

**Contributors**

Jesse St. Laurent

Peter Baer Galvin

Richard Roy

# Data Storage Outlook: 2008 Summary Report

## An IT Strategy Research Report

June, 2008

### Abstract

This research report summarizes the results of 178 in-depth, face-to-face interviews with senior IT decision-makers addressing a broad range of storage requirements and challenges—including storage spending and headcount, top storage concerns, installed storage infrastructure, SAN, NAS, and DAS capacity, plus availability, backup, and security requirements. Extensive quantitative analysis is included, as well as insightful comparisons to third-party data from organizations such as Forrester, Gartner, and *Storage Magazine*. The report also includes nine robust technology-oriented tutorials that cover topics such as the data deletion challenge, thin provisioning technology, Information Lifecycle Management, Virtual Tape Library technology, and WAN acceleration. »

inside front cover  
BLANK

EXECUTIVE SUMMARY	3
• Scope & Methodology	3
• Key Data Findings	4
• Opportunities	7
• State-of-Technology Observations	8
• About Corporate Technologies, Inc.	9
INTERVIEW DATA & ANALYSIS	10
• Current and Projected Storage Spending	10
- Current Storage Spending	10
- Projected Storage Spending	10
- Conclusions	11
• Current and Projected Storage Headcount	12
- Current Storage Headcount	12
- Projected Storage Headcount	12
- Conclusions	13
• Top Three Storage Concerns	14
- Current Storage Concerns	14
- Trends in Storage Concerns	15
- Conclusions	16
• Most Urgent Problems Related to Storage	16
- Current Urgent Problems	16
- Trends in Urgent Problems	17
- Conclusions	17
• Installed Storage Infrastructure	18
- Conclusions	19
• Storage Capacity in SAN vs. NAS vs. DAS	19
- Current and Projected Overall Storage Capacity	19
- Current and Projected Storage Percentages	20
- Current and Projected Mean Storage Capacity	20
- Storage Capacity Growth Rates	20
- Trends in Storage Percentages and Means	22
- Conclusions	22
- <b>TUTORIAL:</b> The Data Deletion Challenge	23
• Utilization of Storage	24
- Allocated Storage Currently Written to	24
- Current Calculation of Storage Utilization	24
- Trends in Storage Utilization	25
- Conclusions	25
- Thin Provisioning Technology	26
- <b>TUTORIAL:</b> Thin Provisioning Technology	26
• Information Lifecycle Management (ILM)	27
- Current ILM Adoption	27
- Trends in ILM Adoption	27
- Conclusions	27
- <b>TUTORIAL:</b> Information Lifecycle Management (ILM)	28
• Availability Requirements	29
- Current Storage Arrays: Snapshot Capability	29
- Current Continuous Data Protection (CDP)	29
- Trends in Availability Requirements	30

**TABLE OF CONTENTS**

- Conclusions	30
- Data Protection Technologies	31
- <b>TUTORIAL:</b> RAID, Snapshots, and Filesystems	31
• <b>Backup Requirements and Processes</b>	33
- Current Components of Backup Strategy	33
- Current Number of TB Backed Up	33
- Current Time Required to Backup	34
- Current Backup Windows Met	34
- Current Percent of Backups Failing	34
- Current Backup Reporting Tools Used	35
- Current Tape Drive Technologies Used	35
- Trends in Backup Requirements and Processes	36
- Conclusions	36
- Tapes, Backups, and Restores	37
- <b>TUTORIAL:</b> The Truth about Tapes, Backups, and Restores	37
• <b>Virtual Tape Libraries (VTL)</b>	39
- Current VTL Adoption	39
- Trends in VTL Adoption	39
- Conclusions	40
- <b>TUTORIAL:</b> Virtual Tape Library (VTL) Technology	40
- <b>TUTORIAL:</b> Deduplication Technology	42
• <b>Remote and Mobile Work Force</b>	43
- Current Adoption of Remote Access Technology	43
- Trends in Adoption of Remote Access Technology	43
- Conclusions	44
- Wide Area Network (WAN) Acceleration	45
- <b>TUTORIAL:</b> Wide Area Network (WAN) Acceleration	45
• <b>Storage Disaster Recovery (DR) Requirements</b>	48
- Current Data Replication to a DR Facility	48
- Acceptable Data Loss (Recovery Point Objective)	48
- Time to Be Back Live (Recovery Time Objective)	49
- Testing of the IT DR Plan	50
- Trends in Storage Disaster Recovery (DR) Requirements	50
- Conclusions	52
• <b>Storing Employee Email</b>	53
- Length of Time Stored	53
- Capacity Dedicated to Storing Employee Email	54
- Trends in Storing Employee Email	54
- Conclusions	55
• <b>Storage Security Requirements</b>	55
- Current Data Encryption on Backup Tapes	55
- Use of Third-Party to House Backup Tapes	56
- Trends in Data Encryption on Backup Tapes	56
- Conclusions	57
- <b>TUTORIAL:</b> Storage and Security	58
END NOTES	59
APPENDIX: KEY VERTICAL MARKET DATA	60
REPORT CONTRIBUTORS	61

