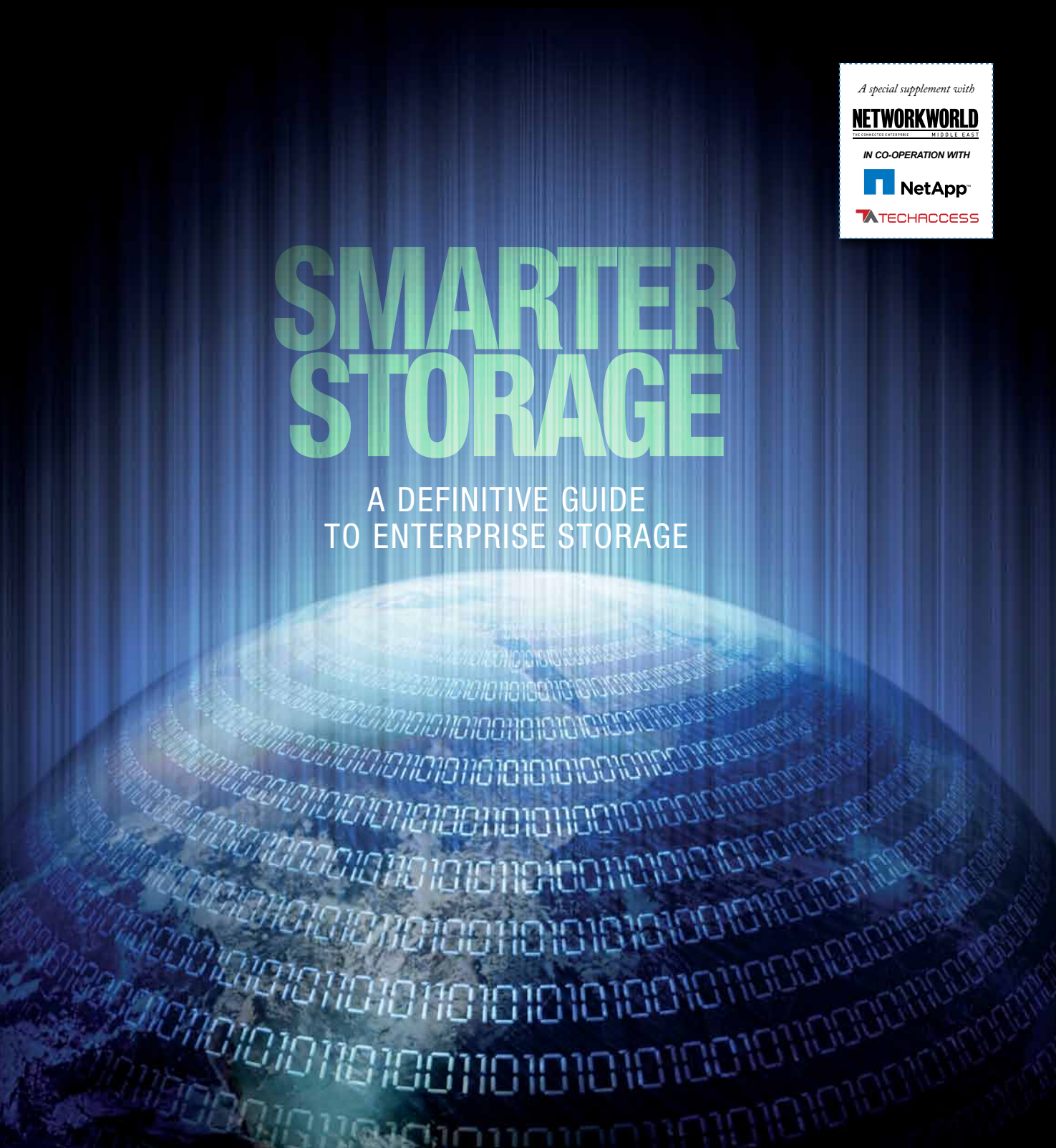
THIN PROVISIONING | DATA DEDUPLICATION

CIO SURVIVAL GUIDE

OCTOBER 2011

GUIDE TO ENTERPRISE STORAGE

A special supplement with
NETWORKWORLD MIDDLE EAST
IN CO-OPERATION WITH
NetApp
TECHACCESS



SMARTER STORAGE

A DEFINITIVE GUIDE
TO ENTERPRISE STORAGE

TECHACCESS
IN TOUCH WITH ENTERPRISE

Delivering technology that builds business

At TechAccess, we believe in Value Addition. We strive to provide state-of-the-art long lasting IT solutions for your business at a price that you can afford.

Our strategic alliances with global IT vendors ensure that we not only provide the latest cutting edge technology but also innovate and develop solutions tailor-made to your business needs.

Partner-Up

For Value Added Distribution, look no further than Tech Access!

Join the TechAccess Partner Up program Today. Whether you are interested in partnering with one, two or more Vendors, our Sales and Marketing teams have in-depth Knowledge about each vendors channel partner program.

We are here to help you with:

- Dedicated Accounts Management
- Pre & Post Sales Support
- Marketing & Sales Support
- Logistical & Financial Support
- Sales & Technical Trainings

Sign up today!
www.techaccess.com/partner_up.php

Partner-Up

DRIVE GROWTH AND DIFFERENTIATE YOUR BUSINESS NOW!

www.techaccess.com
Offices G01-G03, Building No.1
Dubai Internet City, Dubai, UAE
PHONE +971 4391-1820
FAX +971 4391-8875

TECHACCESS
IN TOUCH WITH ENTERPRISE

A VIRTUAL WORLD

IT initiatives such as server virtualisation and data centre consolidation are highlighting the need to modernise backup and recovery systems. Virtualisation, in particular, has disrupted traditional backup processes and forced companies to reconsider their approach.

One issue has to do with resource consumption and efficiency. "Taking a full backup copy of your data is incredibly resource intensive. It takes a fair amount of CPU, a fair amount of memory, and most importantly, a huge amount of I/O," says Dave Russell, research VP at Gartner. "If you were to kick-off full backup simultaneously for six to eight virtual machines on a physical server, it would saturate all the available resources of that physical server."

Another issue is recoverability in a virtualised environment. In some cases a company might want to bring back a full virtual machine. But in a lot of cases, the "disaster" is localised, such as the loss of a file or a corrupt application, and a company may only want to restore that one file, for instance. That kind of granularity has been lacking. "Traditional backup has had a hard time peeking into the virtual machines just to get the data that's required," Russell says.

Backup vendors have been addressing those issues and outfitting their software with new capabilities geared for virtualised environments. Increasingly, those new capabilities are being baked into disk-based data capture and data recovery products. As enterprises look to update their backup software, many are simultaneously moving away from tape-based solutions to disk-based backup

The current generation of enterprise disk-based backup and recovery tools tout features such as data deduplication, snapshots, heterogeneous replication, and real-time backups. Another disk-based backup technology that's getting enterprise attention is the use of storage array snapshots as part of an enterprise backup and recovery scheme.

While disk-based backup is on the rise, tape isn't going away anytime soon.

Six years ago, enterprises backed up directly to tape 63% of the time, Gartner's Russell says. Today, enterprises back up directly to tape -- without any disk - just 13% of the time. But that's not the whole story. Another 65% of enterprises have adopted a disk-to-disk-to-tape approach, whereby they back up to disk first and then write to tape. "That means 78% of the enterprise market has at least a little bit of tape somewhere in their backup process," Russell says.

Even vendors on the cutting edge of disk-based backup acknowledge there's a place for tape.

Today's enterprises can provide more levels of recovery than in the past via a multi-tiered backup and recovery approach. For the most critical systems, a large enterprise might opt to replicate data through high speed fiber-optic lines to a backup data center, for instance. For less critical resources, a company might opt to backup to disk and replicate that disk over the WAN, recognising that recovery might take a little more time, or the backup systems might run on slightly slower hardware. For long-term archiving - often for regulatory requirements - companies may keep using tapes, shipped offsite for storage.

CEO and Publisher	Dominic De Sousa
COO	Nadeem Hood
Managing Director	Richard Judd
Senior Editor	Jeevan Thankappan
Sales Director	Rajashree R Kumar
Art Director	Kamil Roxas
Designer	Froilan A. Cosgafa IV
Webmaster	Tristan Troy Maagma
Production	James P. Tharian
Circulation Manager	Rajeesh M subscribe@cpidubai.com

Published by

CPI

Head Office
PO Box 13700
Dubai, UAE
Tel: +971 4 4409100
Fax: +971 4 4293654
Web: www.cpidubai.com

Regional partner of

IDG
INTERNATIONAL DATA GROUP

5 WAYS TO CUT YOUR STORAGE FOOTPRINT

With the economy still shaky and the need for storage exploding, almost every storage vendor claims it can reduce the amount of data you must store

Trimming your data footprint not only cuts costs for hardware, software, power and data centre space, but also eases the strain on networks and backup windows.

But how do you know which technique to use? First you have to understand how your business uses data and determine when the cost savings of data reduction are worth the resulting drop in performance.

The technique that's best for you depends not so much on the industry you're in as it does on the type of data you store. For example, deduplication often doesn't deliver significant savings for X-rays, engineering test data, video or music. But it can significantly reduce the cost of backing up virtual machines used as servers, for example. Here are five techniques to help reduce your stored-data volume.

1 Deduplication

Deduplication -- the process of finding and eliminating duplicate pieces of data stored in different data sets -- can reduce storage needs up to 90%. For example, through deduplication, you could ensure that you store only one copy of an attachment that was sent to hundreds of employees. Deduplication has become almost a requirement for backup, archiving and just about any form of secondary storage where speed of access is less important than reducing the data footprint.

Data can be deduped at the file or block level, with different products able to examine blocks of varying sizes. In most cases, the more fine-grained assessment a system can do, the greater the space savings. But fine-grained deduplication might take longer and therefore slow data access speeds.

