

White Paper

NetApp Storage Efficiency and Data Protection

Reducing Costs and Reducing Risk

By Lauren Whitehouse

July, 2010

This ESG White Paper was commissioned by NetApp and is distributed under license from ESG.



Contents

Introduction	3
Digital Dependency and Data Growth	3
Managing Cost versus Risk	5
NetApp Storage Efficiency	7
NetApp Data Protection	7
The Bigger Truth	9

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.

Introduction

Companies today are finding a few things inevitable: dependence on data to run the business, continuous data growth, and the need to spend more to manage data and make sure it’s available. [NetApp](#) storage solutions address these challenges with several storage efficiency features—including snapshot, thin provisioning, deduplication, and RAID DP—that drive down the costs of hardware, software, operational staff, and network bandwidth. When combined with NetApp data protection solutions including SnapVault, SnapMirror, and MetroCluster, IT organizations are able to protect more and reduce risk while, importantly, keeping cost metrics in check. This paper will show how NetApp storage efficiency and data protection features allow for a higher level of protection without compromise on risk or cost.

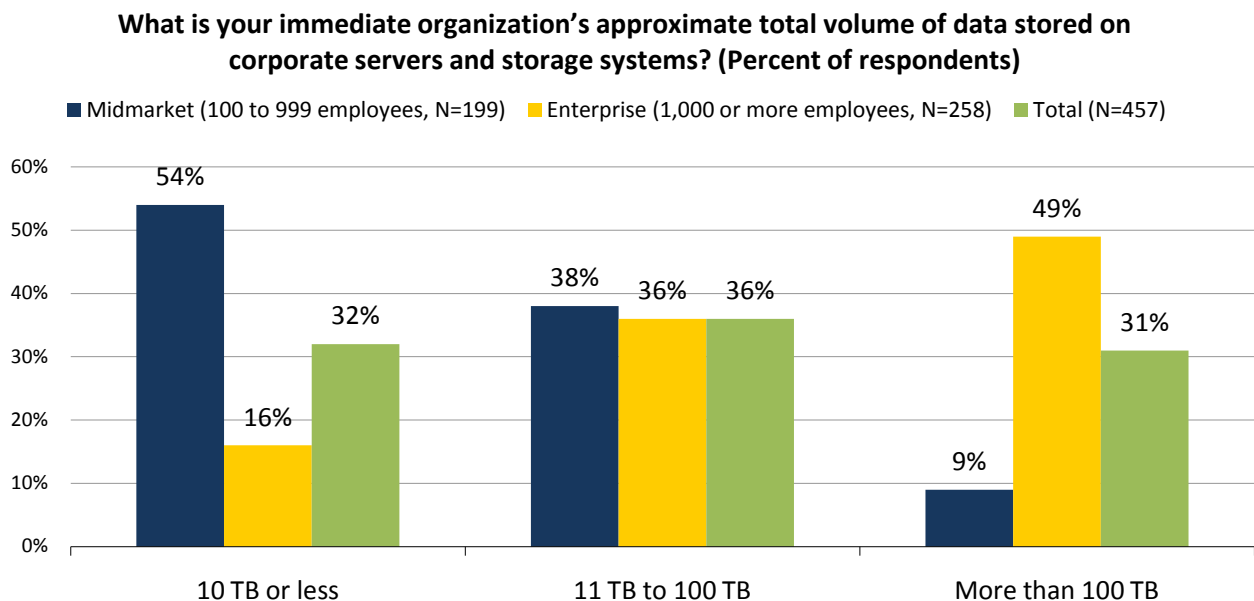
Digital Dependency and Data Growth

Companies are ever-dependent on digital content to conduct business. Everything from financial records, employee information, customer data, and sales orders to video surveillance, e-mail, supply chain data, and third-party logistics information are maintained digitally. As more and more aspects of the business are digitized and automated via the use of computers, it becomes more imperative to protect those digital assets to maintain business continuity.

New digital content is generated every day, increasing capacity requirements. In addition, organizations are often subject to retention requirements, whether due to regulatory or corporate mandates. All too often, however, ambiguity regarding data lifecycle policies force companies to retain data indefinitely, contributing to data growth.