## Executive Summary

### Scope & Methodology

Data Storage Outlook“ conducted in-depth, face-to-face interviews with senior IT decision-makers such as Chief Information Officers and Vice Presidents of IT, **across two waves of research**. In the second wave (**Wave 2**), 129 interviews were completed. In the first wave (**Wave 1**), 49 interviews were completed. During these meetings the interviewees addressed a broad range of storage-related questions designed to gauge their current and projected storage requirements and to identify their storage challenges.

The respondents represent several key vertical markets, including financial services, manufacturing, and life sciences. These organizations are located in the Northeast and represent mid-market and Fortune 1000 companies. Companies range in revenue size from less than \$250 million to multi-billion dollars.

### STRUCTURE OF THIS REPORT

---

This report will provide you with:

- Information on peer activities that will help you benchmark your data storage requirements
- Best-practices that will guide you on a range of storage topics
- Insight into technology options and trends to help drive your strategic planning
- Informative conclusions that explain the implications of key findings
- Robust tutorials to help educate your IT team on important storage topics

We encourage you to contact us with questions about the data and analysis in this report: [info@StorageOutlook.com](mailto:info@StorageOutlook.com) or 781.273.4100. One of our staff members will respond to your inquiries.

This report is designed to provide you with a wealth of data storage information in a variety of formats:

- Robust technology-oriented tutorials that cover topics such as:
  - The data deletion challenge
  - Thin provisioning technology
  - Information Lifecycle Management (ILM)
  - RAID, snapshots, and filesystems
  - The truth about tapes/backups/restores
  - Virtual Tape Library (VTL) Technology

- Deduplication technology
- Wide Area Network (WAN) acceleration
- Storage and security
- Extensive quantitative analysis of the interview data and anecdotal examples
- Interesting market survey data for key verticals such as financial services, manufacturing, and life sciences
- Selected survey data segmented by company revenue size
- Insightful comparisons to third-party data from organizations such as Forrester, Gartner, and *Storage* magazine
- Selected comparisons of research data from **Wave 1** and **Wave 2**

#### RECOMMENDATIONS ON HOW TO USE THIS REPORT

---

- The detailed Table of Contents will help you use this report as a reference guide to a broad range of information on storage issues and technologies.
- Contact us for additional copies so you can provide each member of your team with this information to educate them about recent technology trends.

## Key Data Findings

Following is a summary of the important data findings captured by Data Storage Outlook. In addition to these data findings, informative conclusions which explain the data and discuss significant implications can be found at the end of each major section.

#### STORAGE SPENDING

---

PAGE 10

- The majority of senior IT decision-makers plan to maintain or grow their storage spending over the next 12 months, with an average projected storage budget increase of 6.6%.

#### STORAGE HEADCOUNT

---

PAGE 12

- The majority of respondents project that their storage administrative headcount will remain the same over the next 12 months in spite of the fact that storage capacity and storage budgets are increasing.

**URGENT STORAGE PROBLEMS**

PAGE 16

- Reporting and monitoring, capacity/performance, disaster recovery, and backups are cited as urgent problems not being effectively addressed by current storage solutions.

**STORAGE CAPACITY**

PAGE 19

- In the next 12 months, both SAN (Storage Attached Network) and NAS (Network Attached Storage) capacity are expected to grow as a percentage of total storage, while the DAS (Direct Attached Storage) share of storage is expected to decline. These growth projections vary widely by size of company and key vertical markets.
- Key drivers for NAS and SAN growth include traders' needs for increased data; regulatory issues such as SEC, HIPAA, Sarbanes-Oxley (SOX), the U.S. Patriot Act, and others; ecommerce initiatives; implementation of digital repositories; corporate data retention policies; and support of email environments.