Deduplication can be done preprocessing, or inline, as the data is being written to its target; or postprocessing, after the data has been stored on its target. Postprocessing is best if it's critical to meet backup windows with fast data movement, says Greg Schulz, senior analyst at The Server and StorageIO Group. But consider preprocessing if you have "time to burn" and need to reduce costs, he says.

While inline deduplication can cut the amount of data stored by a ratio of about 20:1, it isn't scalable, and it can hurt performance and force users to buy more servers to perform the deduplication, critics say. On the other hand, Schulz says that postprocessing deduplication requires more storage as a buffer, making that space unavailable for other uses.

For customers with multiple servers or storage platforms, enterprise-wide deduplication saves money by eliminating duplicate copies of data stored on the various platforms. This is critical because most organisations create as many as

15 copies of the same data for use by applications such as data mining, ERP and customer relationship management systems.

Users might also want to consider a single deduplication system to make it easier for any application or user to "rehydrate" data (return it to its original form) as needed and avoid incompatibilities among multiple systems.

Schulz says primary deduplication products could perform in preprocessing mode until a certain performance threshold is hit, then switch to postprocessing.

Another option, policy-based deduplication, allows storage managers to choose which files should undergo deduplication, based on their size, importance or other criteria.

Thin provisioning means setting up an application server to use a certain amount of space on a drive, but not using that space until it is actually needed

2 Compression

Probably the most well-known data reduction technology, compression is the process of finding and eliminating repeated patterns of bytes. It works well with databases, e-mail and files, but it's less effective for images. It's included in some storage systems, but you can also find stand-alone compression applications or appliances.

Dedupe and compression: Better together? Some vendors offer, or will offer, both deduplication and compression. The choice of whether, when and in what order to use both compression and deduplication depends on factors such as whether compression will make it easier or harder for the deduplication software to scan for redundancies, what tier (primary vs. secondary) you're looking to optimize, and how quickly the product can return data to a usable form when needed.

Real-time compression that doesn't delay access or slow performance by requiring data to be decompressed before it's modified or read is suitable for online applications like databases and online transaction processing, says Schulz. The computing power within modern multicore processors also makes server-based compression an option for some environments, he adds.

Deduplication and compression are complementary. "Use compression when the primary focus is on speed, performance, transfer rates. Use deduplication where there is a high degree of redundant data and you want higher space savings," says Schulz.

3 Policy-Based Tiering

Policy-based tiering is the process of moving data to different classes of storage based on criteria such as its age, how often it is accessed or the speed at which it must be available. Unless the policy calls for the outright deletion of unneeded data, this technique won't reduce your overall storage needs, but it can trim costs by moving some data to less expensive, but slower, media.

4 Storage Virtualization

As is the case with server virtualisation, storage virtualization involves "abstracting" multiple storage devices into a single pool of storage, allowing administrators to move data among tiers as needed. Many experts view it as an enabling technology rather than a data reducer, per se, but others see a more direct connection to data reduction.

5 Thin provisioning

Thin provisioning means setting up an application server to use a certain amount of space on a drive, but not using that space until it is actually needed. As with policy-based storage, this technique doesn't cut the total data footprint but delays the need to buy more drives until absolutely necessary.

If storage needs increase rapidly, you must "react very, very quickly" to ensure that you have enough physical storage, says Allen. The more unpredictable your needs, the better measurement and management tools you need if you adopt thin provisioning. Schulz advises looking for products that identify both the data and applications users need to track, and that monitor not only space usage but read/write operations to prevent bottlenecks.

Before choosing a data reduction strategy, set policies to help make tough choices about when to pay for performance and when to save money by cutting your data footprint. Don't focus only on reduction ratios, Schulz says, but remember that you might get more savings with a lower reduction rate on a larger data set.

And don't be confused by vendor terminology. Compression, data deduplication, "change-only" backups and single instancing are all different ways of reducing redundant data. When in doubt, choose your storage reduction tools based on their business benefits and a detailed analysis of your data.

REDUCING STORAGE BLOAT

Enterprise data explosion threatens to overwhelm storage systems, particularly the backup tier. Here's how data deduplication can help

Backing up servers and workstations to tape can be a cumbersome process, and restoring data from tape even more so. While backing up to disk-based storage is faster and easier, and probably more reliable, it can also be more expensive.

One way to get the best of both worlds is to back up to disk-based storage that uses deduplication, which increases efficiency by only storing one copy of a thing.

While the process was originally used at the file level, many products now work at the block or sub-block (chunk) level, which means that even files that are mostly the same can be deduplicated, saving the space consumed by the parts that are the same.

For instance, say someone opens a document and makes a few changes, then sends the new version to a dozen people. With file-level deduplication, the old and new versions are different files, though only one copy of the new version is stored. With block-level or sub-block-level deduplication, only the first document and the changes between the first document and the second are stored.

There is some debate about the optimum process - deduplication of files is not very efficient, blocks, more so, chunks even more so. However, the smaller the chunks, the more processing it takes, and the bigger the indices are that keep track of duplicates. Some systems use variable size chunks to tune this, depending on the type of data being stored.

Deduplication was originally used only for backups - since backups tend to be run regularly and usually contain mostly the same data as the last backup, very high efficiencies can be obtained with deduplication. Now, however, deduplication is beginning to be seen in primary storage and other applications as well, such as the deduplication of snapshots and replication.

There are two main types of deduplication, in-line and post-processing. In-line looks at data as it is sent to the storage system, and only stores a file if it is not already on the system. Post-processing stores the file immediately and then scans all the data on the system at regular intervals to find and remove duplicate chunks of data.

In-line requires less storage, while post-processing requires a 'landing area' where data can be stored until it is deduplicated. On the other hand, since it must handle high-speed streams of data, in-line requires considerably more processing power which is expensive, while storage space is relatively cheap. Post-processing might be scheduled for once a day, following the end of the backup window. Since backups are typically run during the periods of lowest activity, post-processing might be scheduled for the start of the business day. Since the deduplication storage isn't typically used for anything other than backups, this doesn't impact users.

Many companies no longer run backups of data directly - if a database

The snapshots or replication are added features that may then be used to restore files deleted accidentally, which is much faster than restoring from a backup

is in use, it must be locked to run a backup. With the 24x7 availability requirements many businesses have, the simpler process is to take a snapshot of the data or use the replication capability of SAN storage, then run a backup from the snapshot or replica. Some vendors have added deduplication to the snapshot and replication functions of their storage, so that only the differences (deltas) between the last snapshot and the current one are stored. This makes it possible to take regular snapshots of data without increasing the space required very much.

The snapshots or replication are added features that may then be used to restore files deleted accidentally, which is much faster than restoring from a backup. Some organisations may even decide to use snapshot or replication instead of backups. The difficulty here is that SAN storage is often expensive, and snapshots of replication is an added feature with extra licensing costs. Backup appliances with deduplication can be a less expensive way of protecting data.

Finally, vendors are beginning to bring deduplication technology to primary storage. This is a more complex process than deduplicating backups. A backup is written once and then not changed, while primary storage has many users making changes or creating new files throughout the day. The issues of keeping track of files in use, ensuring that duplicate files aren't lost if the 'original' is deleted and redundancy for the indexing system that keeps track of duplicates so that a loss of data in the index doesn't lose user data are all problems that are not simple to solve.

An issue with deduplication generally is that as the size of the data being deduplicated grows, the amount of memory necessary to process the files or blocks generally also grows, which can limit the overall size of a system - the total size of storage is limited by the memory of the processor. This may mean that it is simpler to have a backup appliance for every few servers rather than one large backup appliance for all servers.

It can also be difficult to estimate the actual capacity of a deduplication appliance. If the data being stored is mostly the same, and is also compressible, it is quite possible to get 20T to 100TB of data into a system with four 2TB drives (raw capacity of 6TB with RAID).

On the other hand, if the data being backed up is usually different from backup to backup, and not very compressible - an extreme example would be a working area for video files, holding several different large video files (which are already compressed) that change daily, then there might be very little gain from deduplication.

A DIFFERENT APPROACH

Five ways to successfully manage dynamic storage infrastructures

Don't look now, but the data centre is transforming. Seeking greater efficiencies, enterprises of all sizes have steadily evolved from application-based silos to virtualised environments. Now this evolution is taking the next leap forward as innovative enterprises move toward next-generation cloud computing models that deliver IT as a service (ITaaS) – via both internal and external cloud services.

Delivering on-demand cloud services through a multi-tenant architecture offers attractive benefits, including greater agility, data mobility and scalability with reduced capital expenditures and greater control over service costs. But it also presents some significant management challenges, especially when it comes to managing the dynamic storage resources required by ITaaS infrastructures.

To tap the full potential of cloud computing to drive efficiency, IT managers must maintain – and, in many cases, improve – service quality and availability. But managing storage resources in these highly complex, dynamic and heterogeneous environments can be a major challenge.

One challenge facing customers is that many simply lack the required insight into their storage environments. A prime example of this can be seen through a recent customer visit. The customer was reengineering virtual machines within the data centre to reduce server costs. Unfortunately he was unaware of the impact this change would have on storage growth over time.

A quick analysis showed almost all of the gains the customer was achieving through the move to a dynamic server infrastructure were being negated by the increased storage infrastructure costs. This was a direct result of the customer simply not having the requisite optimisation, capacity planning and visibility into the storage environment.

Meeting these challenges requires a next-generation approach to storage management focused on reducing the risk of data center transformation. This approach should address the following five critical success factors:

1 Global storage visibility: How can you manage your storage environment if you can't see it all? You need tools that provide a single, end-to-end view of storage – both physical and virtual – across the entire multivendor, multiprotocol environment, including host-to-storage access paths. Extending this visibility beyond the hosts to the applications themselves is essential to fully understand the impact of storage service levels on your business applications. The result is a "service-level" view of your storage infrastructure that is critical for ensuring the highest levels of service performance and availability, according to defined service level policies, while managing storage resources for maximum efficiency.

