WHITE
PAPER

IBM Tivoli Storage Manager 6:
Driving Greater Efficiency and Cost Reduction

By Lauren Whitehouse

April, 2009

Table of Contents

Introduction

The design principles of IBM TSM are evidence of the company's foresight concerning efficiency and optimization of data management. TSM's data lifecycle management, integrated archiving, tiered storage architecture, and efficiencies in moving and managing data for operational and disaster recovery are its most notable aspects. Today, data growth's impact on IT environments and, specifically, backup and recovery is the primary driver for IBM to further enhance the TSM platform. TSM's new features can have a considerable effect on the performance, scale, capacity, and management optimization of the backup/recovery and archiving environments—delivering even greater levels of efficiency, and, ultimately, cost reduction.

Situation Analysis

IT organizations are plagued by continued upward trending of data growth, the pressure to maintain regulatory compliance and service level agreements (SLAs) for data availability, and spending cuts driven by the current economic situation. As a result, IT is hyper-focused on cost savings and operational efficiency, but not at the expense of introducing risk or impacting value in their data protection environments.

Data growth is the real culprit here as it necessitates greater investments in IT infrastructure. To minimize risk, IT further perpetuates the problem by making copies of data for operational and disaster recovery. In fact, ESG research respondents cited data protection as the application most responsible for storage growth over the next 24 months.[1] Retention policies—whether to meet compliance, eDiscovery or business intelligence purposes, or to improve recovery objectives—only exacerbate the situation, especially if data is required to be maintained on accessible media, such as disk.

How can IT deal with the conundrum of delivering cost reduction and time savings while maintaining the same levels of SLAs in the face of unprecedented data growth? Optimization investments today that focus on efficiency and cost savings could help organizations emerge from the economic downturn better equipped to keep pace with unabated data growth and demanding compliance requirements and SLAs.

Tivoli Storage Manager

IBM TSM is a client/server software solution providing backup/recovery, archive, hierarchical storage management (HSM), and disaster recovery (DR). TSM has been a leader in helping organizations protect, retain, and assure availability of data for more than 15 years for tens of thousands of customers.

TSM provides automated data protection in a unified recovery platform. It is compatible with a wide range of operating systems, hypervisors, storage devices, and applications, such as database, enterprise resource planning, and messaging environments. Added capabilities for managing desktop/laptop and remote and branch office (ROBO) data with continuous data protection, as well as rapidly rebuilding a server from scratch after a catastrophic failure with bare machine recovery (BMR), can be integrated with TSM.

At the heart of TSM is the TSM database, which maintains information about all clients, files, policies, and scheduling aiding in the backup process. It also allows TSM to satisfy recovery requests—the TSM server queries the database and selects the appropriate file(s) for recovery based on metadata.

[1] Source: ESG Research Report, *Enterprise Storage Survey*, December 2008.

TSM is a unique data management solution. It started out as ADSTAR Distributed Storage Manager (ADSM)—the first disk-based backup solution—and later became part of the Tivoli brand. TSM's tiered storage architecture consists of pools of heterogeneous storage devices—disk, tape, and optical. Data is staged on disk to speed backup jobs and is later migrated to another tier in the storage pool based on policies. Unlike other backup applications that manage tapes, TSM manages backup objects—files, sub-file components, directories, volumes, etc.—and their associated management policies.

Two of TSM's design decisions are based on driving the efficiency of data stored. One is archiving—creating a copy of a file as a separate object in the storage repository to be retained for a specific period of time with the option of deleting the original copy of data on the source system to make additional space available on the client system. The other is progressive incremental backup strategy. After an initial full backup, only those files that have changed since the last backup are retained. This creates efficiency in the amount of data transferred and stored.

Other features ensure that recovery is optimized. Policies dictate the maximum number of backup versions and expiration settings for backup objects. As backup objects are expired, the media contents become fragmented. On tape, fragmentation wastes space and slows restores; therefore, periodic reclamation is performed to reorganize data on media, ensuring that data is stored efficiently on fewer tapes and that media is optimized for recovery. Collocation is another technique that involves relocating data objects from one volume to another to align data objects that belong together. This improves performance and decreases the time required to perform a restore by minimizing the number of sequential access volumes.

The forethought in TSM's architecture and design to create efficiencies in backup and recovery makes it stand apart. As a result, modernization techniques employed in other backup vendors' environments, such as disk-based backup (already part of the TSM fabric) and data deduplication (to eliminate redundancy), will have less of an impact in TSM environments.

TSM 6 Enhancements

IBM TSM 6 includes features that optimize TSM's performance, scalability, manageability, and storage efficiency. These enhancements help organizations respond to today's challenges resulting from relentless data growth, increasing requirements, and budget constraints.