**UTILIZATION OF STORAGE**

PAGE 24

- Nearly half of respondents have 50 to 74 percent of their storage written to a percentage that has changed little across the term of the research.
- The overwhelming majority of respondents calculate their utilization percentages via estimates a reliance that has increased substantially over the two waves of research.
- Many IT decision-makers are under the false impression that their existing storage products offer them access to automated tools to track storage utilization.
- IT organizations can defer and potentially eliminate storage costs by using thin provisioning technology.
- IT organizations often waste floor space, spin idle disks, and cool excess storage capacity in their data centers.

**INFORMATION LIFECYCLE MANAGEMENT**

PAGE 27

- ILM (Information Lifecycle Management) referring to setting policies for which type of data should reside on which type of storage is not widely understood by IT professionals. However, once this concept was explained, a near-majority of respondents state they are evaluating and considering use of ILM.

**AVAILABILITY REQUIREMENTS**

PAGE 29

- Among respondents using storage arrays with snapshot capability, nearly half of them use differentials which increased significantly across the two waves of research.

**BACKUP REQUIREMENTS AND PROCESSES**

PAGE 33

- Respondents report a multi-component backup strategy, with most mentioning tape, followed by snapshots, full copies, and 2nd tier disk.
- A majority of respondents state they are barely meeting their backup window, or not meeting it at all a percentage that has increased significantly over the term of the research.
- Nearly all respondents state that backup reporting tools are in place, although Data Storage Outlook is of the opinion that such tools have no/low actual reporting capability.

**VIRTUAL TAPE LIBRARIES (VTL)**

PAGE 39

- Although only one-fifth of respondents report that VTL is a part of their storage environment, its use has increased over the term of the research.
- VTL offers a potential solution to the never-ending backup treadmill challenge, by providing a tape interface to a disk-based system, allowing easy deployment into an existing backup infrastructure.

**REMOTE AND MOBILE WORK FORCE**

PAGE 43

- Technology preferences for remote access of corporate data have changed across the term of the research, with highest growth reported in the use of a centralized file server via Wide Area File Services (WAFS).

**STORAGE DISASTER RECOVERY (DR) REQUIREMENTS**

PAGE 48

- Respondents are equally divided on replicating data from their production environment to a DR facility.
- A slim majority of respondents report they have the IT infrastructure to support their Recovery Point Objective (RPO).
- More 85% report they are meeting their Recovery Time Objective (RTO) although equally split between just barely and easily.
- Over one-quarter of those surveyed admit they do not have an IT DR test plan.

**STORING EMPLOYEE EMAIL**

PAGE 53

- The term of email storage is lengthening, with those reporting indefinitely increasing from 24% in the first wave of research to 35% in the second wave.
- Storage dedicated to employee emails has increased dramatically 58% from the first wave of research to the second wave.

---

**STORAGE SECURITY REQUIREMENTS**

PAGE 55

- Surprisingly, only 12% of IT decision-makers currently encrypt the data on the backup tapes they send off-site.
- The majority of companies currently use a third-party to house their backup tapes, and they find it is easy to retrieve these tapes when they need them.

## Opportunities

- Almost all computer facilities face challenges around data retention and deletion. Deciding which data to retain versus delete remains an inefficient process it is often less costly to simply add additional disk space than to figure out which data can safely be deleted. *Refer to the tutorial: The Data Deletion Challenge.* Page 23
- IT organizations can defer and potentially eliminate storage costs by using thin provisioning technology. *Refer to the tutorial: Thin Provisioning Technology.* Page 26
- Information Lifecycle Management (ILM) can increase efficiencies and decrease costs by setting policies for which type of data should reside on which type of storage. *Refer to the tutorial: Information Lifecycle Management (ILM).* Page 28
- There is an opportunity to understand the implementation differences among RAID, snapshots, and filesystems, so they may be used optimally in a given environment. *Refer to the tutorial: RAID, Snapshots, and Filesystems.* Page 31
- Backup and restore windows are often challenges. Tape drives based on new technology SAN-attached tape drives, near-line storage, and snapshots can resolve these issues. *Refer to the tutorial: The Truth about Tapes, Backups, and Restores.* Page 37
- As long as data footprints continue to expand, storage will expand and backups will continue to be a problem. Virtual Tape Library (VTL) technology offers a potential solution to this problem with easy deployment into an existing backup infrastructure. *Refer to the tutorial: Virtual Tape Library (VTL) Technology.* Page 40
- Exponential expansion of data is fueling the growth of deduplication technology, which promises to remove redundant data so only unique sequences are stored. The technology is currently emerging, but once mature, it should offer compelling economics. *Refer to the tutorial: Deduplication Technology.* Page 42
- Companies have long wanted to allow more concurrent data to flow between locations, and also to speed up that data. Wide Area Network (WAN) Acceleration technology accomplishes this goal, providing LAN-like speed across the WAN. *Refer to the tutorial: Wide Area Network (WAN) Acceleration.* Page 45

