WHITE PAPER

STORAGE MANAGEMENT IN A VIRTUALIZED WORLD

Taking advantage of the latest technologies and techniques to address growing data storage needs

Executive Summary

Data storage is growing explosively. External disk storage shipments amounted to 5,429 petabytes in the third quarter of 2011 alone — the sixth consecutive quarter of double-digit growth for disk storage, according to research firm IDC.

This growth, combined with new storage initiatives, such as unstructured data, virtual desktop interfaces and virtualization, is making the job of managing storage an increasingly important part of IT. Although server and application virtualization have eased the job of IT management, data storage (especially virtualized storage) has lagged behind, creating inefficiencies and higher costs in the data center.

To keep costs and rapid growth under control, storage must be managed effectively. Fortunately, storage managers have a range of tools and products available to rein in storage growth, improve efficiency and cut costs.

The key factor in the new storage landscape is integration. To get the maximum benefit from myriad storage solutions, IT managers need to tear down silos and apply a unified approach to all the storage in an organization while applying best practices across the whole spectrum.

TWEET THIS!

Table of Contents

A Smarter Approach

Integrating enterprise storage with modern storage management technologies provides a number of important advantages in the virtualized enterprise. The biggest advantage of modern, unified storage management is that it lowers the overall cost of storage — often drastically. By treating the organization and its disparate data storage solutions as a unified whole, a modern storage management approach can cut the need for installed disk space by 50 percent or more and drastically reduce the time and effort needed to manage storage.

The problem with a conventional storage management approach is that it forces the IT group to manage many different factions of data storage with various different tools. This lack of integration results in waste, inefficiency and expense.

Much of old-style storage management entails simple jobs, such as provisioning disk space and managing user accounts, repeated over and over again. While there are tools available for these jobs, they have to be applied repeatedly to silos of data storage scattered throughout the organization, wasting valuable time.

What's worse, the level of integration among conventional storage management tools varies widely. Some work together fairly well. But in other cases, the IT storage manager must use several different tools to do even simple jobs.

Modern storage management is built around a unified, integrated toolset with a common interface that allows the manager to perform all necessary tasks on all storage in the organization from a single, central location.

One of the most remarkable results of a modern unified storage management system is that it drastically reduces the total amount of storage needed by an organization. By treating all of the organization's storage resources as a unified whole, less physical storage is needed to support applications. Storage management systems accomplish this through techniques such as thin provisioning (which cuts down the amount of storage assigned to each disk while allowing for future growth) and deduplication (which eliminates redundancies in stored data).

Studies have shown that allocated but unused storage accounts for an average of 40 percent or more of the total storage on a typical disk. Eliminating redundancy can cut the amount of stored data by 90 percent or more. Therefore, modern storage management techniques result in enormous savings in disk space.

Cloud Storage

Storage in the cloud provides an important new tool in the storage administrator's toolkit. Cloud storage services provide scalable, reliable, instantly available storage with no capital investment.

Because the enterprise pays only for the storage it actually uses, it can more easily control costs. In addition, maintenance and tasks such as data backup typically become jobs for the cloud provider, freeing the organization's storage administration resources for more business-critical tasks.

Cloud storage is especially suited to backup and disaster recovery because copies of the data are stored offsite in the cloud. However, recovery time — as well as other characteristics of cloud storage — is dependent on the bandwidth connecting the stored data to the organization.

Bandwidth limits how fast data can be restored. If a large amount of data needs to be recovered (for instance, in the event of a bare-metal restore), the recovery time may be unacceptably slow. Many cloud storage organizations offer alternative methods of recovery for large amounts of data, such as express-mailing disks or tapes containing the data.

Storage management also makes it easier to perform capacity management and other long-term management tasks. Because storage resources are unified, it is easier for the IT team to keep track of growing storage needs and decide what kind of storage to add and when to add it.

Rather than trying to predict storage growth on multiple servers, the storage manager can deal centrally with the overall quantity of storage and allocate it to applications or servers as needed. These storage management systems also make it easy to reassign storage capacity on the fly as needed and to add storage to the overall system.

Effective Storage Management

The most important characteristic of modern storage management is integration. Unified storage systems, such as EMC's VNX and NetApp's FAS family of solutions, are much more than software applications used to manage and report on the state of an organization's disks.

Rather, they are complete, holistic systems that aim to handle all elements of storage, from the basic storage architecture to storage optimization to details of the management interface. In combining all the elements in this fashion, these storage systems leverage different components to maximize efficiency, ease of use and cost savings. Other characteristics of an updated approach to storage follow.