ESG research found that organizations are managing large and expanding storage environments. As shown in Figure 1, the total volume of data residing on corporate servers and storage systems ranges from less than 10 TB (32%) to more than 100 TB (31%), with the remaining 36% of organizations reporting between 11 to 100 TB of total data to manage. Storage volumes of 26 to 50 TB are the tipping point between midmarket and enterprise-class organizations, with 70% of midmarket organizations falling into the 25 TB or less range and 64% of enterprise organizations citing storage volume at more than 25 TB.¹

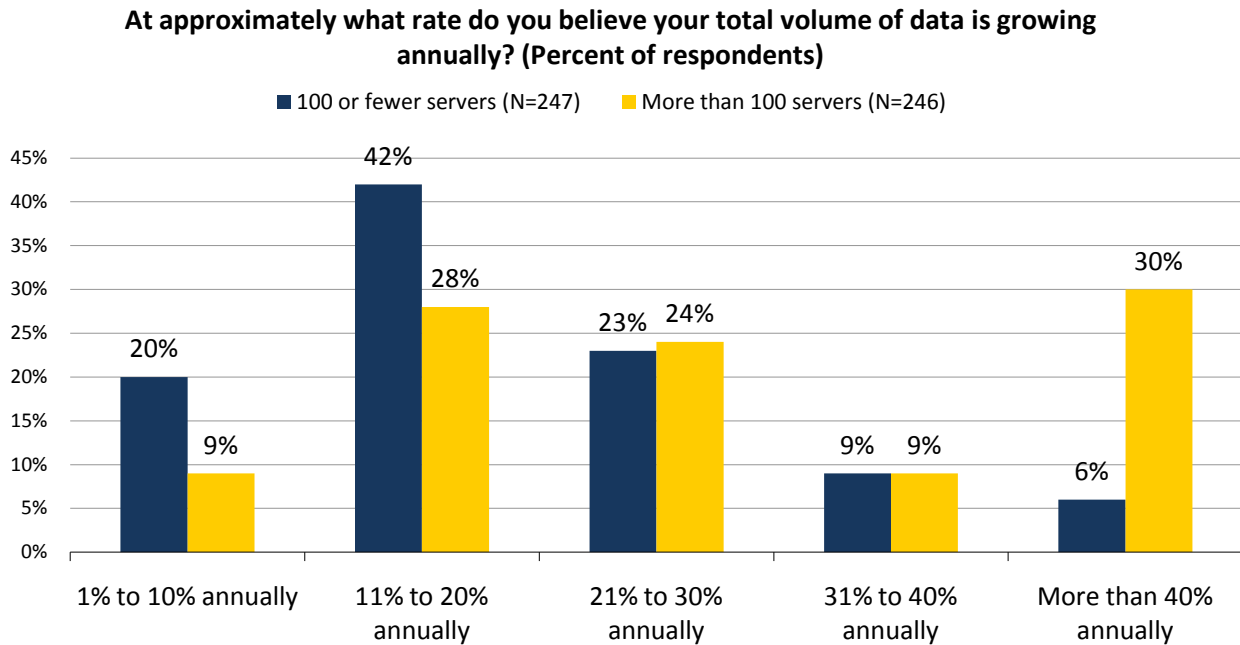
Figure 1. Total Volume of Data Stored on Corporate Servers and Storage Systems, by Company Size



Source: Enterprise Strategy Group, 2010.

¹ Source: ESG Research Report, [Data Protection Market Trends](#), April 2010. All subsequent statistics come from this report unless otherwise cited.

Figure 2. Annual Data Growth Rates, by Number of Production Servers



Source: Enterprise Strategy Group, 2010.