- Regulations and corporate disasters are pressuring companies to increase their security across the board. There are many aspects to storage security and many new technologies that can decrease unauthorized access and increase logging of all access to data. *Refer to the tutorial: Storage and Security.* *Page 58*

## State-of-Technology Observations

- Two recent studies one from Google (refer to End Note 1) and a second from Carnegie Mellon University (refer to End Note 2) have been published with findings that may profoundly change many beliefs about storage. For example, the Google study finds that higher temperatures are not associated with higher disk drive failure rates. The CMU study finds little difference in replacement rates between SCSI, FC, and SATA drives suggesting that disk-independent factors such as operating conditions affect replacement rates more than component-specific factors. Data Storage Outlook will make every effort to take such findings into account in the third wave of research, as the most significant findings are validated in the marketplace.
- Emerging technologies such as thin provisioning, deduplication, encryption, and virtualization are increasingly coming less from the established large storage providers, and more from VC-funded startups, the most promising of which will be acquired. Such a VC as R&D scenario presents support, implementation, and migration challenges both to the acquiring enterprise, and to those adopting the technology.
- Data Storage Outlook finds many companies polarized on their adoption of technology. Those conservative on technology adoption tend to stay with tried and proven approaches, typically from mainstream vendors. Those more aggressive on technology adoption tend to experiment with newer technologies, often from emerging companies. An interesting middle area between these two extremes is represented by a handful of newer technologies that can solve storage problems including driving storage costs down far better than traditional solutions, and do so with a risk acceptably in line with the expected return. Some examples of these "middle area" technologies include disk-to-disk-to-tape backups, lower-cost replication, WAN optimization, and deduplication. These solutions are discussed throughout Data Storage Outlook.

Consider Disaster Recovery, where newer technology is bringing better solutions to older storage problems. Well-established suppliers of DR solutions would have companies believe that Disaster Recovery is expensive and suitable for only large enterprises. However, improvements in WAFS, low-cost communications, and built-in replication have resulted in innovative DR solutions that are relatively easy to implement and affordable for any-size organization.

## About Corporate Technologies, Inc.

Corporate Technologies developed Data Storage Outlook™ in order to ensure that our storage-related products and services meet the needs of our clients and the marketplace, and provide senior IT decision-makers with timely peer data. It has become part of our ongoing market research activities.

For over a decade, Corporate Technologies has been helping companies tap the power of technology to drive the success of business. We are experts in every aspect of IT and pride ourselves on solving your toughest IT challenges. We deliver a full spectrum of cost-effective, reliable infrastructure hardware, software, and professional services in the areas of storage, networking, security, systems, support, and data integration.

For more information contact: [info@StorageOutlook.com](mailto:info@StorageOutlook.com) or call 781.273.4100.

## Interview Data & Analysis

### Current and Projected Storage Spending

#### CURRENT STORAGE SPENDING

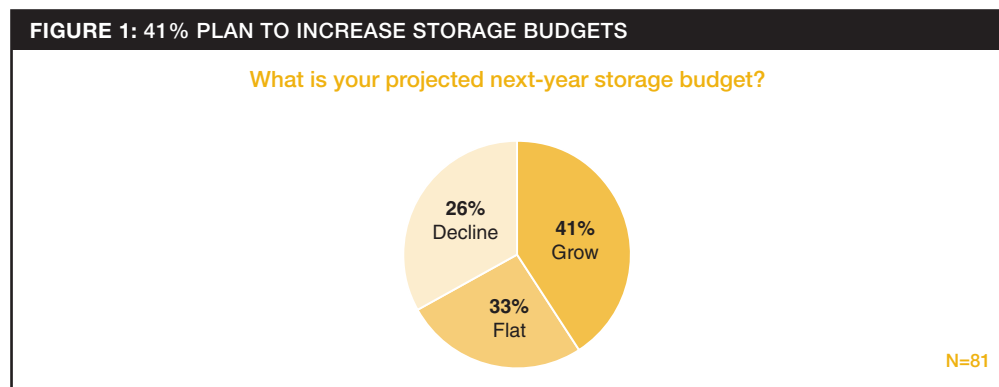
Data Storage Outlook finds that while storage budgets among respondents average \$593,310, they vary considerably by company, from less than \$100,000 to greater than \$2 million.

