs by 17%, Sullivan says.

There's just one problem: Up to now allocating data between those three types of storage has been a manual process. EMC, like other storage vendors, has been working on developing automated "storage tiering" tools. Sullivan says storage administrators can expect to see new products that automate the

Segregating storage and servers also allows IT departments to speed up operations, such as data backup, because these tasks do not compete for the same system resources. In addition, a SAN can leverage an array of high-speed network protocols, including Fibre Channel. This improved performance is especially important today because heightened network use minimizes backup windows.

In addition, a SAN simplifies the management of business continuity processes, strengthening a company's resiliency in the event of disasters. Companies can use a SAN to mirror critical information at distributed locations, which helps ensure quick recovery without any downtime.

The first step in exploring a virtualized storage solution for your company is to consult with an experienced IT storage partner. That's why many IT managers turn to Dell. With a rich history of implementing SANs for a number of companies of all sizes, Dell is a premier end-to-end virtualized storage consolidation provider.

A Dell SAN solution can help you to:

- + Consolidate storage in a central location
- + Expand storage capacity and performance
- + Scale up your storage solutions as you grow
- + Use fewer servers and reduce facility space
- + Simplify IT operations
- + Work with multiple OS environments
- + Reduce expenditures
- + Increase data availability and protect against downtime
- + Comprehensive Solutions for All Storage Needs

As your IT partner, Dell will help you choose the SAN solution that fulfills your storage requirements now and in the future.

If you need an affordable, high-performance modular disk storage solution, you might want to look at the Dell PowerVault™ MD3000i. Or, if you need to keep up with the demands of your increasing data and application growth, the Dell AX4-5I might be the answer. For businesses and remote offices that need virtual server deployments, the Dell EqualLogic™ PS4000 Series provides an affordable solution.

Every company has different storage needs. That's why we invite you to call us at Dell, so we can discuss your particular needs. Together we'll analyze your situation and create a customized solution that can bring you the efficiency, productivity and cost savings that you desire.

movement of data some time in 2010.

Ultimately, Fibre Channel is on its way out, says Sullivan. "I don't think it's too far in the future when we'll have storage systems with no Fibre Channel, SSDs for performance and 1 terabyte, 2 terabyte and 5 terabyte SATA drives," he says.

3 | Consolidate your data.

Finally, a sure-fire way to gain efficiencies is to invest in deduplication, compression and archiving technologies, which can substantially reduce overall storage capacity requirements. "Going through consolidation can cut energy consumed by a storage system from one third to one half," Sullivan says.