



ITPMG Insights

Storage Sizing

A Strategy for Gracefully Accommodating Storage Growth

Author: Nick Allen

July 2008

Introduction

Demand for storage is essentially infinite -- limited only by budget and manageability. Typical growth rates are 50% per year, but 300% is not uncommon. So users must develop strategies for accommodating such growth. These strategies need to accommodate both flexibility and efficiency.

A Case Study

Facing sudden and unexpected growth, a university client is looking for help in developing a long-term storage strategy. How can the client begin to develop such a strategy. Our approach to developing the client strategy consists of two parts: 1) Data reduction technologies and efficiency techniques that can be considered today, and 2) actions to take to develop a long term strategy.

Here are some planning assumptions based on conversations with the client:

- There are two campuses located about 30 miles apart and each campus's IT center is the backup for the other.
- The client does not have a formal disaster recovery plan that extends past these two locations except to create periodic backup tapes.
- Green IT is not a priority.
- Regulatory or non-regulatory data archive compliance is not a priority.
- IT staffing levels are not a major concern.
- Server virtualization is not a priority.
- Current storage performance is acceptable and no higher performance applications are anticipated.

Data Reduction Technologies and Efficiency Techniques

Data Reduction

Data reduction reduces the amount of data stored and sent over remote copy links. A variety of technologies have emerged over the last few years to accomplish this. They include:

- Single Instance Store – store a document or file only once, but make it appear that many copies exist. The client already employs this for e-mail.
- Thin Provisioning – makes a host think it has more space available than it really has and only provides real storage on demand. Most vendors now provide this capability.
- Data deduplication – fundamentally a compression technology, but one that can deliver compression ratios of 20:1 to 40:1 – and sometimes more. Several startups have pioneered this technology with a focus on backups. Most of these startups have since been acquired by larger vendors who are integrating the technology into their products' lines. Data deduplication can be done on the server/source side or on the target/storage side. Shortly, we expect to see offerings that also do deduplication in the background.

- Simple Compression – in common use on tape drives today. This technique typically delivers about a 2:1 ratio. Some disk array vendors are introducing it into their products. A file that has been “single-instanced” might benefit from simple compression.

Improved Utilization

Storage consolidation is one of the best ways to improve storage utilization. Much of the client’s storage is already consolidated on large file servers or SAN attached block servers. However, large pockets of un-consolidated storage are starting to show up, e.g., the Macintosh-based multimedia projects. The client needs a strategy for possible consolidation on these pockets.

Although server virtualization, when combined with storage virtualization can also help drive up both server and storage utilization, the client is not contemplating server virtualization at this time.

File and Disk Virtualization

One key benefit of storage virtualization is to improve availability by reducing disruptions when changes are made to the storage environment. The file virtualization market is more mature than the disk market. The Client noted that disruption was a major concern.

Tiered Storage

The disk drive market has evolved to where there are essentially two kinds of drives for datacenter storage: 1) expensive high performance, lower capacity drives that have Fibre Channel or SAS (Serial attached SCSI) interfaces, and 2) less expensive, lower performance drivers with SATA (it really does not matter what SATA means) interfaces. Both kinds of drives are based on similar technologies, but SATA drives are generally less reliable and require RAID-6 or better to make them a viable solution.

Users typically deploy a minimum of two tiers. The top tier uses the more expensive drives to service high availability and higher performance requirements. The second tier uses SATA drives. Lower tiers can use even slower SATA drives or be on different backup schemes. Tape or special purpose compliance archives are also common at the bottom tiers.

Deciding what data to put where, when to migrate it and when to delete it is a major effort and at the heart of any storage strategy.

Virtual Tape

Virtual tape makes disk drives look like tape drives. The most common reason to deploy virtual tape is to speed up backups and especially restores. Essentially virtual tape acts as a cache buffer for tape files. Most virtual tape subsystems support making copies to real tapes for offsite storage. Note that virtual tape has been around for over a decade in mainframe environments where its main purpose is to improve tape cartridge utilization. Data deduplication can play a big

role here, but as the client already uses disk snapshot copy and remote replication it is unlikely virtual tape would provide much advantage, however, it is worth examining.

Developing a Storage Strategy

What should a Storage Strategy Contain?

Start your strategy by determining what the drivers for change are. The main drivers, for most companies these days, remain compliance and rapid data growth. You may well have other business drivers that are specific to your organization such as new departments or a new ERP application. The details of these types of change drivers are often unknown until they occur. Certainly, experience has probably taught you, that estimates of their impact will be understated so always plan for the maximum rather than the minimum. The key to coping with changes is to make your infrastructure flexible.

It is a good idea to standardize on common sets of hardware, partially to achieve that flexibility and partly to keep running costs down. A good rule of thumb is to stick to two or three hardware vendors, as that keeps things manageable, while retaining an element of competition.

When making hardware purchases people claim that price is not the most important factor in making a decision. The most important factors are:

- Product features
- Services
- Price.

In general, very large companies are segregating their data by application on discrete subsystems, while large to medium companies are consolidating their data into a single large subsystem. Most companies are using, or seriously considering SAS/SATA drives for backup and archiving, and even for primary data.

Despite repeated rumors that tape is dead, it is still important for many operations, and not just for traditional backup and recovery. Compliance and Data Archiving are big tape users these days.

Only a methodology that facilitates the gathering of required baseline information can validate an existing strategy and the associated processes. This allows for the leveraging of core or critical areas of storage strategy that in turn will produce an effective plan. While there are many variations to such a methodology, at a minimum it must include:

- Data profiling.
- Business expectation definition.
- Process/policy review.
- Operational review.
- Strategy definition.

