White Paper

Symantec Backup Exec 2010

Addressing the Root Causes of Inefficiency in Data Protection

By Lauren Whitehouse

December, 2009

Contents

Introduction

The continued upward trending of data growth, spending cuts driven by current economic factors, and pressure to meet increasingly inflexible service level agreements (SLAs) continue to beleaguer IT organizations. With data protection processes, IT is challenged to deliver cost and time savings while maintaining or improving backup and recovery SLAs—an especially difficult task considering that ESG research respondents cited data protection as the application most responsible for storage growth over the next 24 months.[1] Such overabundance of data is often the result of a lack of policies focused on data retention and destruction.

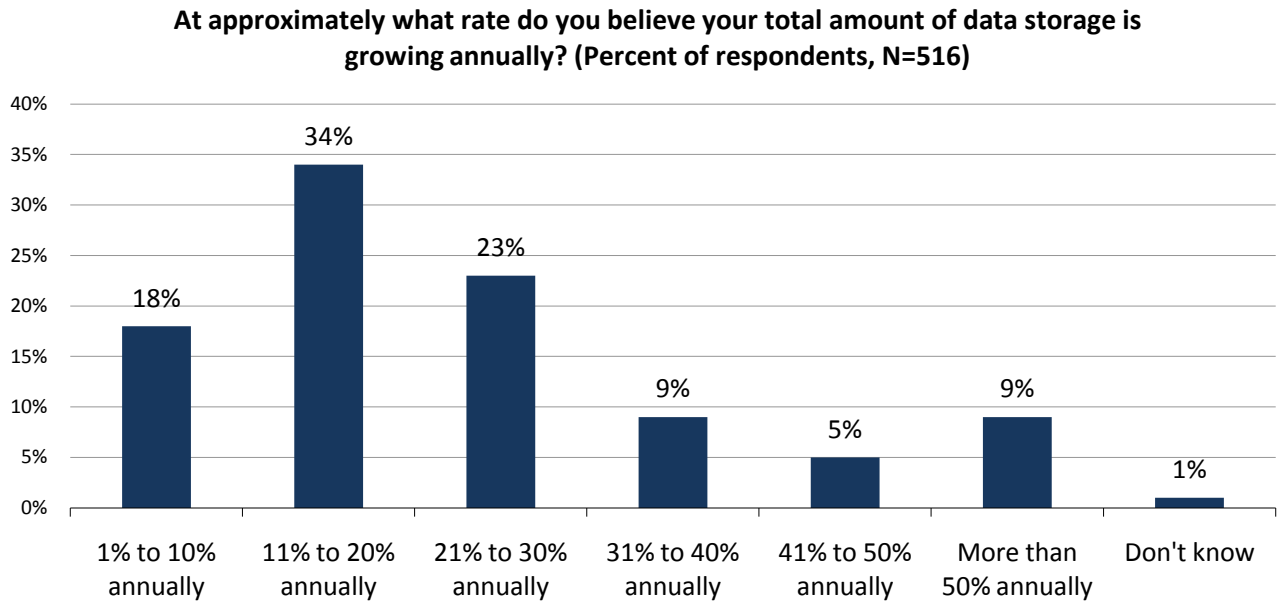
That's where Symantec Backup Exec 2010 comes in: one of the major themes associated with its introduction is storage optimization. In addition to the efficiency benefits of reducing capacity requirements for primary disk storage via integrated archiving, Backup Exec 2010 also includes block-level data deduplication to eliminate redundancy in data protection processes. The combination of these features introduces capital and operational cost savings, greater operational efficiency, and improved SLAs.

Root Causes of Storage Inefficiency

Contributing Factors

Data growth and poor data lifecycle management are two of the root causes of the storage capacity, bandwidth, and operational inefficiencies in IT. Many mid-market organizatios are experiencing double-digit data growth, making it difficult to keep pace with management, storage, and data protection. ESG research found that 34% of survey respondents report annual growth rates in the 11-20% range and 23% expect 21-30% percent growth per year. Still another 23% believe their storage capacity is growing at over 30% annually (see Figure 1).[2]

Figure 1. The Data Growth Burden



Source: Enterprise Strategy Group, 2008

Growth rates for primary storage directly impact the data protection environment. As the volume of primary data expands, so too does the amount of time it takes to make backup copies. For many organizations, the "window" of time to perform daily backup jobs is under pressure from the 24/7 nature of today's businesses. Many organizations struggle to complete backups within the prescribed off-peak time period so that the process doesn't impact production workloads.