Companies with annual revenue in excess of \$750 million report storage budgets averaging \$1.135 million. Companies with annual revenue less than \$250 million report diverse storage budgets, ranging from less than \$100,000 to over \$1 million.

Storage budgets also vary by industry, with financial services firms reporting average storage budgets of \$879,000 larger than other vertical industries surveyed.

#### PROJECTED STORAGE SPENDING

Three-quarters of senior IT decision-makers interviewed for the Data Storage Outlook study plan to maintain or grow their storage spending over the next 12 months. This finding is consistent with a recent Gartner survey, which finds that between 70% and 80% of respondents expect their 2007 storage budget to increase or remain the same for storage hardware, software, and services (refer to End Note 5).



Source: Data Storage Outlook—Corporate Technologies, Inc.

Overall, 41% of respondents report that their projected storage budgets will increase, versus stay flat or decline (Figure 1). This finding is also consistent with recent industry analyst reports. A Forrester Research 2006 study of IT decision-makers reveals that 45% of companies plan to grow their spending on storage (refer to End Note 6). Data Storage Outlook also finds that 47% of financial services firms plan to increase their storage spending going forward.

Data Storage Outlook finds that overall storage budgets are projected to increase an average of 6.6%, and will comprise an average of 13.8% of respondents' overall IT budget. The projected growth rate is validated by a recent *Storage* magazine survey, which finds budgets rising by about

5.2% (refer to End Note 7). The percentage increments documented by Data Storage Outlook vary considerably by company from less than 10% to over 50%. Interestingly, the storage budget as a percentage of the IT budget is relatively constant, across both company size and vertical industry segments.

**CONCLUSIONS: CURRENT AND PROJECTED STORAGE SPENDING**

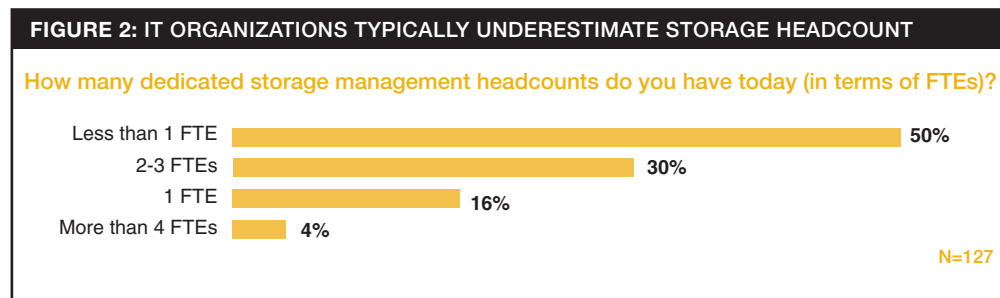
*Storage budgets are growing because storage capacity is growing. Not only is data multiplying at an astonishing rate, but so also is companies unwillingness to delete data. (Refer also to the tutorial: The Data Deletion Challenge.) Even in companies most notably smaller companies that have traditionally had only a central IT budget, Data Storage Outlook finds a trend towards establishing line-items for storage budgets, and towards clearer definitions as to what components comprise such budgets.*

## Current and Projected Storage Headcount

### CURRENT STORAGE HEADCOUNT

The following observations are based on interviews conducted with IT decision-makers for the Data Storage Outlook study, as well as our general field experience. Due to the fractionalized nature of storage management responsibilities, it is often difficult for senior IT decision-makers to calculate how many Full Time Equivalents (FTEs) are dedicated to storage. Storage has become the responsibility of many IT administrators. For example, there might be an organization with 12 people who each spend some fraction of their time involved in storage-related activities. It is typical for an IT organization to greatly underestimate the time it spends managing storage since multiple people are involved in this process.

Data Storage Outlook finds that 66% of the respondents have only 1 FTE or less dedicated to storage management, while the remaining respondents vary widely in their estimates from less than 1 FTE to more than 4 FTEs (Figure 2).



