NETAPP WHITEPAPER

DATA CENTER STORAGE TRENDS, CHALLENGES, AND SOLUTIONS

Mike McNamara, NetApp
January 25, 2007 | WP-7011-0107

TABLE OF CONTENTS

OVERVIEW

The data center has become the heart and soul of a company. It provides far more than just the technology aspect of a business. But the average data center has grown extremely complex. New technologies, fast growth, acquisitions, the online data explosion, and increased security concerns have driven complexity up and utilization rates down. In the recent years there have been an increase number of threats like terrorist attacks, state-wide power outages, and the hurricane and tornado season which have only exacerbated the challenges facing data centers.

While priorities may change over time, one remains constant: the need to do more with less. This paper addresses eight key trends and challenges that impact data centers now and will impact them into the future: connectivity, tiered storage management, thin provisioning, storage system resiliency, application availability, unified storage, security, and integrated data management.

Organizations that have the most revenue and are most heavily dependent on online systems, have the highest potential loss of revenue (up to $2M per hour, or $500 per employee) from application and network outages. The storage trends, challenges, and solutions outlined in this paper will help data center managers to plan and adapt their storage resources and policies.

CONNECTIVITY

Based on industry data from Gartner Group, growth for Fibre Channel (FC) slows, but it will still remain the dominant storage network technology for at least the next five years, even though Internet Small Computer System Interface (iSCSI) over Ethernet will restrain its growth. Continued backward compatibility and the increase in speed (4Gb, 8Gb) continue to characterize FC technology for the data center. Ethernet-based iSCSI represents the biggest challenge to FC dominance and 10GbE helps to narrow the gap. InfiniBand (IBA) offers a low-cost switching technology, but it is not a mature storage networking technology and only a few vendors offer native IBA storage. Serial-attached SCSI (SAS) offers a low-cost alternative to FC for low-to-midrange direct-attached storage, server/storage (DAS) cluster environments, and, potentially in the future, small SANs, but will not impact FC in the data center.

TIERED STORAGE MANAGEMENT

Organizations need ways to enhance storage capacity utilization and optimize storage costs based on the value of data to the organization. By implementing tiers of storage within data centers, organizations can overcome some key data management challenges such as increased storage requirements placed on primary (expensive) storage while secondary storage remains underutilized, inability to effectively place data on different types of storage based on its relative business value, and the high cost of backup in the absence of appropriate classification of data, resulting in excess data protection for data that is not mission critical.

Although certain products attempt to provide tiers of storage, including traditional HSM solutions, their narrow scope in terms of vendor platforms supported or the disruption they cause to users limits their benefit. Some key deterrents to implementing tiers of storage include:

- High user downtime resulting from both tedious data migration procedures and long restore windows for archived data
- Significant administrative effort in restoring user access to migrated data
- Lack of solutions that can migrate data across heterogeneous storage platforms
- Lack of centralized management of distributed data to reduce administrative complexity

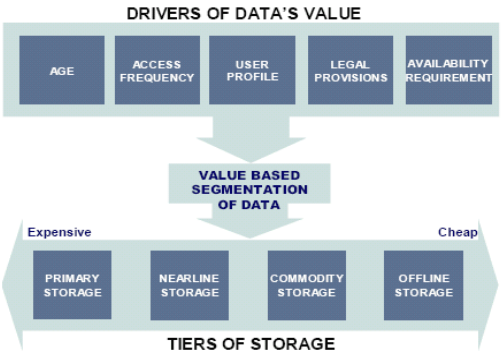

Figure 1) Tiers of storage

- Inability to organize data intelligently and present it to users logically

A data management solution needs to overcome these limitations to deliver a simple yet powerful solution to implement tiers of storage across multiple storage systems, independent of their location or performance characteristics (see figure 1). A tiered storage solution for the data center should provide:

- High level of automation
- Centralized management of heterogeneous storage
- Policy-based migration of data across storage tiers
- Nondisruptive data movement
- Business view of data independent of physical storage tiers

THIN PROVISIONING

Data center IT managers and storage administrators routinely report using just 30% to 40% of their total disk capacity. Whether the problem stems from a direct-attached storage infrastructure with its inherent islands of stranded capacity or inefficient and inflexible data management software, utilization won't improve unless the storage architecture is improved. The good news for data center managers is that by enabling maximum storage utilization, the right storage architecture can dramatically improve capacity/cost ratios to satisfy IT management and users and the corporate financial managers.