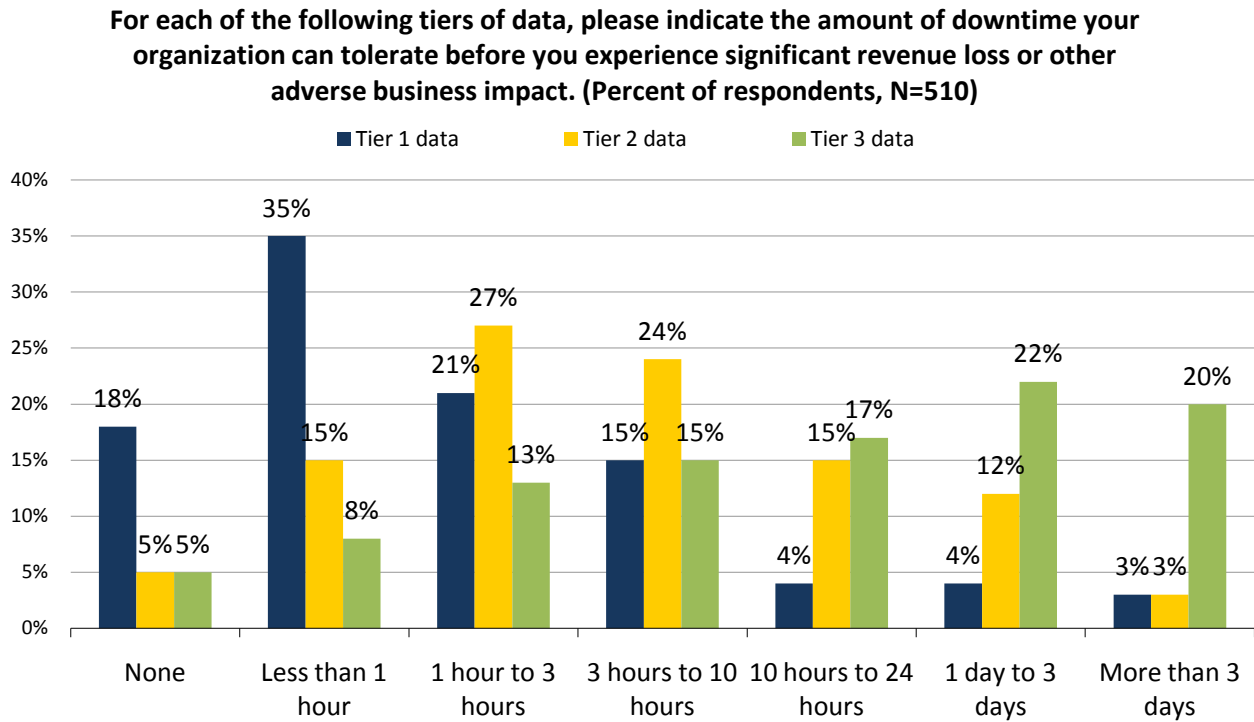
Annual growth rates are also impressive, with the majority (58%) of ESG research respondents citing growth rates in the 11% to 30% range. Figure 2 shows a clear relationship between the number of production servers an organization has deployed and the expected annual growth rate of data. The majority of organizations with fewer than 100 production servers (62%) anticipate growth rates of up to 20%, while the majority of organizations with 100 servers or more (63%) predict annual growth rates of more than 20%. Compounded annually, the total volume of data doubles every three to five years.

The impact of data growth can be seen in a few ways:

- **Increased CAPEX.** Ongoing investments in storage capacity, compute power, and network bandwidth.
- **Increased OPEX.** As more infrastructure is added, requirements and costs for power, cooling, and data center floor space go up. Furthermore, it all has to be managed, potentially impacting operational staff costs. And as more data is moved within and between sites, bandwidth requirements increase—as do expenses.
- **Challenges with business continuity and data protection.** More data to protect means higher costs. As each new server or storage system is deployed, high availability configurations and/or media for backup and recovery are often required.

Downtime introduces higher risk for missed business opportunities, employee or customer dissatisfaction, lost data, lost revenue, damage to the company’s reputation, lost productivity, and even legal liability. ESG research found that over 70% of organizations can experience only three hours or less of downtime for Tier 1 data before suffering adverse business impact (see Figure 3). Similarly, 71% of organizations cited 10 hours or less of downtime tolerance for Tier 2 data, and 58% of organizations cited 24 hours or less of downtime for Tier 3 data. Businesses have to be responsive and resilient for any eventuality—seamlessly managing growth while minimizing risk. Oftentimes organizations have to triage their data—providing greater levels of protection to some tiers than others. Vulnerability is likely introduced when the volume of data to be protected exceeds what can be accommodated, forcing IT to adjust protection strategies to lower standards.