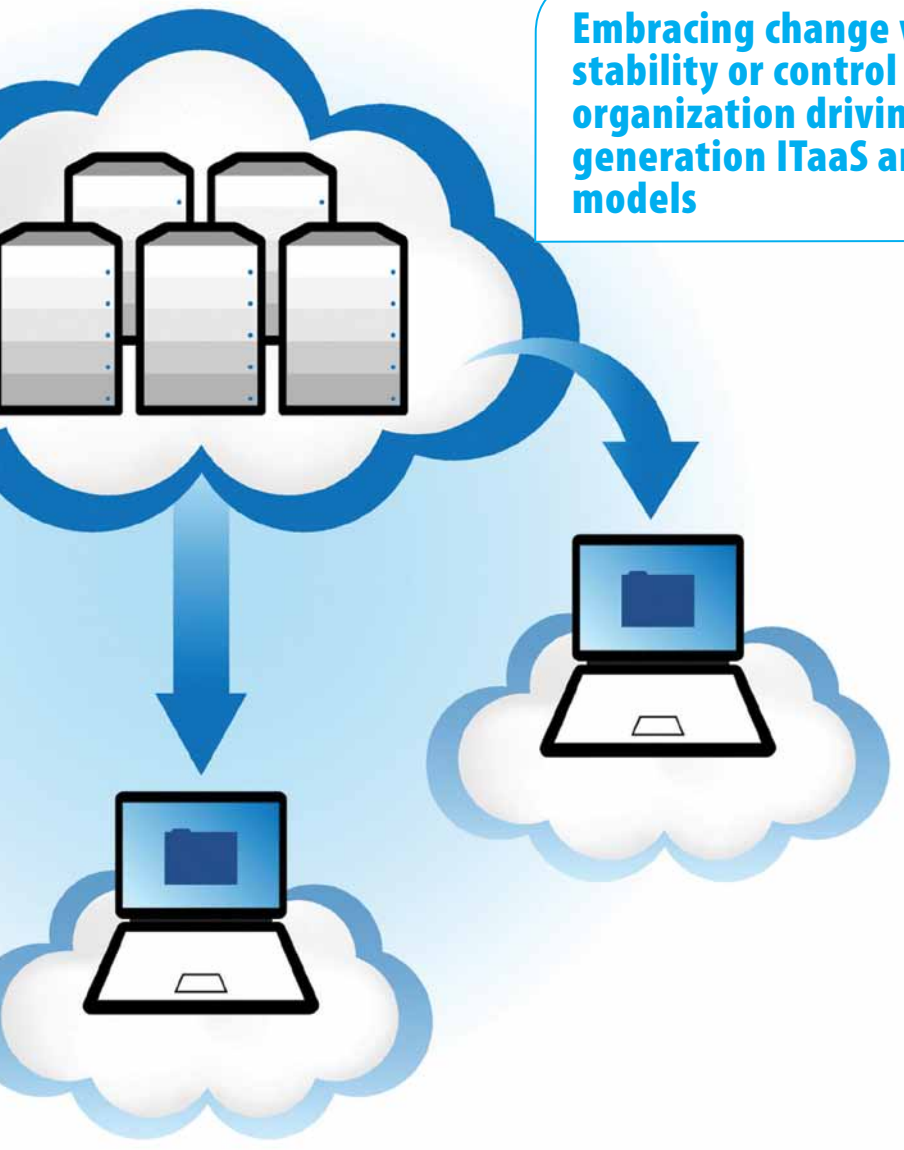
2 Comprehensive, continuous monitoring: Periodically checking the health of dynamic individual storage systems won't do. You must be able to monitor all shared resources continuously – and store that information in a single data warehouse. Proactively monitoring for service violations and latency issues is crucial to identify potential problems before they impact

service quality or availability. And having the ability to generate reports on your entire end-to-end storage infrastructure goes beyond administrative capabilities, giving you powerful management insight for optimising shared resources and controlling costs.

3 Centralised change management: In the evolving data center, nothing is so constant as change. Yet every change introduces additional complexity and potential risk. You need a solution for planning configuration changes that encompasses the entire infrastructure. Having the ability to perform "what if" configuration simulations is especially important, enabling you to see through the complexity and clearly understand the impact of proposed changes. This enables you to identify potential service-level violations to your business applications, so you can make informed decisions before changes are made. This is critical for managing risk, whether you're deploying a single new storage array, planning a major migration or simply upgrading existing firmware levels.

4 Global capacity management: Optimising utilisation of shared storage is crucial to achieving the full efficiency benefits of virtualisation and cloud computing. IT managers need tools that provide continuous, global visibility into storage resource allocation across the heterogeneous data center. Leveraging a centralized data warehouse for all storage utilisation information facilitates meaningful analysis of capacity consumption trends, forecasting trends and chargeback. When new capacity is needed, providing near-real-time visibility of available service tiers helps accelerate provisioning and avoid potential conflicts.

Embracing change without compromising stability or control is the goal of any IT organization driving the adoption of next-generation ITaaS and cloud computing models



5 Elegant simplicity: Tools designed to help you manage complexity shouldn't add more complexity to your environment. Investing in technologies that monitor, measure and analyse equipment from any vendor, based on any protocol, helps avoid a "patchwork" approach of storage management tools that don't provide a unified view of your entire infrastructure. Also, given the dynamic and non-transparent nature of virtual infrastructure environments, it's important to use agent-less solutions that don't require a lot of care and feeding.

Embracing change without compromising stability or control is the goal of any IT organisation driving the adoption of next-generation ITaaS and cloud computing models. The key to managing the risk of data center transformation is having a clear picture of what's really happening across the entire dynamic, complex environment and how it affects the service levels and performance of your business applications.

Taking advantage of storage management technologies that adhere to the five success factors described above can help you achieve this visibility -- and realise the promise of an optimised, efficient, next-generation data centre.

THIN PROVISIONING

While provisioning all the capacity of an external disk to a given application, known as full provisioning, ensures the app has plenty of growth potential, it results in poor utilisation rates, a costly problem that can be addressed with thin provisioning technology

Research shows that storage utilisation rates achieved by most companies is 40% or lower. That means buyers are acquiring more capacity than they really need and the very existence of that extra capacity requires more space apond cooling.

Furthermore, the traditional method of provisioning leads to increased management workloads due to the fact that the extra but unused capacity still needs to be monitored and managed. If applications reach their capacity limits and IT managers have to re-provision capacity, complex management tasks can be involved. More management requires more human resources, further driving up costs associated with storage management.

Additionally, if an application is taken offline to reprovision capacity, it is then unable to serve business needs and can lead to revenue loss.

Thin provisioning provides a way to address these limitations. By automatically allocating system capacity to applications as needed, thin

provisioning technology can help achieve up to 90% storage utilisation, while at the same time significantly reducing power consumption.

Thin provisioning allows users to allocate a large amount of virtual capacity for an application, regardless of the physical capacity actually available. At initial setup, thin provisioning does not physically allocate capacity to the prescribed data volume, and the actual space is used only when data writes occur.

This on-demand method for capacity allocation not only optimizes storage utilisation, but also greatly simplifies capacity planning and management. In order to help users easily monitor capacity utilisation, storage systems automatically issue notifications when the total capacity utilisation is reaching the threshold set by the user. If users wish to expand capacity, they can do so non-disruptively.

With traditional provisioning, it is difficult to move data across logical partitions in a storage architecture. If thin provisioning is applied, storage capacity from different logical partitions can be consolidated, enabling it to be dynamically allocated. From the opposite perspective, this means that the storage controller can move data dynamically across logical partitions based on how resources are designed to function.

Furthermore, thin provisioning opens the door for other advances in storage design, includingautomated storage tiering. Storage tiering involves grouping data into different categories and assigning these categories to different types of storage media in order to optimise storage utilisation.

Automated tiering ensures applications have access to the performance levels they need. High-performance applications can be assigned to high-performance tiers featuring drives such as SSDs or SAS, while applications requiring less performance can be assigned to lower tiers featuring low-performance drives such as SATA.

This ensures that no storage resources are wasted and that applications can function properly. In addition, this technology helps automatically migrate data based on usage patterns. If data in higher tiers has not been used for an extended period of time, it is demoted to lower tiers. Conversely, if data in lower tiers is frequently accessed, it is promoted to higher tiers. Storage efficiency can be greatly improved with this technology.

In terms of tangible benefits, the overarching result of using thin provisioning is it helps significantly reduce costs

The benefits in a nutshell

In terms of tangible benefits, the overarching result of using thin provisioning is it helps significantly reduce costs. With thin provisioning, utilization can be greatly increased and you can get more out of your existing capacity, reducing the need to add new HDDs (hard disk drives). This can generate significant cost savings, which can be particularly important for companies with limited budgets.

These benefits are highlighted by the figure below. With thin provisioning, capacity is dynamically allocated to applications from a consolidated storage pool, eliminating the need for the allocated but unused capacity in full provisioning.

The benefits of delaying new HDD acquisitions are further magnified by the fact that HDD prices continue to decline. Delaying acquisitions thereby becomes even more worthwhile.