Unified Storage

Unified storage management systems, at their most basic level, are considered unified because they can handle both file- and block-level data on the same device, resulting in myriad savings and efficiencies.

"Unified" in this case also usually means a management system that can handle multiple common storage protocols, such as Network File System (NFS), Fibre Channel and iSCSI. Because of this level of unified support, the storage device becomes a one-size-fits-all system capable of handling most of the storage protocols in use today.

Basically, there are two types of storage in the modern data center. There is file-oriented storage, in which the basic unit is the individual file; and there is block-oriented storage, which is based on blocks of storage without reference to files. File-oriented storage generally uses network-attached storage (NAS) technology, while block-oriented storage is most commonly associated with storage area networks (SANs).

In the absence of a unified approach, the combination of file- and block-oriented storage in the same system introduces a fundamental separation. In general, file-oriented systems can't handle block-oriented data, and vice versa. This means that many organizations have to maintain two sets of storage systems — or more, if they have equipment that uses different protocols.

Unified storage eliminates the need for separate systems by translating file-oriented storage into block-oriented storage. While this can be done by a CPU, it is computationally intensive. Therefore, the job is usually handled by a special processor in the storage controller, or by a gateway appliance.

The overall result of a unified storage system's suite of features is that it saves both time and budget for the organization. Like virtualized servers, the virtualization of unified storage hides as much of the complexity as possible from users and storage administrators. The result is faster, easier management, coupled with the cost savings of consolidating both hardware and software.

Vendors of unified storage systems boast that even complex jobs, such as provisioning an e-mail system under Microsoft Exchange, can be done in a few minutes by minimally trained personnel. Even a complete setup of a storage system from scratch using multiple protocols can be done in a few hours.

The Case for Unified Storage

There are several important advantages to unified storage including the following:

- **Simplicity of management:** The most obvious advantage of unified storage is that it simplifies the handling of the organization's data. By hiding the complexity of different protocols and different kinds of data behind a layer of virtualization, unified storage takes a load off the storage administrator.

  In fact, in many cases, the storage administrator doesn't have to be involved at all. Through wizard-based interfaces and other tools, just about anyone can configure, deploy and manage enterprise storage through the unified approach.

- **Fast and easy scaling:** Because of the level of virtualization built into unified storage systems, they can quickly and easily scale up or down as the storage demands of the organization change.

- **Reliability and high availability:** It is a truism that virtualization demands high reliability and high availability. Unified storage systems provide these benefits through features such as dual controllers and redundant pathways, mirroring cache or nonvolatile RAM (NVRAM) to protect against failure.

- **Serviceability and support:** Virtualization enhances the serviceability of unified storage systems by making it easy to switch out parts when they develop problems. Data can be moved around in storage to improve performance or work around failures. The level of integration and the unified user interface make it easy to support changes (even extensive changes) in the storage system as requirements change.

Data Tiering

Modern storage comes in different degrees of speed, capacity, price and technology. Generally, the faster a device can store and retrieve data, the more expensive it is.

Tiered storage capitalizes on this dynamic to offer high-performance storage at minimal cost by mixing different kinds of storage into tiers. By carefully choosing the appropriate kind of storage for each type of data, IT departments can optimize performance while holding down the overall cost to store all that data.

Many kinds of data lose value over time. As data ages, most tiered-storage systems can automatically migrate that data from one tier to the next according to rules set out by the organization. In other words, older, less valuable data moves to slower, less expensive storage because it's less likely to be needed; while newer, more important data remains on faster storage devices so it may be more easily accessed.

The top tier (Tier 0) in a fully developed tiered-storage system is usually a solid-state disk. SSDs are extremely fast devices with relatively low capacity (typically 256 gigabytes or less) per device. Comparatively speaking, SSDs are expensive, roughly ten to twenty times the price per gigabyte of conventional disk storage. However, when used appropriately, SSDs can compensate for their cost by increasing the performance of a system.

SSDs are typically used in applications, or parts of applications, for which high performance in reading and writing is paramount, such as indexes for databases, video editing or online applications. In many cases, despite their cost, SSDs offer an economic alternative to other methods of increasing performance, such as adding servers.

Of course, not every tiered-storage system uses SSDs. The idea behind tiered storage is to build a customized stack to give the necessary performance at minimum cost. In many cases, a tiered-storage system will start with traditional hard-disk drives.

There are several varieties of hard disks available, ranging from fast iSCSI devices through larger, more economical SATA disks. Here again, there is usually a direct relationship between performance and price. By using cheaper disks for less critical storage, administrators can achieve significant savings without compromising needed performance.