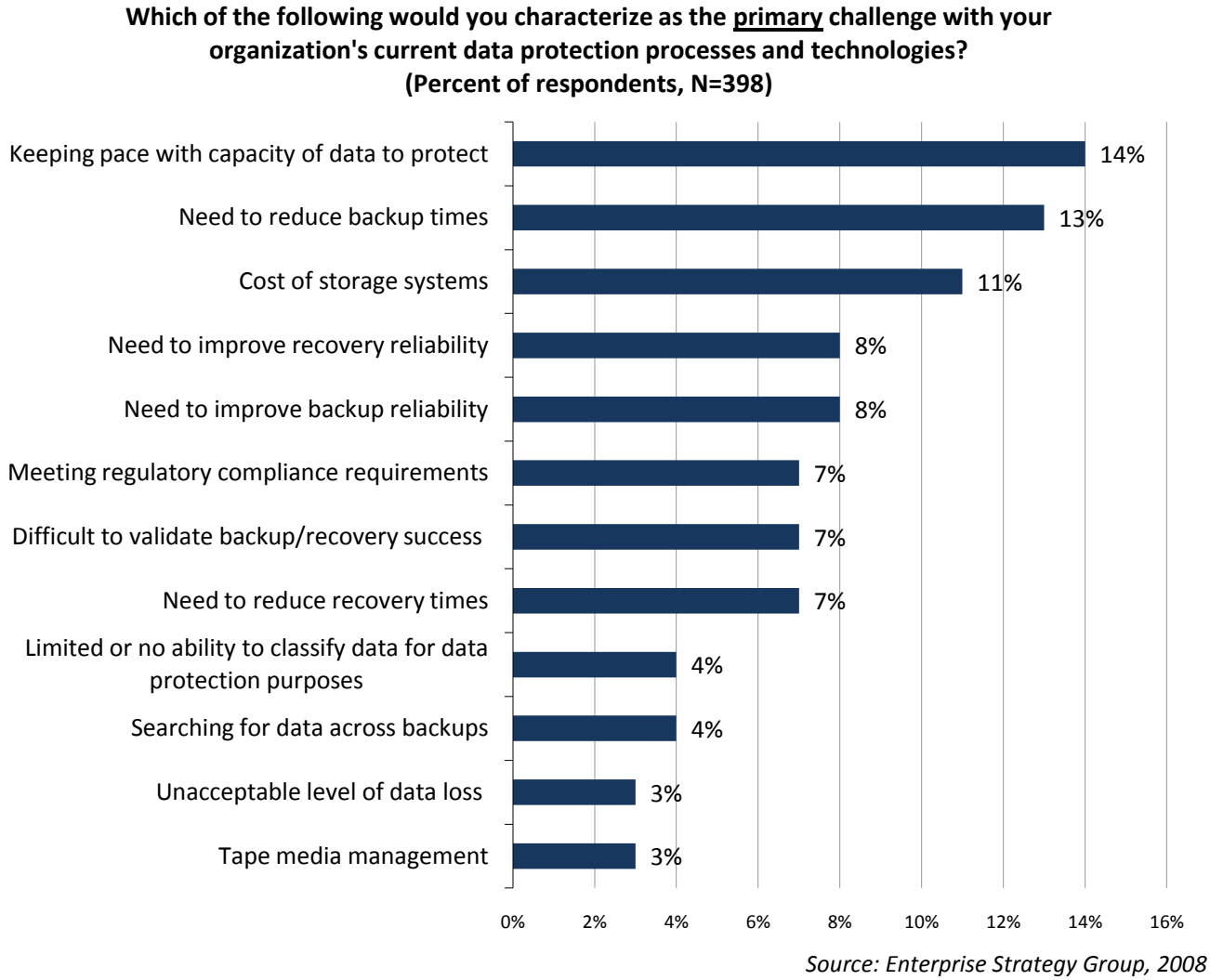
[1] Source: ESG Research Report, *Enterprise Storage Survey*, December 2008.
[2] Source: ESG Research Report, *Medium-Size Business Server and Storage Priorities*, June 2008.