To evaluate storage architectures, a myriad of factors must be considered that impact utilization: operating system efficiency, provisioning techniques, volume management, data protection, and backup facilities. How each service is implemented directly impacts the ability to achieve optimal storage efficiency while still delivering on application and business objectives.

The first step in the implementation of any storage system is the allocation of space to servers and applications. Most storage systems require the storage administrator to pre-allocate specific physical disk space to applications and once allocated, that space is no longer available (free) to be used by other applications that may need it. The problem is that in the early stages of deployment, the storage administrator seldom knows the exact requirements of users and applications, and most administrators have no way to assign storage space to applications without "locking in" specific disk drives to volumes and LUNs.

For example (see figure 2), if a 500GB volume is allocated to an application with only 100GB of actual data, the other 400GB has no data stored on it. That unused capacity still belongs to that application and no other application can use it. As a result, the unused capacity of that 500GB is wasted storage and money, and even though all of the storage capacity is eventually used, it could take many months or years to do so. Thin provisioning eliminates this waste. Using the same example, the system administrator provisions 500GB to the application with only 100GB of actual data. With thin provisioning, the unused 400GB is still available for other applications. This approach allows the application to grow transparently and at the same time ensure that capacity is not wasted.

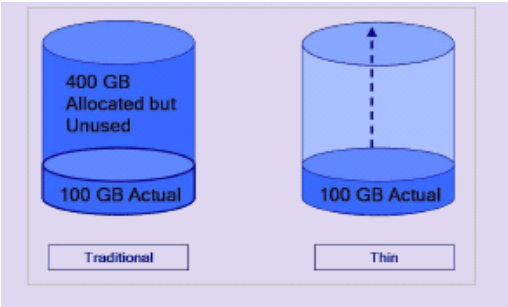

Figure 2) Thin provisioning

Thin provisioning is essentially just-in-time storage. The application thinks it has 500GB of storage, but the storage system only gives it the capacity as it needs it. The rest of it stays in the pool and system administrators can set thresholds to be alerted when to add more disks into the storage pool. Thin provisioning benefits data centers by improving storage utilization up to 65% to 85% and reducing storage costs and complexity.

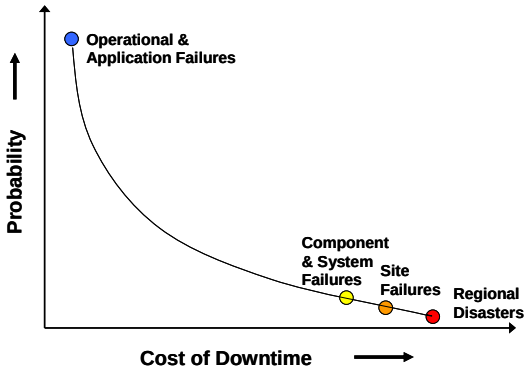

Figure 3) Causes of downtime (Source: Gartner Group)

STORAGE SYSTEM RESILIENCY

Enterprise data centers must provide high levels of application availability and consistent data integrity to support business-critical applications around the world. Data center managers continually wrestle with the challenges of avoiding unplanned downtime to ensure application data is available and of avoiding data corruption to ensure that application data is correct and up to date. Compromises to either data availability or integrity can have disastrous consequences for a company's bottom line and reputation.

While regional disasters and site failures get the most attention by virtue of causing the most pain, the most common causes of unplanned outages are local errors due to operational failures followed by component or system faults (see figure 3). To achieve 99.999% application availability requires a highly reliable storage environment that prevents downtime and data corruption whatever the cause.

Two industry trends—data center storage consolidation and the widespread adoption of larger-capacity storage—make high availability a more urgent priority for storage and IT managers. With consolidation, even higher availability is required as more data and applications are at risk. At the same time, increased adoption of SATA storage with larger disk capacities increases the risk and probability of failures. Productivity growth, increased global competition, and stringent regulatory requirements make additional demands. As a result, data centers require a comprehensive portfolio of storage resiliency technologies that help support very high levels of application availability. To protect against business interruption, storage resiliency should be built into every aspect of the storage solution.