Figure 3. Downtime Tolerance



Source: Enterprise Strategy Group, 2010.

Managing Cost versus Risk

The business value of applications and data differentiates protection strategies. As seen in Figure 3, applications and data are often categorized into a few classes or categories—tiers or Gold/Silver/Bronze designations are fairly common. Classifying applications and data allows IT to establish baselines that broadly distinguish service levels and protection strategies.

Typically, applications and associated data must meet certain criteria. ESG defines a simple three-tier model with the following definitions:

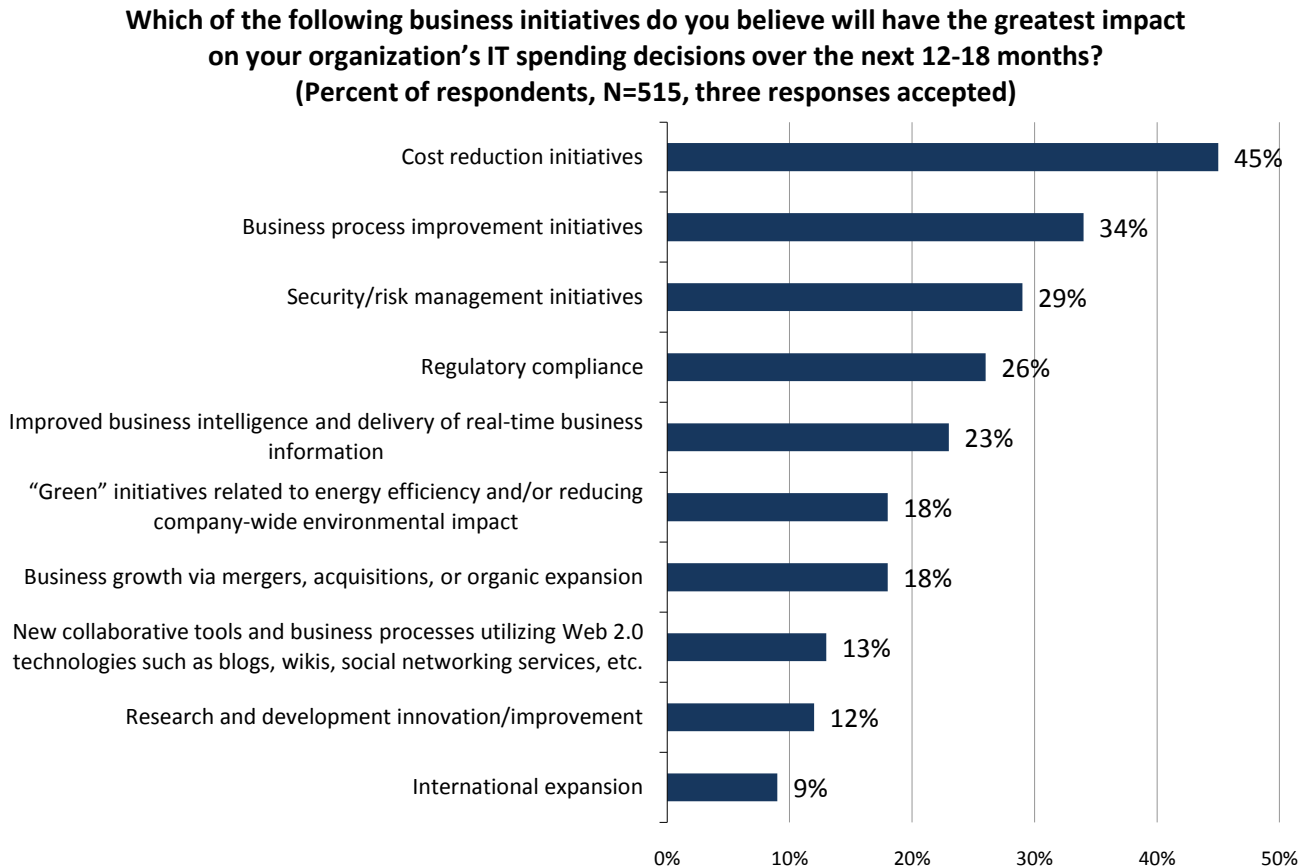
- **Tier 1:** Mission-critical data and applications with the highest requirements for both availability and performance.
- **Tier 2:** Data and applications requiring good performance and reliability, but not at the level of mission-critical data. This can include a mix of less-critical application data and older mission-critical data.
- **Tier 3:** Typically archived data, so performance is less essential, but the data must be retrievable.