The storage capacity glut is often exacerbated by data protection processes which retain multiple copies for data recovery purposes. Often, daily, weekly, and monthly copies are maintained both on- and off-premises. In many cases, duplicate data residing on multiple systems is backed up and replicated. Unchecked, the volume of backup data can easily exceed that of primary data.

The lack of corporate governance regarding the retention and destruction of data further contributes to the problem. An overwhelming majority of corporate data is unchanging and infrequently accessed, yet is often maintained on premium data center real estate and is afforded the most stringent data protection policies—perpetually—even though its usage pattern has changed (and copies of the same data exist in several places). Not only does this persistent data create ongoing capacity requirements and a drag on production systems' performance, it causes backup congestion and network bottlenecks.

The result? Managing and protecting increasing volumes of data—oftentimes with the same or less IT staff—is constraining IT's ability to meet business needs. ESG research found that respondents were most challenged with keeping pace with the capacity of data to protect, followed by backup window issues and storage costs (see Figure 2).[3]

Figure 2. Data Protection Challenges



Source: Enterprise Strategy Group, 2008

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

What's at risk? Service level agreements (SLAs) negotiated between IT and its business constituents regarding data recovery time and recovery point objectives (RTOs and RPOs) become difficult to meet. Costs rise for storage capacity, network bandwidth, and data center environmentals such as power and cooling. There can also be operational consequences as more manpower is required to protect and manage workloads.

Addressing the Problem Without Optimization

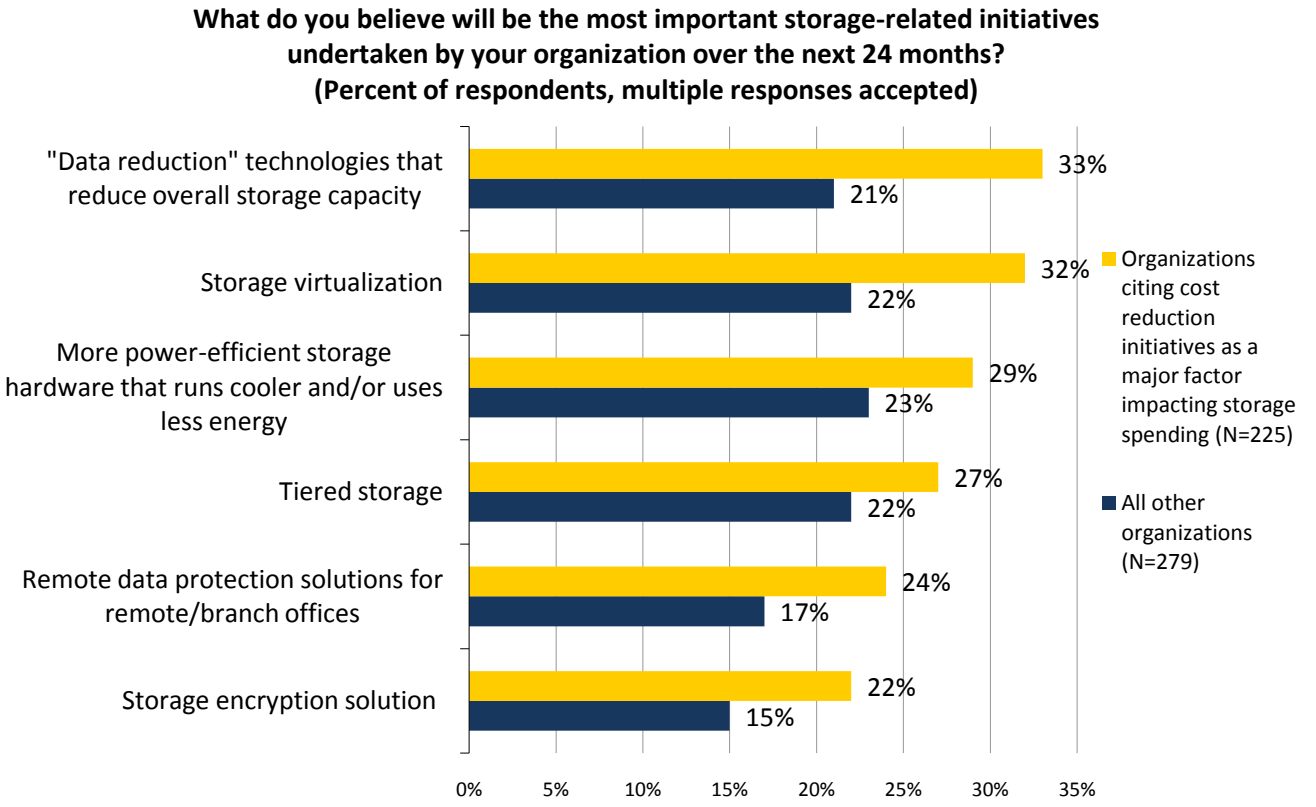
IT organizations often respond by treating symptoms instead of dealing with root causes. Making continual investments in infrastructure to support annual growth, such as ongoing primary and secondary storage capacity purchases, is just one of many IT band-aids.