Data Profiling

Data profiling consists of gathering critical intelligence regarding "where information lives." Technical attributes such as:

- Platform
- Category
- Source
- Flow
- Confidence
- Security

are identified and documented. Many technical requirements for the storage strategy will ultimately be derived from the information gathered during this process.

Many of the attributes identified during this process are key to realizing tangible cost/benefits within the final strategy. As an example, the category attribute encompasses the definition of data as any of the following:

- Lifeblood: Business stops when unavailable.
- Essential: Business slows when unavailable, and customers are affected.
- Critical: Business slows when unavailable, and internal resources are affected.
- Low priority: Business continues as normal when unavailable.

These data definitions are tailored to meet an organization's specific needs or standards.

In defining data in this way, key decisions can be made regarding how data is stored, archived and accessed. In many cases, organizations treat lower priority data no differently than lifeblood data, resulting in inefficient utilization and higher storage costs.

Continuing to identify the details of each attribute provides the information intelligence necessary to begin to make well-informed, educated decisions in each of the core storage strategy focus areas.

While most data attributes are self-explanatory, security does deserve special emphasis. Many organizations will find that as they begin to profile data, security introduces some unique challenges that often hinder a storage strategy and reduce subsequent cost savings. The organization must be prepared to embark on the somewhat political journey of taking a hard look at security policy and potentially repositioning it as an enabler as opposed to an inhibitor.

Business Expectation Definition

While technical attributes are important, it is the critical business requirements that serve as a major catalyst of the strategic definition process. This is because much of the ROI associated with a storage strategy is derived from the business benefits to be gained.

Querying business units enables the IT organization to identify known issues or inefficiencies from a user perspective. Given that substantial operational inefficiency and costs are derived from the use of information, this phase of intelligence gathering is critical.

During this process, business units are queried, exposing issues of data confidence, accessibility and availability. Engaging the business units also aids in enlisting sponsorship and consensus for change, which are key aspects of defining any strategy.

As organizations progress through the business expectation definition process, it is critical for them to pay attention to inter-business-unit synergy. These synergies are often in the areas of platform or data consolidation, as well as elimination of redundant processes. It is within this phase that inhibitors such as stringent security policies also become apparent

What are the critical success factors to effectively execute this process? They include a detailed, documented, enterprise-wide communication plan, as well as business unit executive sponsorship. As with any information gathering process that affects multiple business units, comprehensive and abundant communication mitigates political risk.

Process and Policy Review

Governance plays a large part in the overall efficient — or inefficient — management and utilization of storage. Improvements in operational efficiency can be realized through effective governance.

Effective policy and process design and implementation results in lower TCO. This promotes management and procurement standards at an enterprise-wide level. Policy within the areas of security, procurement, architecture, acceptable use, archival, backup and recovery, disaster recovery and storage management should all be included within this review process.

Organizations must be aware that significant changes in policy are likely, and as such, the previously mentioned business unit senior management sponsorship is an absolute requirement. Without it, the political process of revising governance will be difficult, if not impossible.

Strategy Definition

Organizations that are successful in completing data profiling, business expectation definition, process and operational reviews, will have amassed the baseline information necessary to craft an enterprise storage strategy. By delving into each of the seven areas of storage strategy — and leveraging the recently amassed information — strategic directions begin to form.

This investigative process will ultimately result in the definition of a number of initiatives and projects the organization will need to execute in order to implement the final strategy. Each area will provide varying opportunity, depending on the organization. Following are several examples:

- **Personnel Alignment** - Organizing for storage presents a number of benefits. For example, storage organizations can effectively influence consistent procurement, which can vary as much as 300% between comparable products. Forecasting and capacity planning become more accurate due to utilization of consistent and repeatable processes. The IT chargeback process becomes more accurate, when storage services such as planning, provisioning and administration are well defined. Disaster recovery becomes more effective as roles and responsibilities for data recovery are clearly within the storage organization. Lastly, storage management and its associated benefits are now owned and orchestrated by a single group responsible for all enterprise storage.
- **Archiving, Backup and Disaster Recovery** - Utilizing data profiles, organizations can now rethink the archiving, backup and disaster recovery processes by ensuring they correspond directly to the criticality of the data. As a rule, lifeblood and essential data are excellent candidates for mirroring and replication technologies, while critical data is better suited for a snapshot. Low-priority data should be either backed up via traditional methods or archived.
- **Storage Management** - Simply stated, storage management provides the tools and processes necessary to do more with less. When you effectively do more with less, benefits are realized.

In addition, storage management is more than network management for storage. Storage management really includes all the traditional proactive management basics but also encompasses data movement and relocation associated with data center consolidation and relocation. Remember, a significant portion of our storage strategy benefits are derived from consolidation. Consolidation implies the cross-platform migration of data, wherein only a few niche players provide solutions.

Policy and Procedure

Policy is less an opportunity area and more an enabler to the other six core focus areas.

Conclusion

A storage strategy does provide valuable opportunity for cost reduction. A storage strategy may also provide tertiary benefits, depending on industry, product base and size. The prerequisite activities of storage strategy definition are straightforward, executable internally or externally and can be political at times.

If desired, the entire process can be componentized and phased, as long as the data profile prerequisite has been completed. Doing so can often provide the cost savings justification necessary to complete the process end to end.

Today's combination of storage hardware and software technologies enable storage-hungry departments to take a holistic approach that addresses data classification, storage capacities and matching the right storage technology to each type of data. Users often employ outside consultants to help develop or at least to review a storage strategy.

To receive additional articles and materials, contact ITPMG at:

information@itpmg.com or

Call 203.743.7538