Finally, there is tape, which provides the cheapest medium for long-term storage of archival information. Most medium and large systems use tape as the final tier in their storage architecture, often using deduplication to minimize the amount of storage needed.

The key to getting the most out of tiered storage is to classify the organization's data and select the appropriate tier and migration path for each class.

Data Deduplication

One way to rein in storage growth is to store the same amount of data in less space. Deduplication accomplishes this by reducing the amount of redundancy in what the organization is storing.

By its nature, data includes a lot of duplication. Words are repeated, formatting information is repeated, and numbers are repeated. Generally, the more broadly data is viewed, the more duplication there exists.

In today's storage management systems, data deduplication technology studies the data stream, locates duplicates and removes them, and replaces the duplicate information with much shorter tokens. Not to worry: The full data stream can be reproduced as needed.

The Effect of DC Virtualization on Storage Management

The nature of data center virtualization and its relationship to enterprise storage make it vital that potential problems be found and addressed early. A problem with one server in a virtualized environment can affect a dozen or more applications and produce a range of counter-intuitive symptoms. Though some of the data center resources may not be in their area of expertise, it can be important for storage administrators to track use statistics from all virtualized machines and spot trends before they become problems.

With virtualization, it is easy to forget that there is real hardware underneath abstract layers of IT. If resource usage and other parameters aren't carefully monitored, demand on a virtualized server can grow until applications start interfering with each other, and memory hot spots and other problems degrade performance.

Virtualized systems can be more complicated to troubleshoot when performance issues arise. For one thing, there are likely to be more resources (servers or applications) in a virtualized environment than physical resources in a nonvirtualized environment. Another issue is that the relationships between the components of the system are likely to be much more complex.

In the end, storage planning and provisioning assume greater importance in the virtualized environment. Because more resources depend on the same block of storage, it is vital that enough storage be made available in every pool and at all times.

Techniques such as thin provisioning make storage work harder and leave less room for error. Storage administrators must be on top of their organization's needs and must plan ahead to add more storage before those needs becomes critical.

This isn't a new technique. Versions of deduplication have been used for decades. For example, most tape drives can compress data by roughly 50 percent and have done so for years. But the combination of new deduplication technology and more powerful processors means today's storage solutions can deal with larger chunks data faster, making deduplication effective for primary (disk-based) storage, as well as backup systems.

The bigger the chunk of data, the better the deduplication. This is because the more data a storage system can process at one time, the more redundancy it can identify and remove.

Although the exact amount of redundant information a system can remove will vary significantly depending on the data itself and other factors, in general, data deduplication can produce significant storage savings. Modern deduplication techniques can compress a data stream by as much as 90 percent or more. That means the IT department must store just 10 percent of the actual data.

Deduplication can be either hardware-based, with a special appliance that handles the process, or it can be done by software. Choosing one method over the other depends on the characteristics of the overall system.

Software deduplication is usually much cheaper, but it puts a considerable load on the server that must handle the deduplication. Hardware deduplication is more expensive, but usually much faster. Storage systems that don't need much deduplication are more likely to use a software-based approach.

Another consideration in choosing a deduplication technology is where in the system the data deduplication and reconstruction will be performed. The easiest and least expensive option is to deduplicate data at the storage device level. For this reason, many storage arrays come with built-in deduplication.

The drawback to this approach is that it doesn't reduce the amount of traffic that must travel over the network, because the data isn't deduplicated until it reaches its destination. This can reduce the overall efficiency of the system and cut into potential savings.

Deduplication in a modern storage system can produce a cascade of benefits, the most obvious being the potentially significant reduction in storage space needed to hold deduplicated data. This reduces the cost of storage and slows the growth in storage demands. However, there are a number of other benefits as well, such as the following.

Greater utilization of fast storage: For one thing, storing less data means information can sit in a higher storage tier for a longer period of time before it passes to a lower tier. When data is held longer in a higher tier, storage administrators can set better recovery time objectives (RTOs). In the event of a problem, admins can recover data more quickly because it's held in a higher storage tier.

Less tape: A related benefit is the possible reduction of tape storage. Because data can be held on disk for longer periods of time, organizations need less tape for accessible data. In the ideal situation, tape becomes purely an archival medium, used only to store information that must be kept for a long time.

In this ideal state, backup data (as opposed to archived data) remains on disk, such as inexpensive SATA disks, and never reaches tape at all. Most organizations won't completely eliminate tape for backup, but deduplication allows them to reduce its use.