To deal with backup window pressure, many organizations insert disk in the backup data path. This greatly improves the performance and reliability of backup and recovery processes when compared to a backup-to-tape approach. It also reduces dependence on tape media in backup and recovery processes, thereby reducing operator intervention for tape handling and troubleshooting. However, adding more infrastructure—in perpetuity—results in higher (and ongoing) costs. The root cause of the problem is not being addressed.

Storage Optimization

Creating efficiency is a top concern for IT organizations. An ESG poll regarding cost reduction initiatives found data reduction technology that optimizes storage capacity to be a top IT priority (see Figure 3).[4]

Figure 3. Top Storage-Related Initiatives for Cost Reduction



Source: Enterprise Strategy Group, 2009

[4] Source: ESG Research Report, *Data Center Spending Intentions*, January 2009.

Optimizing storage to reduce overall storage capacity can be accomplished in a few ways. Data deduplication identifies and eliminates redundant data. After the data is initially seeded on a secondary storage device, subsequently written data is examined for redundancy. Replicate data is not written again; instead, only a pointer to the unique data is stored.

Archiving persistent data involves moving unchanging data from more costly primary storage to another, less costly tier in the storage hierarchy. It preserves a record for long-term retention, pruning data from primary storage in the process.

These technologies can be used independently; however, the combination of technologies amplifies results.

Deduplication

Data deduplication optimizes backup storage by identifying and eliminating redundant data within a defined domain. The cost implications of reducing the capacity of data transferred and stored in data protection processes can be significant. Deduplication has the potential to drive the cost of disk-based backup closer to that of tape, making it more feasible to introduce disk-based data protection. As previously noted, leveraging disk in backup and recovery processes improves performance, reducing the time required to complete these tasks. Since the footprint of backup copies is smaller with deduplication, more workloads can back up to disk, improving RTOs. Alternatively, the frequency of backup copies on disk can be increased, improving recovery point objectives (RPOs).

Reduced capacity requirements reduces incremental capacity purchases. On the other hand, the disk capacity "recovered" via deduplication allows organizations to lengthen retention policies for data on disk, improving the likelihood that data will later be recovered from disk. The savings also impact sustainability efforts since a reduction in disk storage lowers power and cooling requirements.

Deduplication has impact beyond storage capacity. Transferring one-tenth of the data[5] in backup and recovery processes, including the LAN and SAN traffic between the source production system and backup storage as well as WAN transport between remote and branch offices and primary and disaster recovery sites, could yield significant bandwidth savings.

Archiving

Archiving for storage optimization should not be confused with the practice of keeping backup copies on tape in an offsite location for periodic recall. With archiving, a copy of a file is created as a separate object and stored in a disk-based archive "vault" based on a time or threshold policy. It's retained for a specific period of time in the vault and offers the option of deleting the original copy residing on the source system.

This process:

- **Reclaims space on primary systems**, potentially delaying incremental capacity purchases, contributing to performance improvements for source systems and applications, and saving money.
- **Reduces the volume of data in backup/recovery processes**, improving performance (and associated backup and recovery SLAs), lowering bandwidth and capacity requirements, and saving money.
- **Addresses retention and deletion of data**, supporting compliance requirements, minimizing tape operations for long-term retention/preservation of data, and saving money in terms of tape media, operational staff responsible for media handling and tape troubleshooting, and offsite media storage fees.

[5] ESG research found that 48% of survey respondents using data deduplication reported a reduction ratio of 10-20x.