New TSM Database, Monitoring, and Reporting

By switching the underlying proprietary TSM database to IBM DB2, improvements in scalability (the system can now handle more than 1 billion objects per TSM server) and performance (as much as 300% improvement in backup performance) are achieved. The change enables TSM servers to be consolidated over time—an expected 2:1 reduction in the number of backup servers needed—since previous size restrictions have been eliminated.

IBM upgraded the monitoring and reporting capabilities to near real-time and completely redesigned the dashboard that provides visibility into the system. An integrated, common reporting engine based on IBM Tivoli Monitoring (ITM) and Tivoli Data Warehouse (TDW) allows data extracted from the TSM database and maintained in the TDW database to be displayed in the ITM monitoring application. Tivoli Common Reporting (TCR)—the Tivoli standard infrastructure for creating, viewing, and managing Tivoli product reports—provides information on the retrieved data and allows users to mine that data and produce reports. TSM 6 includes new pre-defined reports and the ability to create customized ones, including the option of using open source tools called Business Intelligence and Reporting Tools (BIRT) developed by IBM's Eclipse community. The new monitoring and reporting functionality is a standard feature of TSM—there is no additional charge for these features, including customization of reporting, since BIRT is open source.

Data Deduplication

Data deduplication identifies and eliminates redundant data. In secondary storage processes, such as backup, data is initially seeded on the target storage device (most often, disk) and all subsequently written data is examined for redundancy. Replicate data is not stored twice; instead, a pointer to the stored duplicate data is written (which takes up significantly less space). Deduplication is usually not suitable for data on tape because increased fragmentation of data could greatly lengthen access times.

TSM now has native deduplication capabilities built in to its Extended Edition—at no extra charge. After data is written to the primary disk pool at full disk performance, a deduplication operation can be scheduled to eliminate redundancy at the sub-file level. The process involves dividing the object into logical sections, called chunks, and then identifying duplicate chunks. Duplicates are identified using a hash algorithm to produce a unique identifier and then comparing that ID with a central index to determine if the chunk is already stored. If the chunk is a duplicate, then the index is updated to reference the matching chunk and the space containing the redundant chunk is freed. Deduplication, therefore, allows for more data to be retained on disk for faster access.

Deduplication can be enabled or disabled on a per storage pool basis. Data deduplication may make the most sense for data that will remain on disk for an extended period of time, has a high degree of redundancy (a common operating system, for example), and has access to TSM server CPU and disk I/O resources for deduplication processing. Situations where deduplication may not be optimal include: data that will soon be migrated to tape; TSM servers that do not have sufficient CPU and disk I/O resources; and data whose recovery could not risk delay due to accessing chunks that are not stored contiguously.

Data deduplication offers the benefits of reducing the capacity of data stored. The combination of backup policy, retention settings, data type, and rate of change will ultimately impact deduplication ratios. In a TSM environment, the impact of deduplication may be less than in other backup environments due to progressive incremental strategies that avoid duplicating non-changing backup data in the first place.

Faster Backup for NAS N Series Products

NAS filers provide affordable networked storage, but often present a challenge for backup. Unlike traditional network servers, NAS devices run a closed operating system, so a TSM client cannot be created or installed. TSM Extended Edition uses Network Data Management Protocol (NDMP)—an open standard protocol for backing up network-attached storage—to perform high-performance backup and recovery outboard of the TSM client and server. NDMP was designed to enable NAS devices to directly communicate with tape and other backup devices

TSM 6 leverages NetApp's Snapshot Difference API to specify the differences between two snapshots, allowing TSM to perform backup of only the changed files—an incremental backup. This eliminates the need for TSM to perform file scanning of filers, greatly improving backup performance and enabling file-level recovery from image-level backups.

TSM 6 also takes advantage of NetApp's SnapMirror to Tape technology, enabling customers to use TSM NDMP support to back up FAS and IBM N series volumes directly to tape or across the network to any tape device managed by TSM, improving backup speeds tenfold.

VMware Support Improvements

VMware Consolidated Backup (VCB) allows the backup of multiple virtual machines to be offloaded to a dedicated physical host serving as the backup proxy. By offloading backup from the virtual machines and ESX server, backups can be completed LAN-free and eliminate overhead at the source virtual machine. VCB captures the virtual machine disk image (.vmdk file) via a backup proxy and TSM can target the virtual machine disk image for backup to TSM storage.

TSM 5.5 worked with VMware's Consolidated Backup (VCB) image-level backups, but required users to do scripting themselves. The new version now automates the process within the GUI.

Item-Level Recovery for Microsoft Exchange and Active Directory

Recovery of an individual element in Microsoft Exchange required a cumbersome, two-step backup/recovery process. If rapid backup was required, a full Exchange, or "store-level" backup, was performed. Unfortunately, it's extremely difficult and time-consuming to recover an individual mailbox or item (e.g., message, contact, task, journal entry, appointment, etc.) from a store-level backup. To speed recovery of individual mailboxes or messages, a first pass store-level backup and then a second pass individual mailbox backup, or "brick-level" backup, must be executed—effectively doubling the time and capacity needed to protect Exchange at a granular level. Given the hassle and the time it takes for the two-step backup process, it's typical for brick-level backup to be executed less frequently or more selectively.