Quicker retrievals: However IT chooses to deduplicate and store its data, retrieval times are usually much shorter. True, the data has to be restored, which takes a good deal of processing power. But the amount of data that has to be recovered and moved over the network is usually much less. Given a good deduplication system that is properly sized for the storage system, the lower volume of data more than balances out the extra time needed to reconstruct the deduplicated data.

Less network traffic: As mentioned earlier, a major benefit of data deduplication (if deduplication is done at the data source) is that it reduces the amount of data traveling over the LAN. This means the demand for network bandwidth grows more slowly over time, and the organization doesn't need to spend as much on networking equipment to maintain response time.

Of course, these benefits come at a price. In effect, data deduplication means trading storage space for processing power. Finding and eliminating redundancy on the scale of a modern deduplication system takes a lot of processing. In addition, the deduplication hardware or software must be sized carefully in order to provide enough performance so that the storage and retrieval processes don't become bogged down.

Data Storage Virtualization

Storage virtualization does for data storage what server virtualization does for other resources. By treating the storage system as a whole, it enables the overall system to operate more efficiently and with fewer wasted resources. To accomplish this, it relies on sophisticated software to squeeze the best possible performance out of the available resources.

Storage virtualization starts by taking a holistic view of system storage. Rather than looking at storage in terms of disks or applications, it treats all the storage in a system as a unified storage pool.

One example of storage virtualization is thin provisioning, which virtualizes empty disk space so that a system can assign more disk space to applications and virtual machines than it actually has. Thin provisioning is based on a simple, well-known fact: There is always more storage than what's in use. In fact, if the amount of storage in use gets

anywhere near the amount of storage available, alarms start going off, and administrators scramble to add even more storage space.

In a typical "thickly provisioned" storage system, excess storage space is assigned to each application, so there is a lot of space that never gets used. On an average server, this amounts to a lot of unused space. Typically, a well-managed server will only use 40 percent of its available storage.

Thin provisioning puts that other 60 percent to use by virtualizing the unused space among applications as needed. The applications think they have the extra storage they need, but it is actually shared among all the applications on the system.

The total allocated space is larger (often much larger) than the amount of available storage, but because most of that storage is never used, it doesn't matter. As the applications and their data grow, a storage management system keeps track of how much space the applications actually use and alerts administrators when more physical space is needed.

In a SAN, storage virtualization is often accomplished using a switch or appliance in the network. The device performs the mapping of physical-to-virtual locations and translates I/O requests from the system to the physical location. The system "sees" a unified pool of storage, while the switch or virtualization appliance makes the connections with the actual physical space.

Virtualization also makes it possible to relocate objects in physical storage by changing the metadata that translates the virtual location to the physical. This allows the IT team to rearrange the storage space on the fly, moving data to a different disk or adding additional storage. Servers and applications don't even realize that it's happening.

With storage virtualization, because the system is dealing with a single pool of storage, many other management functions are also easier to accomplish. For example, data backup is simplified because the IT group deals only with a single contiguous unit of storage.

Similarly, this team is able to perform replication invisibly to the server and take snapshots of data more easily. Thanks to its numerous advantages, data storage virtualization plays a central role in modern storage management.

Storage Management Software

Organizations will want to adopt a software solution that brings these helpful components all together and offers advanced functionality to meet today's data requirements.

Data Discovery and Storage Management

Discovery is a relatively new task for storage management software (SMS). Although there are many specialized products to help organizations recover records and documents that are required for legal and regulatory reasons, SMS has become an important launchpad for discovery processes.

Discovery has different requirements than mere data archiving. Archiving is aimed at preserving data in large, relatively unstructured chunks. Teasing out specific records from an archive is usually secondary.

Discovery has to be far more granular. The aim of discovery is to retrieve specific records on specific topics with a minimum amount of wasted effort. The data that needs to be retrieved, often in response to a legal request, is usually recorded and stored months or years before the discovery request.

And often, the data must be collected from multiple sources within a short, prescribed time period. Automating the discovery process is key to helping organizations meet their obligations. Today, more storage management solutions include discovery features.

Storage management software (SMS) isn't new, but the modern storage landscape has changed the job it needs to do.

Today, SMS must handle new technology challenges, such as virtualization, as well as new data tasks, such as information discovery. Moreover, even event logging and analysis (classic jobs for SMS) have become more difficult as organizations require greater detail and more sophisticated analysis.

Monitoring what is happening in a storage system has always been a fundamental job of SMS. However, changes in the environment have morphed monitoring requirements almost beyond recognition.