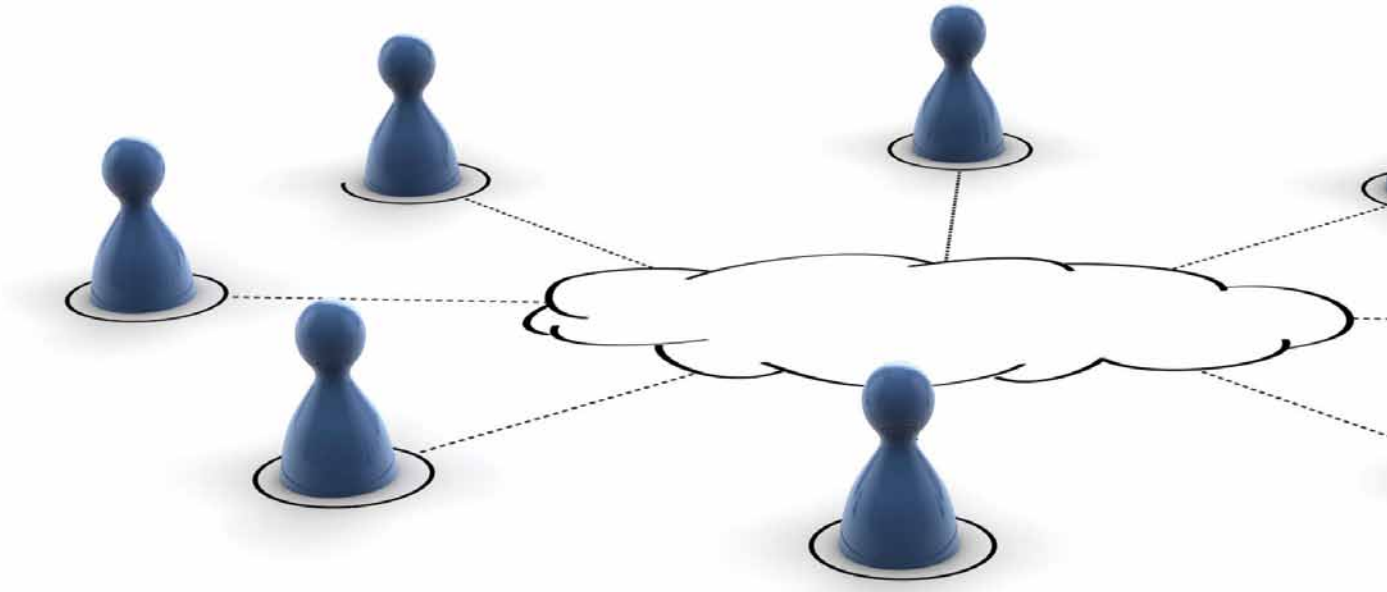
Another major benefit is power consumption can be reduced considerably, leading to more cost savings. Since you do not need to buy as many HDDs as before, the rack space needed for a storage solution can also be reduced.

Storage management also becomes much easier with thin provisioning. There is less to manage with a smaller number of HDDs, and you spend less time allocating capacity to applications. Furthermore, when capacity limits are reached, capacity can be added nondisruptively, eliminating the hassles associated with downtime.

If you end up turning to thin provisioning for the benefits outlined, here are a few tips to help you get the most out of the technology:

- Claim as much virtual capacity for applications as possible: This reduces management tasks and ensures applications have access to sufficient capacity as long as physical resources are available.
- Monitor physical storage capacity utilisation: Even though large virtual capacities can be allocated, each storage pool still has physical limits. Monitoring physical capacity utilisation is thus extremely important.
- Set physical capacity utilisation thresholds based on pace of data generation: By creating notification mechanisms for when these thresholds are reached, sufficient time will be available to properly plan storage expansion. Setting a correct threshold is most important. For example, if an application fills 10% of capacity within only a few days, setting the utilisation threshold at 90% could lead to major problems, as the user will not have enough time to plan capacity expansion in order to ensure that an application can continue to operate normally.

The benefits of thin provisioning are numerous and can generate considerable cost savings. With thin provisioning technology now widely available in the storage market, it should be a key consideration for companies planning to acquire new storage solutions.



KEY STORAGE TRENDS

What is reshaping the storage landscape?

With 15 petabytes of new data being created every day, and the volume of data continuing to grow, the ability to simply store data in a secure and easily-accessible manner presents a challenge for IT departments and vendors.

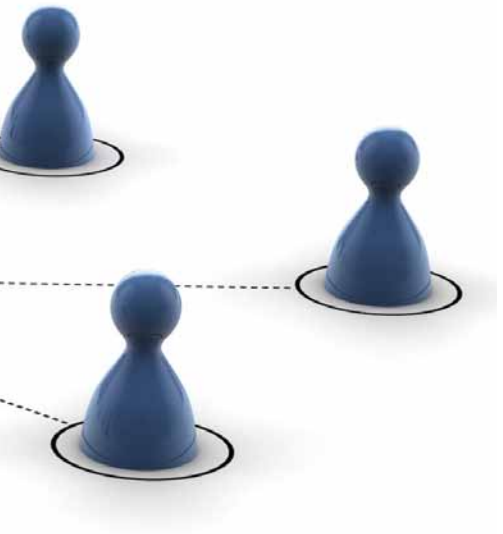
Traditional methods of storage are proving costly, inefficient and unable to keep pace with client needs; and new, more efficient and intelligent methods of storage are beginning to take their place. The new face of storage for this year looks something like this:

1. Storage as a cloud service gains momentum.
As the volume of data grows at an exponential rate – with sources like IDC predicting a growth rate of nearly 60 percent per year – we're beginning to see a variety of different cloud storage offerings become attractive, including private, public and hybrid models. All three offer enterprises large scale, better planning capabilities and more flexibility to meet unique user demands; but with growing concerns over the vulnerability of critical information on public clouds, the majority of investment this year will go to the private model.

In fact, Gartner predicts that Global 1000 IT organisations will spend more money building private cloud computing services through 2012 than they will on offerings from public cloud-computing service providers. As those customers begin to seek out vendors who can not only store their critical information safely, but can also manage and host their cloud environments, those vendors who are investing in the cloud now will come out on top in the year ahead.

For enterprises that are driving the shift to the cloud, one thing quickly becomes apparent: status quo won't do

2. Goodbye storage status quo
For enterprises that are driving the



shift to the cloud, one thing quickly becomes apparent: status quo won't do. Old processes of storing and managing information can't continue to support the new model. The increasing data deluge matched with the economic downturn is forcing innovation.

New solutions are appearing that integrate virtual storage with other virtual system elements to create fully virtual systems, and there is going to be an increase in software aggregation that will lead to large scale, high performance or high availability solutions for a lower cost. Customers are also going to be seeking improved system efficiencies with real time compression capabilities, de-duplication and data migration.

3. Converged infrastructure

These solutions will lead to a converged infrastructure that is held together through service management software. As categorisation and automated data placement are used to ensure the right resources are being applied to the correct business need, the demands on storage will only increase. Customers will seek vendors that can offer a layer of service management that is able to integrate with servers, storage, software and networking under one offering.

Customers need solutions that can optimize efficiency in the use of infrastructure assets; improve response to hardware failures, new policies or new business opportunities; and adjust operations of the infrastructure against the business goals and policies, all under one roof.

4. Smart data categorisation

There's also a growing need for more intelligent data categorisation, as the amount of data outgrows IT departments' budget to contain it, and as regulatory requirements become stricter for data security, retention and access. Data must be managed to meet these needs and to fulfill its potential for being used as a tool for business innovation. It will be categorised for optimal placement–with "hot data" that is available and ready to access and "cold data" that can be stored for later use–and it will be tiered to allow clients to manage critical and secondary data differently.

Once categorised, data can then be accessed using a mix of various storage technologies, including solid state drives, flash, disk and/or tape storage, depending on the requirements.

5. New roles for tape. Yes, tape.

We're also seeing new opportunities for tape storage. While disk-based storage has grown in popularity recently, tape storage is experiencing resurgence thanks to the increase of digital archives. In fact, tape will be the preferred medium for 80 percent of all data in electronic archives, according to industry experts.

Many factors are driving continued use of tape: Data can be stored on tape at one fifth the cost of disk; it's 290 times less expensive in terms of energy costs; it can last 30 years or more; and the continued commitment from vendors to keep the technology relevant ensures tape will not become obsolete.

Better storage tech could trim energy, cooling costs

The migration to smaller drives and the use of solid-state drives (SSD) are among several key technologies that will lead to much lower data center power and cooling costs over the next several years, IDC said.

According to IDC , more efficient storage technologies should allow power and cooling costs in data centers to level off by 2014.

"There is no doubt that the economic straits of 2008 and 2009 modified the attitudes and behaviors of IT managers and system [manufacturers], greatly accelerating interest in and the adoption of more cost-efficient storage strategies," said David Reinsel, group vice president for Storage Systems at IDC.

Reinsel said the power and cooling plateau is not likely a sustainable trend, and that ever-growing capacity requirements will result in "renewed growth in energy costs."

But for now, the migration to smaller 2.5-in hard disk drives (replacing 3.5-in drives), the continued adoption of SSDs and the uptake in technologies that make more efficient use of existing capacity, such as data deduplication, compression and thin provisioning, will reduce overall energy costs, IDC said.

According to IDC's report, enterprise data storage systems remained a key area of investment for CIOs and IT managers throughout the economic downturn, driven by continuous pressure to store more data. External storage system shipments increased 38% from 2008 to 2009. At the same time, hard disk drive shipments for external storage systems increased 10%.

At the same time IT managers were buying more storage, interest in storage efficiency technologies grew because IT budgets turned flat or declined, IDC said. While technologies such as data deduplication and thin provisioning are most effective when first deployed, migrating to components that consume less power, and generate less heat, offers important energy savings over the long term.

STORAGE VIRTUALISATION

In just a few short years, storage virtualisation has proven its worth in the large enterprise and traveled that well-worn path from pricey boutique solution to affordable commodity

As a standard feature in all but the most modest mid-tier storage arrays, storage virtualisation soothes a wide range of storage management woes for small and mid-size organisations. At the same time, dedicated solutions from top-tier vendors deliver the greatest ROI to large shops managing large SANs with intense data availability requirements.

Storage virtualisation creates an abstraction layer between host and physical storage that masks the idiosyncrasies of individual storage devices. When implemented in a SAN, it provides a single management point for all block-level storage. To put it simply, storage virtualisation pools physical storage from multiple, heterogeneous network storage devices and presents a set of virtual storage volumes for hosts to use.