Now, TSM 6 allows for image-level backup as well as individual mailbox and item-level recovery for Microsoft Exchange environments. Recovery of a user's mailbox or items within a mailbox that may have been accidentally deleted is now possible. Locating items to recover based on a number of criteria, including sender, subject, attachment name, folder, delivery date/time, or text strings within the message body, is available.

A similar capability exists for Active Directory. TSM 6 enables individual object recovery from System State backups of user and group objects. Support for un-deletion (reanimation) of objects from the Active Directory's tombstone assists in the manual recovery of individual objects—without having to shut down/restart the Active Directory server.

Value Delivered in TSM 6

The TSM platform is already a base of efficiency for data management as a unified platform for backup/recovery, archiving, HSM, and DR. The solution was designed to incorporate multiple functions, operating systems, applications/databases, storage devices, and recovery strategies.

Solving the Distributed Data Backup Challenge

The TSM platform is also an aggregation point for data management of endpoints, such as desktops, laptops, and ROBOs. IBM Tivoli Storage Manager FastBack is a standalone solution for Windows server environments, offering continuous backup and near-instant recovery. FastBack can support ROBO data protection, replacing local tape-based backup solutions which may require operator intervention. FastBack backups can be integrated with TSM by allowing TSM to back up a mount point from FastBack.

IBM Tivoli Continuous Data Protection for Files (Tivoli CDP for Files) is a data protection solution specifically targeted at Windows desktops/laptops. It offers transparent real-time replication and traditional backup services as a standalone or TSM-integrated solution. CDP for Files detects changed files using journaling and stores copies on local and/or remote disk. Since it uses native file system format for stored copies, users can recover files themselves. Both CDP for Files and FastBack drive down costs of managing highly distributed data by consolidating backup for edge data in the core data center.

Intelligent Data Management

With exponential data growth rates, efficiently managing and protecting data continues to be a major challenge for IT organizations. Many are uncertain about what data should be retained or deleted, often resulting in unnecessary costs. More and more, customers want to tier their storage environments and retain data according to its value. This approach allows organizations to reduce operational expenses, decrease risk, and more efficiently manage data throughout its lifecycle of usefulness to the organization.

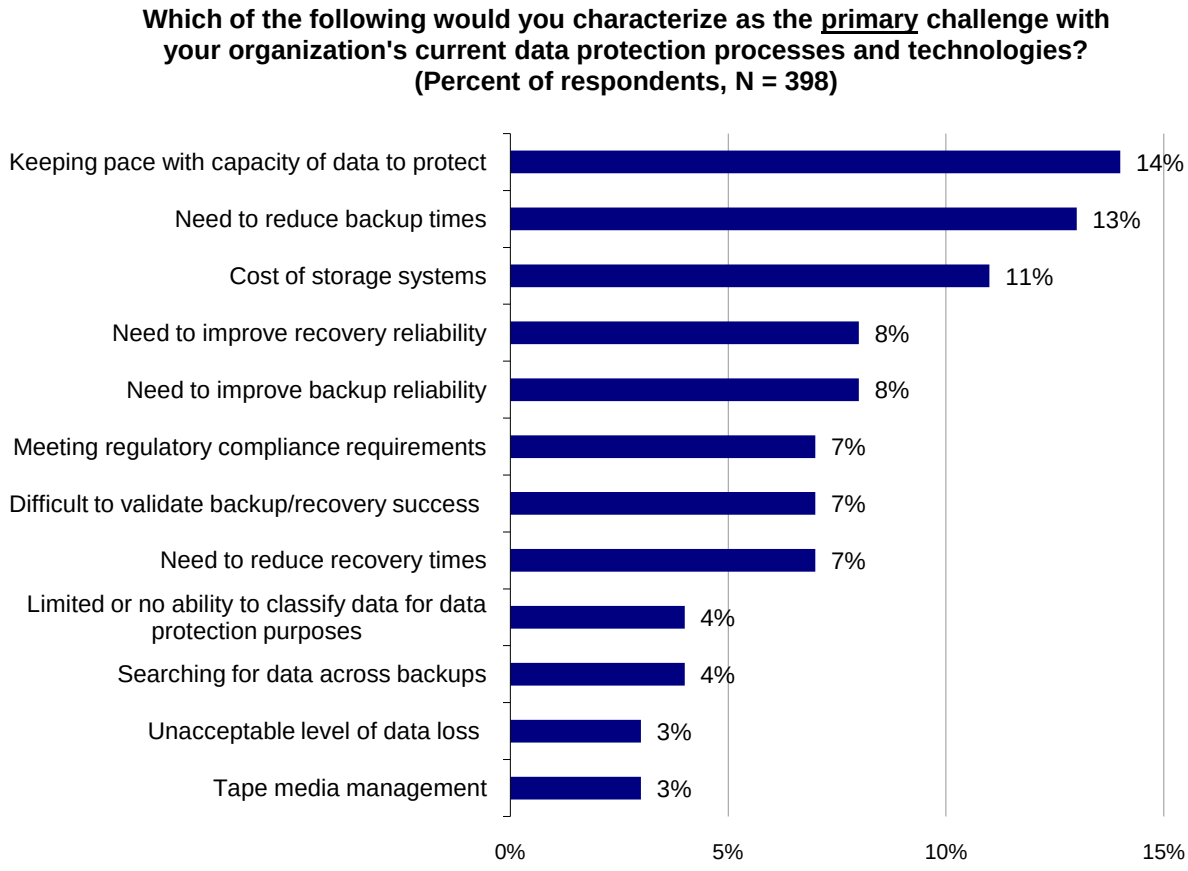
It is this concept that IBM uses in TSM. TSM maintains storage pools in a hierarchical structure with a policy engine specifying how and when files should be automatically moved among storage pools during their lifetime and, optionally, when files should be deleted. For example, backup data can be initially stored on TSM disk