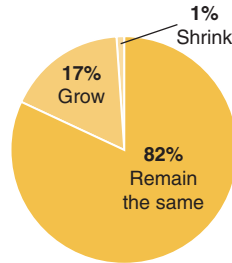
Source: Data Storage Outlook—Corporate Technologies, Inc.

### PROJECTED STORAGE HEADCOUNT

The study reveals that 82% of these IT decision-makers anticipate that their storage administrative headcount will remain the same over the next 12 months (Figure 3) in spite of the fact that storage capacity and storage budgets are increasing. These IT organizations will need to simplify the management of their storage environments (e.g., through storage consolidation initiatives) in order to operate with a headcount that will not grow. This finding is corroborated by a Gartner survey measuring IT spending intentions, which finds flat IT staff budget plans, suggesting that these U.S.-based organizations are still unwilling to significantly bolster the number of IT professionals in the ranks (refer to End Note 8).

**FIGURE 3: MAJORITY EXPECT STORAGE HEADCOUNT TO REMAIN SAME**

In the future (next 12 months), do you anticipate your storage administrative headcount will grow, shrink, or remain the same?



N=127

Source: Data Storage Outlook—Corporate Technologies, Inc.

In comparison, 17% of respondents expect their storage administrative headcount to grow over the next 12 months. Some of the reasons cited for this growth include bringing on new Disaster Recovery centers and new projects. Only 1% expect their storage headcount to shrink.

**CONCLUSIONS: CURRENT AND PROJECTED STORAGE HEADCOUNT**

*Increasing storage budgets coupled with flat storage administrative headcounts mean that companies are becoming more efficient about managing their storage they are finding ways of growing their storage per headcount available. Technology primarily storage consolidation initiatives is the force behind such efficiency improvements. However, these improvements do not suggest that storage management is a solved problem. Until administrative headcounts increase to keep pace with increasing storage budgets, companies will have to continually improve their efficiency it is a never-ending problem.*

## Top Three Storage Concerns

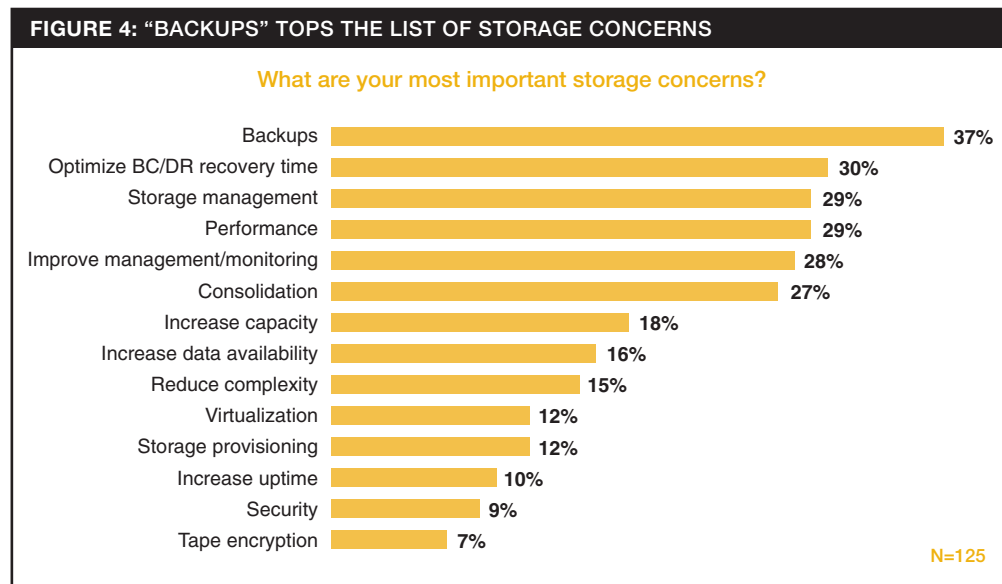
### CURRENT STORAGE CONCERNS

Senior IT decision-makers were asked to identify their current top three storage concerns from a list of fourteen possible issues (Note: Although respondents were given the option of specifying other issues, very few did so.) Their responses are shown below (Figure 4):

- The most frequently mentioned issue in the top three is backups, cited by 37% of the respondents as one of their top three storage concerns.
- The next most frequently mentioned issue is, not surprisingly, optimize business continuity (BC)/ disaster recovery (DR) time, cited by 30% of the respondents as one of their top three storage concerns.
- As expected, storage management and performance are both cited by 29% of the respondents respectively as one of their top three storage concerns.