For each class of application/data, risks must be assessed. Once the implications of downtime and data loss are understood, IT organizations can use these factors to determine application availability requirements and the protection mechanisms that should be applied to ensure compliance.

Of course, once the desired quality of service parameters are defined and recommendations are made for providing them, the negotiations begin. The probability and consequences of each risk must be balanced with the cost of mitigating it. For example, what is the cost of lost revenue versus the costs incurred to minimize the risk of downtime for an application or data integral to generating revenue?

Figure 4 shows the top business initiatives impacting IT spending in 2010.² Managing cost (cost reduction initiatives), “cycle time” (business process improvement initiatives), and risk (security/risk management initiatives) are most often at the top of the list since they have the greatest impact on business success.

Figure 4. Business Initiatives Impacting 2010 Spending Decisions



Source: Enterprise Strategy Group, 2010.

Storage efficiency is maximizing the use of storage capacity. Given the aforementioned data growth rates, it’s a topic on every IT administrator’s mind. Taking advantage of storage efficiency techniques to better manage data growth allows IT to, ultimately, save money. Budget savings may be realized through deferring new storage systems or capacity expansion purchases, or by lowering operating costs.

When it comes to business continuity and disaster recovery (DR), a number of specialized technologies are available to mitigate risk. Many organizations apply layers of protection, often using a mix of replication or mirroring for primary system/storage business continuity, and snapshot and once-per-day backup for operational and disaster recovery. The result is “over insurance” at the topmost tiers of applications and data and “under insurance” in the others. In keeping with the storage efficiency theme, this approach begs the question: is there a more cost-efficient strategy that can also reduce risk?

The issue of storage efficiency is not limited to primary storage capacity. Copies of production data made for data protection can amplify storage efficiency issues. Capacity glut at the source will impact downstream network and backup storage capacity—both on- and off-premises—as well as the operational overhead associated with additional infrastructure. Therefore, a lack of storage efficiency will impact backup and disaster recovery efficiency.

² Source: ESG Research Report, [2010 IT Spending Intentions](#), February 2010.

NetApp Storage Efficiency

To achieve maximum cost savings, companies need to address data growth in both primary and secondary systems. Data storage inefficiencies that are not addressed in primary storage tend to ripple throughout the storage infrastructure and can increase costs. Storing more data in primary can not only mean more storage is required in primary but also that more management, networking equipment, and bandwidth is needed. It can increase length of time for backups as well.

NetApp's Storage Efficiency portfolio addresses the challenges of dealing with data growth while managing capital and operational costs. NetApp provides several products and features that address space efficiency while driving down the costs of hardware, software, and operational staff in both the primary and secondary storage.

- **Snapshots** – Snapshot copies enable system administrators to create many point-in-time copies of entire data volumes, but consume only a fraction of the space normally required to make backup copies of these volumes.
- **Thin provisioning** – Thin provisioning in FlexVol allows storage managers to virtually allocate capacity to users without physically allocating it. Storage is only physically allocated as it is used (i.e., “just in time”), eliminating the issues associated with over provisioning and waste. In addition, volumes can be sized and resized dynamically and automatically according to policies. Thin provisioning can be used on both primary and secondary systems to safely enable higher levels of storage utilization—a key factor in getting the most out of existing storage.
- **Deduplication** – Deduplication eliminates duplicate data at the block level, using minimal system resources and with nominal impact. NetApp deduplication can occur “upstream” at primary storage and creates efficiency for all points downstream in the data path with minimal performance impact.
- **RAID DP** – RAID DP protects users from having a drive fail while another drive in the same RAID set is being rebuilt. It's a high performance version of RAID 6 that is a more cost-effective alternative to RAID 10 (mirroring) for protecting data against more than one simultaneous drive failure.