space for performance reasons before being moved to tape media. Enabling disk-to-disk and disk-to-disk-to-tape (and other media combinations) for backup allows TSM to provide a data lifecycle management methodology for data management. Migration activity is automated to reduce manual intervention, enabling IT organizations to improve both storage and resource utilization.

Optimization of Backup/Recovery

ESG research revealed that data protection challenges are widespread and varied, resulting in less-than-desirable backup and recovery success rates. Many of the top challenges involve making backup and recovery more efficient—to improve performance and reliability—and ensuring that capacity and costs are kept in check. Organizations are eager to deploy new technologies—such as backup-to-disk, snapshots and replication, and data deduplication—that should help improve overall data protection effectiveness.

FIGURE 1. DATA PROTECTION CHALLENGES



Source: Enterprise Strategy Group, 2008

TSM's architecture and design are steeped in optimization, especially when it comes to the organization, movement, and storage of data to maximize backup/recovery efficiencies. Features such as integrated archiving, progressive incremental backup, and data deduplication combine to solve the challenges presented by ever-increasing data growth. Optimizing the speed of backup and recovery is achieved via TSM's disk staging capabilities, continuous data protection modules, image-level capture with item-level recovery, as well as TSM's reclamation and collocation approaches for reorganizing data to speed recovery.

Further, management of TSM is improved with near real-time monitoring and comprehensive reporting. The visibility provided by TSM allows IT to save time over manual data collection, document capacity usage, initiate trending, resolve vulnerabilities in the environment, lay the foundation for setting up and communicating policies

and procedures, institute better planning, reduce trouble tickets, and more. Through these tools, IT has the potential to deliver business, operational, and financial benefits while increasing customer satisfaction.

A Solution for Now and the Future

The unique needs of one data center are rarely duplicated in another. That's why IBM TSM is packaged to meet the needs of organizations of any size and is designed to scale as the environment does. A high degree of customization—coupled with the ability to grow and adapt in complex, high-growth IT environments—has set TSM apart from the rest.

IBM took this concept a step further when it improved the scale and performance of TSM by leveraging DB2 for the internal TSM database, also improving the manageability of the environment through enhanced monitoring and reporting. IBM TSM is a data protection and management solution that can be tuned for organizations' needs today, while maintaining flexibility and scale to accommodate future growth.

This is noteworthy considering that organizations are more inclined to stick with a solution over time. Only situations where organizations need to reduce backup and recovery times or have general difficulty keeping up with growth in data that needs to be protected are causes, cited by research respondents, to re-architect the data protection environment. Furthermore, organizations surveyed by ESG have a significant interest in single vendor solutions, with 80% citing a preference for standardizing on a single vendor for data protection processes.[2]

Summary

The amount of pain an organization experiences when protecting and managing data today is highly dependent on its rate of data growth and the ability of its data protection solution to keep pace with that growth and ongoing requirements. Budget constraints aggravate the situation, making it necessary to seek out improvements and optimization in the environment.

IBM TSM is a data management platform distinguished from other solutions in its architecture and approach. The solution has several unique and progressive features that help it achieve high degrees of efficiency, which can lead to time and cost savings for its users. TSM 6 carries on this theme with several noteworthy enhancements that allow current and prospective customers to benefit from its revolutionary and well-timed value propositions.

20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com

[2] Source: ESG Research Report, *Data Protection Market Trends*, January 2008.