Backup Exec 2010

Symantec Backup Exec is a leading backup/recovery solution suitable for small and medium-sized organizations. It provides disk-to-disk-to-tape backup and recovery for systems running Windows, Linux, Macintosh, Solaris, and NetWare as well as VMware and Microsoft virtual machines. Backup Exec 2010 includes storage optimization enhancements including data deduplication and file and e-mail archiving.

Deduplication

The block-level data deduplication technology proven and popularized in Symantec NetBackup PureDisk is offered as an option in Backup Exec 2010. In addition, Backup Exec 2010 supports highly optimized integration with third-party target-side deduplication solutions.

Symantec supports multiple deduplication implementations. Users can implement deduplication at different points in the backup data path to accommodate the performance requirements of various workloads. Client-side and media server deduplication strategies benefit from Backup Exec's content awareness—enabling the block size to be automatically adjusted based on the natural pattern breaks of the data.

Client-side deduplication. Deduplication processing occurs at the client/source system, delivering end-to-end bandwidth efficiency as only a fraction of data traverses the LAN, WAN, and/or SAN. Deduplication processes are distributed across clients and deduplication data is shared by all Backup Exec Agents backing up data to the same media server.

Media server deduplication. This approach centralizes deduplication processing at the Backup Exec media server(s), offloading the process from source systems. Data from production servers is transferred to the media server, where it is examined for redundancy. Only unique data is transferred to disk storage.

Appliance deduplication. Backup data from production servers is transferred in its entirety across the LAN, WAN, and/or SAN and duplicate data is eliminated in a device purpose-built for the process at the end of the backup data path. The network efficiency gains of the client-side approach are eliminated.

Backup Exec 2010 supports a common interface for integrating third-party disk targets: OpenStorage (OST). Through this interface, OST-enabled disk targets optimize backup processes. An OST-enabled device performs remote replication for disaster recovery without engaging the Backup Exec media servers, but still tracking all copies in Backup Exec's catalog. Backup copies made by the target device still follow established retention policies. The OST interface can also optimize performance by removing CIFS/NFS overhead for NAS target devices.

Just as third-party target-side solutions can efficiently move data between sites for disaster recovery or backup consolidation, Backup Exec 2010 can facilitate deduplicated copies of backup sets. Backup Exec 2010 facilitates copies of deduplicated data from one or more Backup Exec 2010 media servers to a site running either Backup Exec 2010 media server or NetBackup PureDisk.

Archiving

The Backup Exec 2010 Archiving Options integrate file and e-mail archiving capabilities—based on technology in Symantec Enterprise Vault. Backup and archiving processes are delivered via a single Backup Exec 2010 agent and management interface. Federating backup and archiving allows consolidation of data lifecycle management activities such as retention and expiration.

Some of the key features of Backup Exec 2010 archiving include:

Off-host processing. All archive activity is done at the Backup Exec Media Server, reducing overhead on content sources. Data is archived from backup copies, reducing data movement by transmitting data once during the backup of the content source.

Disk-only archive vault. Disk storage is optimized through single-instance storage (file-level deduplication).

Policy-based archiving. Archive processing is initiated based on size and/or date conditions.

Administrator- or user-initiated retrieval. Self-service retrieval of archived data is accomplished with the Backup Exec Retrieve Web Console.

Content indexing. Indexing of data enables end-users and administrators to perform detailed searches of archived data.

Migration path to Enterprise Vault. As sites grow or eDiscovery needs expand, Symantec enables a transition from the archiving option in Backup Exec 2010 to the more full-featured, eDiscovery-enabled archiving available with Enterprise Vault.

The Bigger Truth

Organizations are challenged to manage greater volumes of data based on growth rates and retention policies. The effect is an increase in capital and operational costs, performance degradation of systems and applications, and an inability to meet backup windows and recovery objectives. Non-compliance, data recoverability, and unmet SLA risks are also introduced.

Symantec is addressing these issues by dealing with the root causes—not just addressing the symptoms. The archiving and deduplication features available with Backup Exec 2010 drive storage optimization for primary and secondary disk storage and drive network optimization across the LAN, WAN, and/or SAN.

The benefits of implementing Backup Exec 2010 with its storage optimization options are very compelling when financial, operational, and business benefits are considered. Backup Exec 2010 can protect more workloads while storing less data and saving more—more time and, importantly, more money.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel:508.482.0188 Fax: 508.482.0218 | www.enterprisestrategygroup.com