While these technologies reduce primary storage costs, and in the case of snapshots reduce risk, they do not by themselves solve the challenge of creating a cost-efficient solution that reduces risk. To address the second part of the equation, lower risk, availability, and recovery features capable of extending the value of storage efficiency to secondary storage must be available.

NetApp Data Protection

NetApp protects primary data and extends the reach of its storage efficiency technologies using integrated data protection services that provide high availability, disaster recovery, backup, and compliance functionality. These embedded services simplify management; because the product set is unified within the storage, there is less management overhead than managing multiple standalone solutions from different vendors. This reduces costs and ultimately eliminates the need for IT to have to make tough choices about what—and what not—to protect.

SnapVault addresses backup and recovery needs. NetApp's snapshot technology forms the core of SnapVault. Leveraging the benefits of snapshots to provide block-level incrementals, SnapVault reduces network consumption, storage capacity, and backup and recovery time for highly efficient disk backup and recovery. After an initial full copy is made, only changed blocks are stored—yet they are always accessible directly in a non-proprietary format. This is an important distinction as it can greatly reduce the time and complexity associated with locating and recovering data.

SnapMirror addresses disaster recovery needs for long distances and does not require special networking equipment or dark fibre, which can be quite costly. As its name implies, SnapMirror synchronously, asynchronously, or semi-synchronously replicates the local snapshot copy from the primary to one or more secondary storage systems over a LAN or WAN in a one-to-one, one-to-many, many-to-one, or cascading configuration. Data deduplicated on primary storage will be replicated in a deduplicated format, reducing both the amount of data sent and the storage required on the DR system. It works across any NetApp platform whereas

other solutions require replication from one kind of storage to the same kind of storage. This means that, unlike its competitors, NetApp can support a replication strategy between different tiers of storage. SnapMirror includes network compression to reduce network consumption to reduce impact on and need for expensive network upgrades. In addition, SnapMirror also has the ability to test recoveries without impact to the production system using FlexClones—providing the added assurance of the integrity of a disaster recovery strategy.

MetroCluster addresses continuous data availability needs by combining array mirroring with clustered storage controllers, enabling IT to gain high availability. MetroCluster enables automatic failover triggered by a component, storage node, or site failure and protects against the major causes of downtime in the data center (power, network, heating/ventilating/air conditioning [HVAC], and human errors). As with SnapMirror, MetroCluster supports deduplication on both cluster nodes, enabling IT organizations to significantly reduce storage costs. MetroCluster supports a variety of clustering configurations: within a data center, across a data center, on a campus, or in a metropolitan area. Implementing MetroCluster creates efficiency in capital expenses invested in high availability and in operational expenses associated with staff costs since it lowers management overhead and provides seamless, automated failover.

NetApp FAS and SnapLock addresses archive and compliance challenges. When archiving of data is required, NetApp FAS can be leveraged as a high efficiency storage platform for long term retention of data. While any archive data stored on FAS can take advantage of storage efficiency technologies like deduplication, compliance data especially benefits. Using SnapLock, administrators can convert volumes into WORM storage for preventing modification or deletion of data in accordance with local or governmental regulations. And because snapshots, SnapVault, and SnapMirror are available, archive data can easily be backed up or replicated offsite to prevent data loss in the event of an outage.

The Bigger Truth

Considering the volume of data under management and current growth rates of 20% or more, it's no wonder that organizations are challenged with keeping pace with the capacity of data to protect. Continuously increasing data growth has an obvious impact on data protection strategies from the amount of infrastructure required to capture and store backup copies, to the ability of IT administrators to create and move copies within and between physical locations, to all the of time and resources to manage the devices, software, and processes.

From a cost standpoint, sometimes choices have to be made. IT has to triage the application and data environments and separate the "critical" from the "important" and "best effort" classes, applying protection technologies and strategies to mitigate risk while also balancing the cost investment.

NetApp has a different vision—one enabled by storage efficiency and Data OnTap data protection technologies—that eliminates the need to choose between reducing risk and reducing costs. NetApp customers get *both*.



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