In addition to creating storage pools composed of physical disks from different arrays, storage virtualisation provides a wide range of services, delivered in a consistent way. These stretch from basic volume management, including LUN (logical unit number) masking, concatenation, and volume grouping and striping, to thin provisioning, automatic volume expansion, and automated data migration, to data protection and disaster recovery functionality, including snapshots and mirroring. In short, virtualisation solutions can be used as a central control point for enforcing storage management policies and achieving higher SLAs.

Perhaps the most important service enabled by block-level virtualisation is nondisruptive data migration. For large organisations, moving data is a near-constant fact of life. As old equipment comes off lease and new gear is brought online, storage virtualisation enables the migration of block-level data from one device to another without an outage. Storage administrators are free to perform routine maintenance or replace aging arrays without interfering with applications and users. Production systems keep chugging along.

Four architectural approaches

In a virtualised SAN fabric, there are four ways to deliver storage virtualisation services: in-band appliances, out-of-band appliances, a hybrid approach called split path virtualisation architecture, and controller-based virtualisation. Regardless of architecture, all storage virtualisation solutions must do three essential things: maintain a map of virtual disks and physical storage, as well as other configuration metadata; execute commands for configuration changes and storage management tasks; and of course transmit data between hosts and storage. The four architectures differ in the way they handle these three separate paths or streams -- the metadata, control, and data paths -- in the I/O fabric. The differences hold implications for performance and scalability.

An in-band appliance processes the metadata, control, and data path information all in a single device. In other words, the metadata management and control functions share the data path. This represents a potential bottleneck in a busy SAN, because all host requests must flow through a single control point. In-band appliance vendors have addressed this potential scalability issue by adding advanced clustering and caching capabilities to their products. Many of these vendors can point to large enterprise SAN deployments that showcase their solution's scalability and performance.

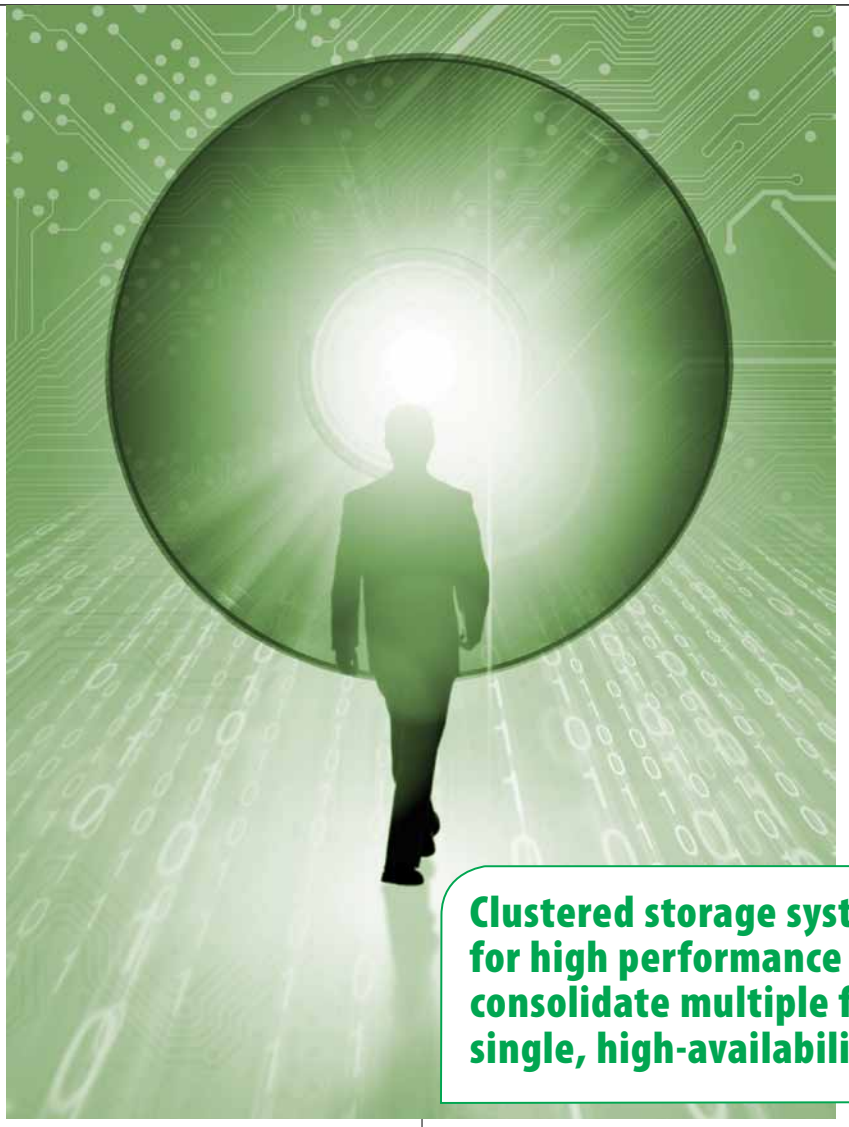
An out-of-band appliance pulls the metadata management and control operations out of the data path, offloading these to a separate compute engine. The hitch is that software agents must be installed on each host. The job of the agent is to pluck the metadata and control requests from the data stream and forward them to the out-of-band appliance for processing, freeing the host to focus exclusively on transferring data to and from storage.

A split path system leverages the port-level processing capabilities of an intelligent switch to offload the metadata and control information from the data path. Unlike an out-of-band appliance, in which the paths are split at the host, split path systems split the data and the control paths in the network at the intelligent device. Split path systems forward the metadata and control information to an out-of-band compute engine for processing and pass the data path information on to the storage device. Thus, split path systems eliminate the need for host-level agents.

Array controllers have been the most common layer where virtualisation services have been deployed. However, controllers typically have virtualised only the physical disks internal to the storage system. This is changing. A twist on the old approach is to deploy the virtualisation intelligence on a controller that can virtualise both internal and external storage. Like the in-band appliance approach, the controller processes all three paths: data, control, and metadata. The primary example of this new style of controller-based virtualisation is Hitachi Universal Storage Platform.

File virtualisation

Just as block virtualisation simplifies SAN management, file virtualisation eliminates



much of the complexity and limitations associated with enterprise NAS systems. We all recognise that the volume of unstructured data is exploding, and that IT has little visibility into or control over that data. File virtualisation offers an answer.

File virtualisation abstracts the underlying specifics of the physical file servers and NAS devices and creates a uniform namespace across those physical devices. A namespace is simply a fancy term referring to the hierarchy of directories and files and their corresponding metadata. Typically with a standard file system such as NTFS, a namespace is associated with a single machine or file system. By bringing multiple file systems and devices under a single namespace, file virtualisation provides a single view of directories and files and gives administrators a single control point for managing that data.

Many of the benefits will sound familiar. Like storage virtualisation, file virtualisation can enable the nondisruptive movement and migration of file data from one device to another. Storage administrators can perform routine maintenance of NAS devices and retire old equipment without interrupting users and applications.

File virtualisation, when married with clustering technologies, also can dramatically boost scalability and performance. A NAS cluster can provide several orders of magnitude faster throughput (MBps) and IOPS than a single NAS device. HPC (high performance computing) applications, such as seismic processing, video rendering, and scientific research simulations, rely heavily on file virtualisation technologies to deliver scalable data access.

Three architectural approaches

File virtualisation is still in its infancy. As always, different vendors' approaches are optimally suited for different usage models, and no one size fits all. Broadly speaking, you'll find three different approaches to file virtualisation in the market today: Platform-integrated namespaces, clustered-storage derived namespaces, and network-resident virtualised namespaces.

Platform-integrated namespaces are extensions of the host file system. They provide a platform-specific means of abstracting file relationships across machines on a specific server platform. These types of namespaces are well suited for multisite collaboration, but they tend to lack rich file controls and of course they are bound to a single file system or OS.

Clustered storage systems combine clustering and advanced file system technology to create a modularly expandable system that can serve ever-increasing volumes of NFS and CIFS requests. A natural outgrowth of these clustered systems is a unified, shared namespace across all elements of the cluster. Clustered storage systems are ideally suited for high performance applications and to consolidate multiple file servers into a single, high-availability system.

Clustered storage systems are ideally suited for high performance applications and to consolidate multiple file servers into a single, high-availability system

Network-resident virtualised namespaces are created by network-mounted devices (commonly referred to as network file managers) that reside between the clients and NAS devices. Essentially serving as routers or switches for file-level protocols, these devices present a virtualised namespace across the file servers on the back end and route all NFS and CIFS traffic as between clients and storage. Network-resident virtualised namespaces are well suited for tiered storage deployments and other scenarios requiring nondisruptive data migration.

File and block storage virtualisation may be IT's best chance of alleviating the pain associated with the ongoing data tsunami. By virtualising block and file storage environments, IT can gain greater economies of management and implement centralised policies and controls over heterogeneous storage systems. The road to adoption of these solutions has been long and difficult, but these technologies are finally catching up to our needs. You will find the current crop of file and block virtualisation solutions to be well worth the wait.

SUPERIOR STORAGE

UAE Exchange accelerates and secures money transfers using NetApp storage solutions

UAE Exchange (UAEX) is a global leader in remittance and foreign exchange. Since the opening of the first branch in Abu Dhabi in 1980, direct operations have spread across five continents. Award-winning services powered by a reliable network that extends across the world enable money transfer and exchange needs to be met at all times. Faced with the increasing mobility of customers and businesses, UAEX relies on innovation and technology to maintain customer satisfaction, which won the company the accolade "World's Trusted Money Transferrer."

The challenge

Accommodate data growth and implement a disaster recovery plan

In 2008, UAEX experienced strong growth in demand for its services, straining the HP EVA3000 storage system that was then in place. "The number of transactions were tripling per year, and the increase in the mobility of workers, money transfers, and payments was not limited to local time zones, requiring 24/7 service uptime. We were reacting with quick rollouts of application enhancements, but the storage capacity remained a constant bottleneck," says Sudhir Shetty, COO of Global Operations at UAEX.