True storage resiliency has two aspects and storage architecture needs to provide both: (1) preventing errors and system failures from happening by means of early detection and self-healing processes and (2) recovering quickly and unobtrusively from errors and system failures when they do happen. A vendor's solution should include tools to predict and fix disk drive faults before they happen, protect against all forms of double disk failure cost-effectively and with minimal impact to performance, maintain data availability in spite of triple disk, enclosure, and storage loop failures, and support synchronous and asynchronous replication, clustered failover (local and remote), and fully redundant, fault-tolerant systems.

APPLICATION AVAILABILITY

Data centers are under relentless pressure to improve application availability, striving for 100% availability for mission-critical applications. Due to the increasing cost of downtime, organizations need to focus on the various causes of outages and adopt a systematic approach to reducing the risks of downtime. Not only do typical infrastructure issues need to be addressed, but the people and process issues also have to be addressed with a plan in place to quickly recover from unforeseen disasters.

Disasters such as 9/11, hurricane Katrina, and the tsunami in Indonesia have caused many organizations to reconsider their approach to risk management and disaster planning. Organizations recognize that disasters and disruptions will occur, but focus has changed from disaster avoidance to disaster recovery.

There are two categories of application downtime: planned and unplanned. Failures of one type or another are, for the most part, unavoidable and can lead to unplanned downtime. Improving application availability not only depends on preventing unscheduled downtime and recovering seamlessly from unexpected hardware and software failures, but it also depends on the ability of administrators and operators to perform their daily tasks without reducing the availability of system resources.

Improved application availability and speedy disaster recovery require a storage architecture that protects against all planned and unplanned downtime and allows quick recovery when downtime does occur. A storage solution must provide a comprehensive offering that addresses all causes of application downtime—preventing operator errors, recovering from operator and application errors, minimizing planned downtime, maximizing system uptime, and recovering from a disaster. All storage vendors deliver availability, but not all focus on delivering solutions for the most frequent causes of downtime: application and operational failure.

Site and natural disasters are less likely than operator error, but they can have a much greater impact. Data centers require a flexible and cost-effective disaster recovery (DR) solution, making it affordable to cover all application tiers under a single DR plan and a solution that puts the DR site to active business use. Application and database administrators need application-integrated solutions that perform frequent and nondisruptive backups in a matter of seconds to ensure RTOs (recovery time objectives) and RPOs (recovery point objectives) are met.

UNIFIED STORAGE

To fully realize the consolidation and management benefits of networked storage, data centers need to deploy a solution with a single set of management tools that meets both the SAN and NAS requirements inherent in most data centers. A unified pool of storage would have higher storage utilization, a single data recovery solution, a single data management model, and greater leverage of IT staff and skills. A unified storage platform would use just one set of software and processes across all tiers of storage. These benefits result in a better return on investment (ROI) and reduced total cost of ownership (TCO).

Unified storage systems should abstract and virtualize the specifics of SAN and NAS into a common form that can be allocated and managed using the same set of tools and skills. All of the internal workings required to deal with the specifics of each networked storage approach (FC SAN, NAS, and IP SAN) need to be transparent to the user (see figure 4).

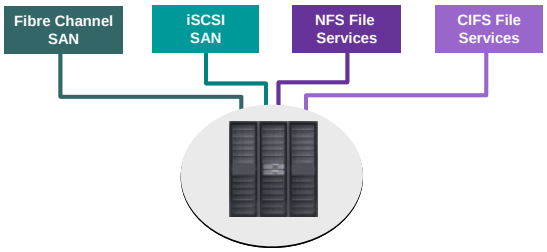

Figure 4) Unified storage architecture

Greater storage consolidation, investment protection, and reduced management have been driving factors in the adoption of unified storage. For data centers in which three (or more) different management paradigms are at play—for example, NAS, midrange IP SAN storage, and high-end FC SAN storage—the ability to combine all approaches under a single unified storage architecture is becoming extremely attractive. Unified storage is a natural bridge because of its ability to support and manage multiple types of requirements within a common pool of storage. Unifying storage and data management software and processes significantly reduces the complexity of data ownership, lowers the risks associated with user errors, and decreases the need for expensive training and integration services.

SECURITY

The advantages of networked data storage technologies such as network-attached storage (NAS) and storage area networks (SAN) are well established, but storing an organization's data on a network creates significant security risks. Data in networked storage environments is significantly more vulnerable to unauthorized access, theft, or misuse than data stored in more traditional, direct-attached storage. Aggregated storage is not designed to compartmentalize the data it contains, and data from different departments or divisions becomes co-mingled. Data replication, backup, off-site mirroring, and other disaster recovery techniques increase the risk of unauthorized access from people both inside and outside the enterprise. Partner access through firewalls and other legitimate business needs also create undesirable security risks.