Next most frequently mentioned, in terms of top three concerns, are improve management monitoring, consolidation, increase capacity, increase data availability, and reduce complexity.

Issues that are not generally important to these IT decision-makers include: virtualization, storage provisioning, increase uptime, security, and tape encryption.

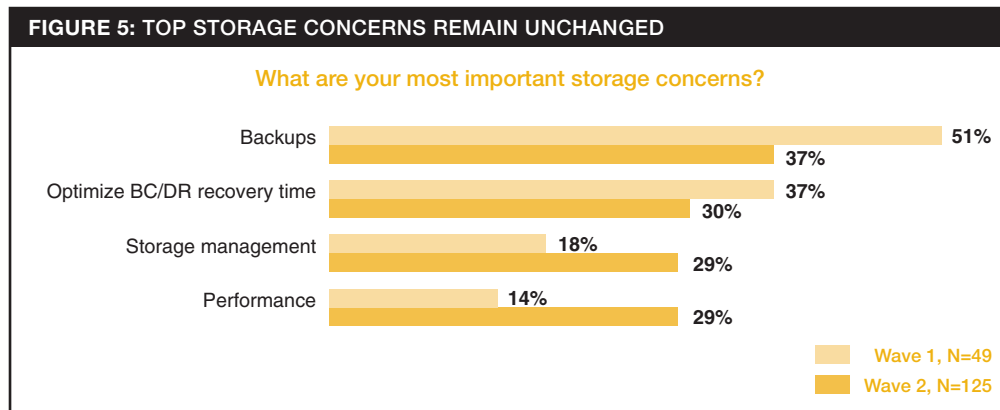


Source: Data Storage Outlook—Corporate Technologies, Inc.

**TRENDS IN STORAGE CONCERNS**

Data Storage Outlook finds that backups have remained the number one storage concern over the term of the research (Figure 5). In Wave 1, more respondents (51%) named backups than any other concern. In Wave 2, although a smaller percentage of respondents (37%) named backups than in Wave 1, it remained higher than any other concern. This finding is consistent with a recent Gartner survey, which reveals that backup and replication software is one of several solutions that will receive the greatest amount of budget during the next one to three years (refer to End Note 5).

Similarly, optimize business continuity/disaster recovery time remained the number two storage concern over the term of the research. In Wave 1, 37% of respondents named it as a top concern, more than all other concerns except backups. In Wave 2, although a smaller percentage of respondents (30%) named optimize business continuance/disaster recovery time than in Wave 1, it remained higher than any other concern except backups. This finding is consistent with a Gartner survey among Small and Midsize Businesses (SMBs), which concludes that disaster recovery is an example of technology areas that have traditionally been too complex and expensive for SMBs but are now affordable and becoming top initiatives in 2007 (refer to End Note 9).



Source: Data Storage Outlook—Corporate Technologies, Inc.

In that same timeframe, performance significantly increased in importance, from 14% (Wave 1) to 29% (Wave 2). Similarly, storage management also increased in importance, from 18% (Wave 1) to 29% (Wave 2) making these two the third most important current concerns among senior IT decision-makers today. Gartner notes that meeting the growing storage performance demands and lack of storage management tools are two of the five top problems IT personnel face when managing storage today (refer to End Note 5).

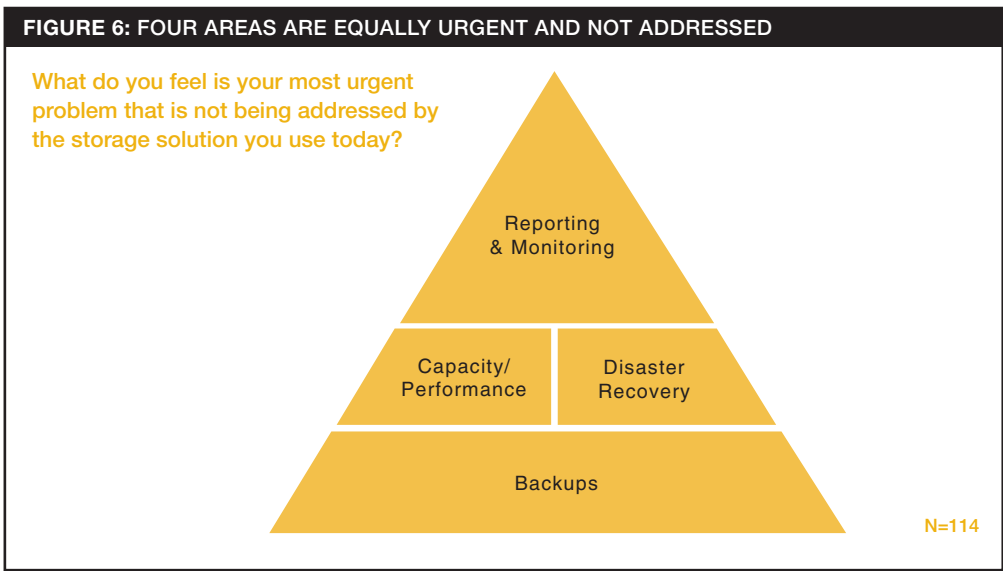
**CONCLUSIONS: TOP THREE STORAGE CONCERNS**