For the CIO of UAEX, Sarath Chandra, it became a top priority to ensure that the company's storage environment could deliver responsive, reliable, and efficient performance to meet the business needs: "The storage solution had been in place for five years and could no longer keep pace with the massive growth both in terms of business volume as well as product offerings. There were other issues beyond lack of scalability, including hardware compatibility problems and dissatisfaction with the technical support of the infrastructure in place. Also, in terms of disaster recovery, we didn't have equal storage solutions at the primary and secondary data centres, which was an additional worry."

As a global leader in remittance and foreign exchange, with direct operations across five continents, reliable 24/7 IT services are required. Any interruption is not only costly but is contrary to the reputation that UAEX built over the past 30 years. The backup and recovery window of the previous system also presented multiple challenges and high overhead for the UAEX IT department, negatively affecting their internal commitment to fulfill required SLAs for their business applications.

With the legacy storage system proving inefficient and difficult to manage, impairing both uptime and the speed at which new business applications could be brought on board, UAEX started to look for a new solution. The company wanted a solution that was designed and built not only to accommodate peak loads with no performance penalty but one that was also flexible and easy to manage. With the solution it chose and implemented, UAEX

now benefi ts in terms of redundancy and capacity from new storage features that the old infrastructure did not offer.

The solution

Standardise on NetApp Unified Storage Architecture

UAEX worked with NetApp Star Partner Fujitsu Technology Solutions and the NetApp team of system engineers to determine which hardware and software options were best suited to fulfi ll the goals set out by the IT team. "We were looking for a stable, scalable, high-performance storage solution which would give us an attractive ROI. Apart from the TCO, we placed particular emphasis on the quality of the post-sales technical support and scalability," says Chandra.

Other storage alternatives, including EMC and HP, were considered, but NetApp's simplified, unified architecture proved the most appropriate and most effi cient storage solution. "We chose NetApp as it scored high on all decisive factors, and we were especially impressed with the sincerity of NetApp's technical expertise in clarifying all our apprehensions by sharing a lot of useful technical details and benefi ts. On the whole, we were looking for a simple, scalable, cost effective solution and NetApp was the best fit," Chandra states.

Two identical NetApp FAS2050s with a raw capacity of 28TB were implemented, one in the primary offi ce in Abu Dhabi and one for data replication purposes in the Etisalat hosting facility in Al Ain. The systems are used for block-level data through FCP and iSCSI protocols. NetApp deduplication software is used to enhance the effi ciency of storing data and hence to gain extra capacity for more services.

NetApp NearStore is used to enable archiving data in the same system without additional investment in a dedicated solution for archiving purposes. NetApp SnapDrive is used to optimise the used capacity on disk through the thin-provisioning feature, in which servers are able to use more capacity than the actual reserved space while enabling the administrator to manage all the allocated disks using a single piece of management software.

NetApp SnapRestore is used to enable quick recovery from any logical data corruption to any point in the past and NetApp SnapMirror is deployed to provide uninterrupted services in case of site failover across cities. UAEX's core business platforms are based on Red Hat Linux and Java applications hosted on HP blade servers. The applications are developed in house based on J2EE standards. The backend storage is on Sybase ASE, with BI on Sybase IQ and Hyperion. For data replication, UAEX depends on Sybase Replication Server.

Business benefits

Improved performance, reliability, and risk management

The high performance, reliability, and ease of management of UAEX's storage environment have greatly reduced the time the team spends on maintenance, allowing them to shift time to other tasks. "We have estimated a gain of 75% in time spent on maintenance tasks when we compare to the legacy system [that was] in place," comments Chandra.

UAEX now relies on NetApp Snapshot technology to create point-in-time copies of its database. With the previous environment, backups were laborious and time consuming for the UAEX IT team. UAEX now takes preventive Snapshot copies of its data in conjunction with SnapRestore, allowing staff members to retrieve data rapidly and restore it to a certain point in the past. This feature makes data management easier and the customer values the benefi ts of UAEX being able to respond to critical technical challenges in a much more effi cient way.

Now there is no need to worry about time consuming data restoration from tapes or even tape retention. With NetApp, the IT team can also rely on the advanced effi ciency features to fine-tune the fi gures for recovery time objectives and recovery point objectives according to the business-critical applications. The team can also replicate sensitive data to a different site located in another city using NetApp thin replication, while further optimizing capacity and bandwidth through NetApp block level deduplication. The ability to optimize the backup and recovery window through NetApp technology offers clear value and benefi ts to meet UAEX's challenging SLAs on IT delivered services.

UAEX now relies on NetApp Snapshot technology to create point-in-time copies of its database. With the previous environment, backups were laborious and time consuming for the UAEX IT team

Achieved ease of use and implementation

The most important benefi ts have been the ease of implementation and the guidance from NetApp engineers. "Though the data migration from our previous storage to NetApp was a massive exercise, at the end the performance of NetApp in the production environment was impressive. For users, there was no negative impact or downtime. As a second-phase improvement, we added storage capacity and implemented NFS and CIFS. All of this was done as if they were absolutely simple confi gurations," adds Chandra.

50% space savings

NetApp deduplication software has had a very positive impact in terms of capacity space savings and thus energy costs. By reclaiming the capacity of every duplicated block, UAEX has registered savings of up to 50%. "NetApp deduplication was an important selling point for us. If we had to deploy 50% more capacity, then we would have to use more cooling power and rack space," confi rms Chandra.

Scalability poised for growth

One of the key pain points of the former solution was its inability to scale. The NetApp solution offers greater fl exibility. "As with any major implementation of this type, considering the demands of growth for the future is a key criterion. We consider that the present solution will hold good for a period of three years. An important advantage of the solution is the ability to scale the capacity without changing the core part, the controllers," says Chandra.

With help from Fujitsu Technology Solutions and NetApp, UAEX now has a storage infrastructure worthy of the company's reputation as a "World's Trusted Money Transferrer." It delivers scalable, reliable, cost-effi cient performance while enabling new applications to meet the demands of a growing customer base.

UNIFIED APPROACH, REAL BENEFITS

ING Bank Turkey improves storage efficiency and meets data protection challenges with NetApp

ING Bank Turkey began its operations in 2007 when the local Turkish bank Oyak was acquired by the Dutch ING Group. Today ING Bank has nearly 330 branches located in Turkey and a 7/24 customer call center in the commercial and consumer banking market segments.

Recognising consumer banking as one of the principle areas of its future growth and expansion, ING Bank has one of the most extensive ATM networks in operation in Turkey. Consumer banking products and services, which are intensively driven by technology, already contribute increasingly more to the bank's business volumes and noninterest income.

ING Bank is positioned to be a bank focused on customer satisfaction in Turkey's rapidly changing banking services market, extensive physical and electronic distribution network, talented human resources, and advanced IT base. Innovation, quality, and an effectiveness-based approach to business represent the axis of ING Bank's growth and development

The challenge

Reduce cost while improving the storage infrastructure

ING Bank Turkey encountered some problems with its storage infrastructure because all data, including the core banking applications, were stored on one storage system. Hakan Gürdal, assistant VP for Small and Medium System Management at ING Bank Turkey, says, "Since all applications had to share the same infrastructure, maintenance and costs were high, performance was impacted, and operational flexibility was limited. Feature integration was almost impossible, and data replication and restores were time consuming. Our project to virtualise most of our servers to maximise space and reduce cost had to be put on hold due to the limitations of the infrastructure."

Faced with these constraints, storage became a key issue for both the IT team and executive management. The bank launched an open request for proposal (RFP) including detailed specifications on the storage requirements that would meet the following goals:

- Improve application availability through hardware redundancy.
- Create a more efficient backup/restore environment with diskto-disk-to-tape and site-to-site backup.
- Make maximum use of available disk capacity through efficient utilisation.
- Centralise management of all storage through one console.
- Leverage consolidation and virtualisation through separation of storage from servers.

"We were clear from the outset that this was not about boxes, but a solution-oriented RFP which would satisfy not only technical specs but also business needs," says Hüseyin Sivri,CIO at ING Bank Turkey. "Our objective was to split the core banking from all the other applications and migrate these onto a new storage environment."

The bank deploys Microsoft Exchange Server, Microsoft SQL Server, and Oracle databases, including data warehousing, and operates Linux and UNIX environments as well as hundreds of servers. Each environment has its own set of requirements.

Three vendors submitted proposals. The project team paid particular attention to the migration, future integration, and replication capabilities of the proposed offers. The bank deploys a multi-data-center concept, with its primary site located in Istanbul. A second data center, also in Istanbul, was to become its exact mirror.

A third data centre, hosted in Izmir, serves disaster recovery purposes. "We needed synchronous mode replication between the primary and secondary sites

and asynchronous replication to the disaster recovery site. Not all vendors could offer these capabilities," says Gürdal.

The bank's project committee of IT experts and executive staff chose NetApp with an 80% consensus. The committee had requested a total cost of ownership (TCO) scenario that placed NetApp ahead of the other vendors and guaranteed a specific TCO over five years.

"Although part of this project was cost driven, we recognised the opportunity to gain essential new management capabilities embedded in the NetApp solution. These tools would enable us to enhance all our environments and especially bring our virtualisation project to life," adds Hakan İnceoğlu, senior vice president of Information Technology, ING Bank Turkey.

The solution

Migrate all applications to NetApp unified storage

NetApp's migration and integration capabilities were immediately put to the test starting with the entire virtual environment, followed by the Microsoft SQL and Exchange migration. The migration process was performed progressively over a 22-week period.

Says Gürdal, "We had 500 virtual servers, the Exchange system supporting 6,500 mailboxes, and 13 file servers with a capacity of 12 terabytes which we needed to consolidate on the NetApp systems with a minimum of interruption to business."