With storage networks, a single security breach can threaten the data assets of an entire organization. Technologies such as firewalls, Intrusion Detection Systems (IDS), and Virtual Private Networks (VPN) seek to secure data assets by protecting the perimeter of the network. While important in their own right, these targeted approaches do not adequately secure storage. Consequently, they leave data at the core dangerously open to both internal and external attacks. Once these barriers are breached—via stolen passwords, uncaught viruses, or simple misconfiguration—data assets are fully exposed.

Businesses that don't encrypt sensitive data will spend a lot of money on corrective measures and reparations because of failure to comply with regulatory or contractual data protection requirements. A unified platform for securing stored data across the enterprise, with support for NAS, DAS, SAN, tape, and iSCSI environments, is the optimal solution for data centers to protect data assets. Providing wire-speed encryption and protecting data at rest with secure access controls, authentication, and secure logging simplify the security model for networked storage. Security appliances must also be deployed transparently within the data center without changes to applications, servers, desktops, and storage.

INTEGRATED DATA MANAGEMENT

In traditional data center IT organizations, application, database, system, and storage administrators each focus narrowly on only a part of the data/storage management problem. Each has different and distinct areas of responsibility and accountability. As a result, end-to-end data management depends upon manual communication between the data administrator and storage administrator and manual mapping of data to storage. This is a disruptive and potentially error-prone process that results in critical errors and lost productivity.

Traditional approaches have left a gap between the management of data and the management of storage. This has resulted in inefficient operations, with considerable duplication of effort and with frequent interruptions to the activities of highly interdependent administrative groups.

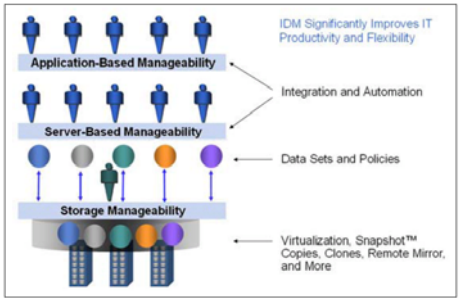

Figure 5) Integrated data management

An integrated data management approach would simplify the data management that encompasses both the management of storage devices and the data that resides on those devices (see figure 5). Using this approach, storage administrators would operate more efficiently and with minimal interruptions by automating routine storage management processes and by linking those processes to specific application requirements. This is accomplished without sacrificing control over the storage environment by defining appropriate, reusable policies that support different QOS requirements for each application.

By creating linkages between application requirements and storage management processes in a controlled environment, system, application, and database administrators could control their data in a language that they understand, without the need for extensive storage management skills. Because the data owners can perform certain data management tasks, their ability to respond to changing business conditions is enhanced. In addition, the use of process automation, role-based access, and policy-based management enables business-centric control of data and reduces the interdependencies between storage and data administrators to deliver dramatic productivity and flexibility gains.

The storage and system administrators still have all of the tools and capabilities they always did, but can now create policies that control capacity allocation, protection levels, performance requirements, and replicas. For example, a storage manager can set up policies for different classes of application. The Tier One application can have up to 2TB of capacity, Snapshot™ policies of once a day, remote replication to a specific data center, and a nightly backup to a VTL target. A Tier Two application that requires a lot of capacity can have up to 10TB of capacity, Snapshot policies of once a week, no remote replication, and monthly backups to tape.

Tight integration with popular data center business applications, allowing application and server administrators to manage data without having special storage management skills and freeing storage administrators from help-desk mode requests, will allow data centers to be more efficient and cost effective.

SUMMARY

Continued FC dominance in the data center, implementing tiers of storage across multiple storage systems, increasing storage utilization with thin provisioning, improving storage system resiliency and application availability, combining SAN and NAS under a unified storage architecture, encrypting sensitive data, and creating linkages between application requirements and storage management are the key trends addressing the challenges data centers face today.

With the explosion of data storage capacity and the increasing importance of data to organizations, data centers need to implement a plan to ensure that their data storage and applications are highly resilient and optimized. An organization's health depends on instant data availability, infallible data security, and the ability to quickly respond to changes. Simplifying data and storage management is a key priority for data centers now and in the future.

WP-7011-0107