For example, server virtualization adds major new requirements to managing storage. Now, it is no longer enough to monitor only storage; the virtualized environment must be monitored as well. Each virtual machine and virtual server has to be tracked and its use patterns analyzed.

But most storage administrators don't have time to hover over their storage pools. They need sophisticated rules-based alarm systems to inform them when their storage

system moves outside specified parameters. An important part of the solution to these challenges is comprehensive, transparent monitoring with associated analytical tools and alarms to make it obvious what's going on in the increasingly virtualized data center.

Resource planning and provisioning are more important in today's storage environment. Administrators need tools to help them stay on top of storage trends so they can add capacity in a timely manner.

A fundamental trait of modern, high-efficiency storage environments is that they do more with relatively less storage. Techniques such as thin provisioning, storage virtualization and deduplication squeeze excess storage out of the system. This is, after all, the definition of efficient storage.

But it also means there is less room for error in providing storage capacity. Modern storage management software helps track utilization and alerts administrators to developing problems in time to fix them before they become apparent to users.

Further complicating matters is the sheer volume of data that SMS provides. Vital information can be easily overlooked when it's buried in an avalanche of log entries. Storage management systems help the administrator by teasing out important trends while keeping the overall flow of information at a reasonable level.

Finally, when evaluating SMS, organizations must consider compatibility. The ideal storage management system will work with all equipment and software from all major vendors. It will support any kind of storage on any major operating system, and do so transparently, allowing the administrator to manage all of the storage functions from a single, unified console.

The good news? Storage management compatibility is a lot better than it used to be. The emergence of storage management standards, standardized data formats for recording and reporting, and more standardized data sets are making it possible to manage more kinds of storage more easily.

The bad news is that compatibility still isn't complete. There remain major differences among SMS packages, in terms of what and how well they actually manage multiple types of storage devices in multiple environments.

Today's storage management solutions are much more sophisticated than the collection of scripts that administrators used to manage storage just a few years ago. However, SMS isn't completely interchangeable. In selecting a storage management solution, administrators must carefully consider both their needs and the details of their storage system and select the SMS that will do the best job for them.

Getting Started

Storage management in a virtualized environment offers a range of options and benefits. An organization's first task is to figure out what it needs and what kind of storage systems will give it the most efficiency with the least expenditure, both in capital and operating costs.

Moving to a new storage management system is an exercise in optimization rather than an upgrade. The organization has to decide on the right mix of technologies, techniques and applications to achieve the best possible performance at a reasonable cost. Simply adding equipment and software won't help. Because of the specialized knowledge required, many organizations hire a consultant to help them define and implement their goals.

Every organization is different; most will require a custom mix of storage resources and strategies. Asking several questions up front will help determine the right mix.

For example, what is the appropriate response time for the organization's applications? Does it need to support a great deal of transaction processing? Does it support a heavily trafficked website? Which applications are considered mission critical, and how fast does the IT team have to get them back up and running in the event of a problem?

Designing a storage management architecture from the ground up requires careful thought and planning. The organization has to balance cost versus return on investment (ROI) and storage performance. Typically, this requires getting input from all stakeholders — users, top management and IT personnel — and reducing the conflicting wants, needs and nice-to-haves into a firm plan that works for everyone. The plan then has to be reduced to a requirements list for the hardware and software needed to implement it.

Once the requirements are set, the next step is to select potential vendors. Often, this is a two-step process, starting with an extended list of possibilities and then narrowing that down to a short list of candidates whose offerings may meet the organization's needs.

First, an organization will want to draft a request for information (RFI) outlining the project and its requirements, and then send it to the vendors on the extended candidate list. When they reply, evaluate the responses and create the

short list of vendors that will be invited to bid via a request for proposal (RFP). The RFP is more detailed than the RFI and will include more information on the operations, specific needs, constraints, assumptions and terms and conditions.

Once the organization has selected a vendor, the next step is to negotiate the contract. Be sure to include all important, relevant terms and leave as little as possible to unwritten understandings. Keep in mind that in choosing a storage management vendor, the organization is seeking a long-term partner to help with its still-evolving storage needs. Contract negotiations shouldn't be adversarial.

The last step is to procure, install and deploy the solution. Both the organization's consultant and vendor should be helpful at this stage. Typically, the devil is in the details, so the IT group needs to pay close attention and make sure everything is working properly before signing off on the project.

Storage management is an ongoing process that doesn't end when the system is installed. The IT department must constantly evaluate the system's performance throughout its lifecycle and lean on its vendors for continuing support.

This is a more-involved process than some IT teams may be used to. However, the savings gained from a properly designed storage management system make it well worth the effort.