Both data centres in Istanbul are equipped identically, with two NetApp FAS3160As, a FAS3170, and a FAS2040. NetApp SnapMirror runs between the two data centers to protect against local disasters such as total power outage or fire. The software mirrors another copy to the FAS3170A in Izmir for extra security, because Istanbul is in an earthquake zone.

All of the systems are SAN connected. In total, there are 10 NetApp systems deployed, for a total capacity of 1.3 petabytes. Additionally, ING Bank Turkey uses NetApp deduplication, thin provisioning, and FlexClone to drive efficiency as well as Snap-Manager and SnapMirror to support continuous data availability.

The long-awaited virtual environment project was now possible. The migration of 500 virtual machines (VMs) to NetApp storage was fully transparent. However, what was noted the most was the integration with VMware vSphere and the resulting useful features. NetApp SnapManager for Virtual Infrastructure helps to simplify management and protection of this growing virtual environment, which is expected to increase to 800 VMs within a couple of months.

NetApp Snapshot technology has also proven to be a very effective tool for ING Bank. "We use snapshots intensively—10 snapshots of the Exchange and file sharing environments and 7 snapshots of the virtual environment, which translates into 500 x 7 copies, and all the backup of these took only 2 hours with NetApp. This is really impressive," says Gürdal. ING Bank Turkey controls the storage from a central console using NetApp Operations Manager. The tool delivers comprehensive monitoring and simplifies management with alerts, reports, performance tracking, and configuration tools, enabling the IT team to align the storage infrastructure for maximum efficiency and availability.

The new storage infrastructure is still in its early days and some fine-tuning still has to be carried out. However, for the IT team, the benefits of the new storage system became evident from an early stage.

Business benefits

One efficient platform does it all

"The migration process of all the environments was completed without any setbacks and at minimum interruption," recalls Gürdal. "When we migrated Exchange, including CEO level mailboxes, we were not sure what to expect, but our worries were unfounded. There was no data loss at all." Once this phase was completed and the structure was stable, the benefits of NetApp's storage efficiency, such as deduplication, FlexClone, thin provisioning, and Snapshot technology, began to kick in.

Virtual enhancements: By using NetApp and VMware software together, server provisioning is now easy, increasing operational efficiency and contributing to cost reduction. "With our plans to expand our virtual environment, this is a critical key to success. Also, by applying NetApp deduplication to our virtual environment, we have been able to save on storage capacity by almost 50%," states Hakan Gürdal.

Efficient disaster recovery and reduced backup times: The major processes associated with data protection have now been simplified and accelerated. With effective use of NetApp SnapManager, SnapMirror, and Snapshot software, the backup window has been significantly shortened. NetApp Snapshot has enabled the bank to create high frequency point-in-time copies and to reduce the time needed to restore files or e-mails. Beyond improving the system uptime, this tool has enhanced overall data availability for its users while saving the IT team time. SnapMirror is used to mirror the multiple Snapshot copies between the data centers.

By using NetApp and VMware software together, server provisioning is now easy, increasing operational efficiency and contributing to cost reduction

"In the Exchange environment alone, we have registered a 70% time saving. With our former traditional mechanism, it took us up to 12 hours to back up. Today, we do it in 3," says Gürdal.

With NetApp, ING Bank Turkey now has a storage infrastructure worthy of the firm's world-class reputation, one that delivers scalability, reliability, and high-level data protection and performance while enabling new processes to meet the rigorous demands of consumer banking. "With the unified approach across the entire product line, NetApp offers tremendous flexibility to configure and control our storage landscape in line with business needs. Thanks to NetApp, we are managing systems efficiently and easily," concludes Sivri.

AGILE BY DESIGN

Thomson Reuters unleashes new products with a shared IT infrastructure built on NetApp

Thomson Reuters is the world's leading source of intelligent information for businesses and professionals. The company combines industry expertise with innovative technology to deliver critical information to leading decision makers in the financial, legal, tax and accounting, scientific, healthcare, science, and media markets, powered by the world's most trusted news organization. Headquartered in New York, with major operations in London, England, and Eagan, Minnesota, Thomson Reuters employs some 55,000 people in more than 100 countries and reports revenues of $13.1 billion (2010).

The challenge

At Thomson Reuters, it's all about the "Knowledge Effect"—that is, putting the right information into the right hands to help clients produce amazing results. And just what does it take to deliver the Knowledge Effect? Rick King, chief technology officer for the company's Professional Division, says a lot of data and the ability to access it intelligently, quickly, and affordably: "We use a lot of storage—in fact, 16 petabytes of information is stored worldwide. But the data on that storage gains value only when our customers can find what they need, when they need it, and have confidence in the results. So the second half of the equation is sophisticated applications and search tools. To design and deliver those, we require a robust shared IT infrastructure, one that's flexible, scalable, and available to our product developers every day, hour, and minute of the year. NetApp is one of the companies that we've come to rely on for our storage needs and to enable the dynamic IT infrastructure that is vital to game-changing innovation."

Across the organisation today, Thomson Reuters uses NetApp storage to support a wide array of business systems and research platforms. The Thomson Reuters shared IT infrastructure and cloud are built on NetApp storage, the VMware vSphere virtualisation platform, and Cisco networking switches. This infrastructure provides a scalable and multi-use foundation for Oracle -based

content and metadata systems for Thomson Reuters' Westlaw legal research service that is used by 98% of the largest firms in the United States.

A major enhancement to the Westlaw service was one of the most recent—and "infrastructure-challenging"—projects to leverage the Thomson Reuters and NetApp partnership. King comments, "Our strategy for WestlawNext was to engineer a premium product that would make doing a legal search as easy as doing an Internet search. For that, we needed to store and search even more data, return results faster, and develop a human-centric interface."

The Solution Scaling out its existing shared IT infrastructure gave Thomson Reuters a jumpstart on the project, enabling rapid deployment of resources to engineering teams developing the WestlawNext back-end search dynamic, as well as front-end applications for enhanced accessibility and usability. "Our infrastructure strategy for WestlawNext required dramatically scaling shared storage," says Mick Atton, vice president and chief architect, Thomson Reuters Professional Division. "NetApp worked with us throughout the product development process to profile environments, identify capabilities and constraints, and recommend best practices to help us achieve product strategies."

Blair Linville, head, Shared Infrastructure and Operations, Thomson Reuters Professional Division, adds, "NetApp storage gives us tremendous flexibility for scaling and provisioning resources to pursue new ideas. If a concept doesn't pan out, we can just as easily redeploy capacity to support alternative efforts."

Rapid resource deployment, agile software development principles, and automated cloud-based testing contributed to an on-time and on-budget project launch. Cary Felbab, vice president, Technology, Thomson Reuters Professional Division, remarks, "At times it seemed like we were bending the laws of physics to meet project schedules and accelerate time to market. One key to our success was setting up automated scale testing using our private cloud infrastructure—the same one our business users access during the day—to simulate real clients hitting the system from the Internet. Every night we ran multi-thousand-user tests to measure performance. Leveraging cloud-based services let us do more testing faster and with fewer IT resources.

The Thomson Reuters team emphasizes that NetApp storage is an important element in a highly innovative technology stack. Inherent reliability and replication technologies help achieve availability targets and efficiencies from deduplication and compression enable expansion of content sets without requiring budget-crippling upfront capital investments. Shared storage improves both resource utilisation and economic viability.

The Thomson Reuters team emphasizes that NetApp storage is an important element in a highly innovative technology stack

Business benefits

Functionality indistinguishable from magic

The WestlawNext you-centric legal research system, launched in 2010 and named by the American Association of Law Libraries as its New Product of the Year for 2011, delivers breakthrough research capabilities. Highlights include:

- Enhanced search. WestlawNext enables non-literal searches on some 5 billion documents from heterogeneous sources. Advanced search algorithms deliver responses in an average time of 2.5 seconds—twice as fast as product developers' original target—with consistently on-point information, typically within the first 10 documents returned on a search request.
- Organisation tools—Clients can easily create and share folders, add notes and highlights to documents, and take advantage of other timesaving features to better organise legal research.
- Simplicity. Thomson Reuters experts

in artificial intelligence, data mining, machine learning, and natural language processing helped design the simple yet powerful interface. With its human-centric focus and single-box search field, WestlawNext offers a clean, uncluttered dashboard from which to work.

"The original Westlaw was hands down the dominant product in the market," maintains Felbab. "But good companies don't sit still, so with WestlawNext we set out to search 50 times more data and return results twice as fast. We've achieved those objectives and more, solving problems that were not possible to tackle just 5 years ago. And we're making it easier than ever for users to find information. But that simplicity belies the magic going on behind the scenes. For every customer search, the WestlawNext engine does 15 in the background. We're delivering in seconds or minutes what manual research processes could only produce in 2 or 3 hours."

Felbab points out that the Thomson Reuters infrastructure directly contributes to the magic: "NetApp NFS performance and on-the-fly scalability allow us, for example, to search on a significantly larger body of content, as well as to seamlessly handle the unpredictable workloads of a new launch. In just the first year, we scaled to bring on some 6,000 new users, exceeding our firstyear adoption forecast by 50%."

Dial-tone availability

"This system has worked flawlessly," states Felbab. "In many cases our clients are working to court deadlines, so they rely on 24/7 access to WestlawNext services. The replication processes NetApp helped us architect enable us to meet or exceed our 99.99% uptime target to deliver dial-tone availability."

Unlike Internet searches, a WestlawNext search must be absolutely accurate. If a system fails, Thomson Reuters can't temporarily substitute older or alternative content. So to ensure complete accuracy, a reliable, redundant infrastructure and NetApp shared storage were deployed that allow Thomson Reuters to quickly repoint a replacement server to content stores in the event of a server failure. NetApp NFS performance is also essential to being able to rapidly retrieve search indexes if a refresh of cached data is required.

NetApp Snapshot and the replication software that builds on it—including NetApp SnapRestore technology—are other tools that are heavily leveraged. These technologies allow Thomson Reuters to quickly—in seconds, not minutes—roll back to a consistent state when needed.

Innovation made affordable

Building WestlawNext on the shared IT infrastructure and cloud has helped Thomson Reuters achieve efficiencies essential to the economic viability of the new service. King highlights savings in:

- Administration. By standardizing on server, storage, and other infrastructure technologies, IT staff spends less time becoming experts on a myriad of platforms. Instead, they're able to focus on most effectively applying well-understood functionality to meet technical and business needs.
- Data center costs. "The virtualization process has cut 20–25% off our power consumption," states King. "Last year, negative power growth actually allowed us to return significant savings to the company. By eliminating older, inefficient technologies of existing resources via the shared IT infrastructure with NetApp storage, we avoided building a new two-megawatt data center at an estimated cost of $65 million. Most importantly, we're spending our money on expanding the business rather than growing a technology footprint. With a traditional infrastructure, we simply would not have been able to deliver on WestlawNext performance and capacity objectives at an affordable price point."

NetApp Snapshot and the replication software that builds on it—including NetApp SnapRestore technology—are other tools that are heavily leveraged

Will you marry me, WestlawNext?

To date, some 20,000 law firms, corporate law departments, and government law departments have upgraded to WestlawNext. Current clients include 36% of Am Law 100 law firms, the corporate law departments of 36% of Fortune 100 companies, and 97% of ABA-accredited law schools.

Customer response has been overwhelmingly positive. From praise for the "search anything" box to reports of dramatically simplified research and even a marriage proposal, legal researchers' eedback suggests unprecedented appreciation for the game-changing functionality.

Felbab comments, "For the first time ever, we're seeing emotional reactions to a product. NetApp helped us put in place a rock-solid infrastructure that allowed us to dream big, to tackle ambitious goals, and to deliver a service that is changing the way law firms do business."

Partnering to unleash new products King concludes by describing the critical role technology partners play in delivering differentiating innovation: "We use a lot of vendors, but consider only two providers to be strategic technology partners. NetApp is one. If we encounter a problem, it's fixed. If we tell them there is going to be a key launch like WestlawNext, they are engaged.

From the first day we started working with the company, the NetApp team has demonstrated both business curiosity and understanding. They've worked with us to optimize and tune performance, to quickly leverage new storage functionality. We've consistently found NetApp teams to be smart, creative, hard working, and entrepreneurial—those characteristics make them map well to our own project teams.

FLASH MEMORY: IT IS CACHE VS. CASH

Where speed of data access is a determining factor for enterprise applications, the emergence of flash memory is a development which cannot be ignored

It's just massively faster than anything that can be served from a mechanically spinning platter accessed by a read-write head. That's not all: lacking those mechanical aspects, it's also more reliable. These two factors make a potent cocktail for performance storage – but there is a drawback. Speed, reliability and cost are the holy triumvirate for most enterprise computing systems and form an equation which has to be balanced in the light of technical requirements and available budgets. Go up on speed and reliability, and cost inevitably spirals, too.

Flash memory is expensive. It is only for this reason that it isn't yet the de facto standard for the disk drives which are the building blocks of any storage technology. However, the cost is continually coming down, driven not only by the ramping up volumes in the enterprise storage space, but also through the enormous numbers of relatively low-volume consumer devices which are entering the market. It's a simple matter of economies of scale: a 16GB solid state USB key today costs as little as 120 to 130 AED's. Five or ten years ago, a 16MB memory stick cost roughly the same amount.

It is an industry expectation for 2011 that flash memory will see increased use in enterprise applications as companies seek solutions to ensure high performance despite ever-increasing volumes of data. However, it remains expensive, with a per-megabyte price which is far beyond that of 'traditional' disk drives.

Martyn Molnar, Regional Sales Director Middle East, NetApp

While it would be great to simply introduce the mass replacement of IDE drives with solid state devices, the cost is prohibitive. This demands an innovative approach to the introduction of flash which delivers the benefits of massive speed, without the drawback of massive cost. The answer, then, to reducing cash outlay while still gaining the benefits of flash memory lies in cache.

Caching provides a way to decouple storage performance from the number of disks in the underlying disk array to substantially improve cost and at the same time decrease the administrative burden for performance tuning. For example, some storage systems use NVRAM as a journal of incoming write requests, allowing the system to commit write requests to non-volatile memory and respond to writing hosts without delay.

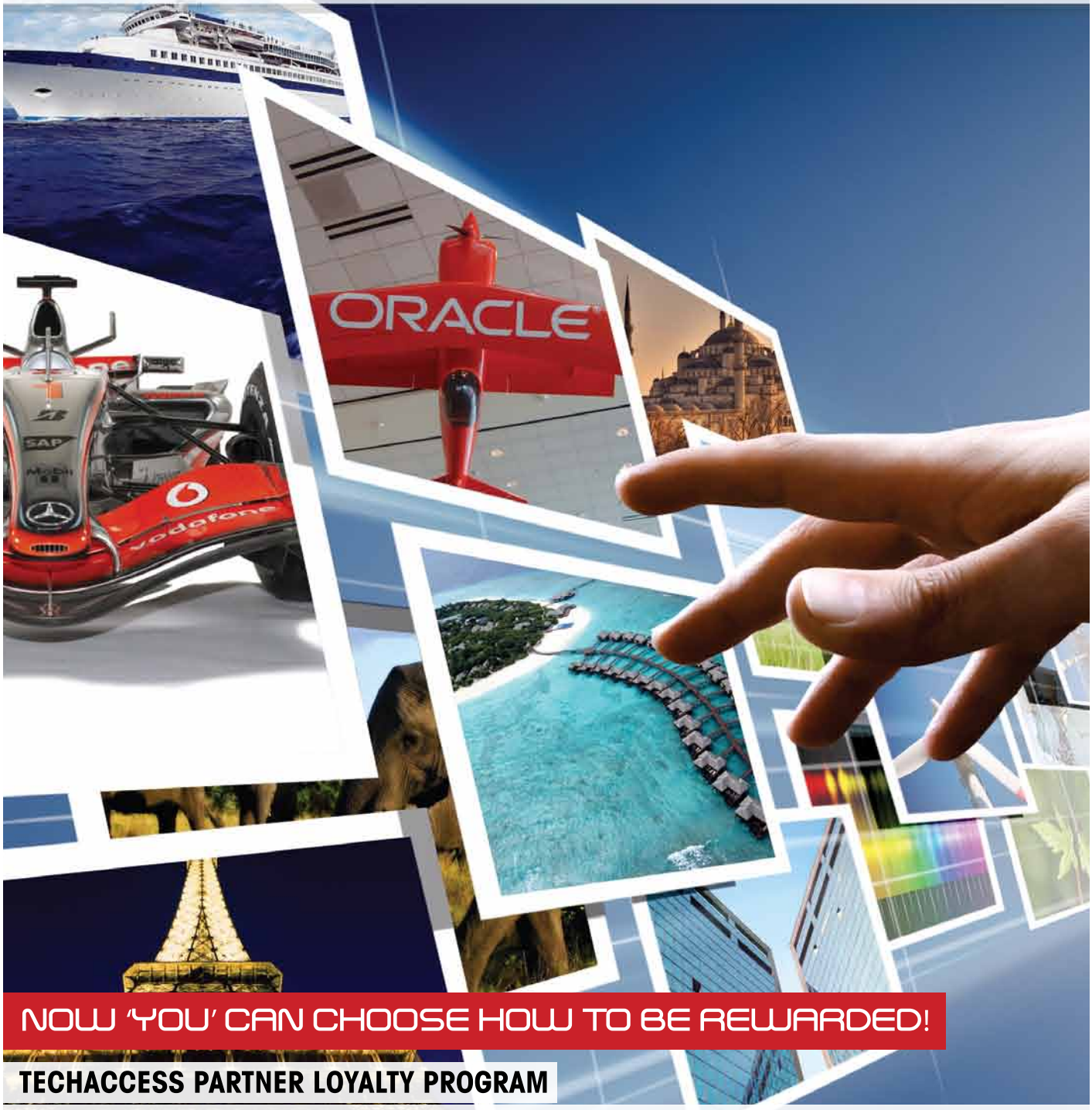
For read caching, a multilevel approach can be employed. A first-level read cache is provided by storage system memory. Special algorithms decide which data to retain in memory and which data to pre-fetch. At a second level, an appliance can be used (called Flash Cache) which allows for the creation of a large low-latency block pool. On top of that, a third-level can be provided, again with an appliance, which creates a separate caching tier in your storage infrastructure, scaling read performance beyond the boundaries of a single storage system.

While this represents a considerable simplification of how such a system works, the takeout is quite straightforward: through caching, it is possible to get optimal bang for the storage euro. It is, in effect, a form of 'tiering' within the storage subsystem itself which allows for optimal performance to be achieved without a massive outlay of cash. Using second-level read cache, for example, it is shown that it is possible to reducing the number of drives needed for a given level of performance by as much as 75% and even allowing the replacement of high-performance disks with more economical options.

Even as flash increasingly makes its way into the enterprise, it will do so driven by innovation and careful application which seeks to achieve the optimal balance between that equation of speed, reliability and cost. And right now, cache is the best way to optimise cash.

PARTNER LOYALTY PROGRAM

TECHACCESS
IN TOUCH WITH ENTERPRISE



NOW 'YOU' CAN CHOOSE HOW TO BE REWARDED!

TECHACCESS PARTNER LOYALTY PROGRAM

Welcome to the 'Best of Both Worlds'; The TechAccess Premier Partner Loyalty Program.

A unique program designed to offer a breadth of rewards on both business and leisure options to let **'You'** choose how to be rewarded.

To Learn more, please Log on to www.techaccess.com/loyalty

Offices G01-G03, Building No.1 Dubai Internet City, Dubai, UAE
TEL +971 4391-1820; FAX +971 4391-8875

***Terms and condition apply.**