*The top storage concerns identified by respondents remain problematic because successfully addressing them remains problematic. Backups are a never-ending issue driven by the fact that data continues to grow at an unprecedented rate as soon as an organization buys the storage infrastructure to handle current and near-term needs, that growth in capacity requires yet more backup infrastructure (the perpetual backup treadmill). While Disaster Recovery is becoming more common effective DR implementations are not. DR designed to meet both a Recovery Time Objective (hours required to be back live) and a Recovery Point Objective (data loss tolerated) can readily accomplish what a given company needs to achieve, but this apparently is not leveraged by many companies. Performance remains a troubling issue because the gulf between server performance and storage performance is widening. Storage devices are so significantly slower than servers that storage today often represents the performance bottleneck for whole IT infrastructures.*

**Most Urgent Problems Related to Storage**

**CURRENT URGENT PROBLEMS**

Respondents identify four problems that are both *most urgent* and *not being effectively addressed* by their current storage solution (Figure 6) reporting and monitoring , capacity/performance , disaster recovery , and backups.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Reporting and monitoring clearly refers to the lack of automated tools frustrating these respondents they are typically looking for tools that will provide them with a view of multiple products in their environments.

Capacity/performance continues to be a perceived problem because storage environments are growing so fast and managers are without the tools to measure storage utilization accurately. So it is difficult to know if there is a bona fide capacity problem or a reporting/process problem.

Disaster recovery remains a problem because, surprisingly, a majority of those interviewed (52%) report that they are still not replicating data to a DR facility.

The importance of backups is underscored by the number of respondents reporting that they are barely meeting backup windows 39% nightly, and 38% weekly.

#### TRENDS IN URGENT PROBLEMS

Data Storage Outlook finds that the four problems identified above reporting and monitoring , capacity/performance , disaster recovery , and backups were also identified as most urgent in Wave 1 of the research. The percentage of respondents naming these concerns as urgent has remained relatively constant, suggesting that senior IT decision-makers have made little traction in solving these storage problems over the term of the research.

#### **CONCLUSIONS: MOST URGENT PROBLEMS RELATED TO STORAGE**

*Respondents clearly have unused storage available, but it is impractical to reallocate storage from one host to another host in most storage arrays. Although such reallocation is technically feasible, it is highly complex and risky. A better solution would be to seek alternate remedies such as thin provisioning technology, which allows IT decision-makers to defer the cost of storage until users actually need the capacity.*

## Installed Storage Infrastructure

Figure 7 provides an overview of IT decision-makers' vendor preferences for the following products (listed in order of most prevalent in their respective IT environments): disk SAN, disk NAS, tier 2 disk, tape libraries, servers, operating systems, and databases.



Source: Data Storage Outlook—Corporate Technologies, Inc.

**CONCLUSIONS: INSTALLED STORAGE INFRASTRUCTURE**

*It is clear that IT decision-makers have had, and continue to have, a preference for purchasing their storage products from vendors with leading market share and mindshare. At the same time, these respondents also say they prefer to select best-of-breed products and operate multi-vendor environments. It is worth pointing out the disconnect here best-of-breed products are not always leading market share.*

## **Storage Capacity in SAN (Storage Attached Network) vs. NAS (Network Attached Storage) vs. DAS (Direct Attached Storage)**

### **CURRENT AND PROJECTED OVERALL STORAGE CAPACITY**

The majority of senior IT executives surveyed for this study have production IT environments where a combination of Direct Attached Storage (DAS), Network Attached Storage (NAS: refer to End Note 3), and Storage Attached Network (SAN: refer to End Note 4) is in use.

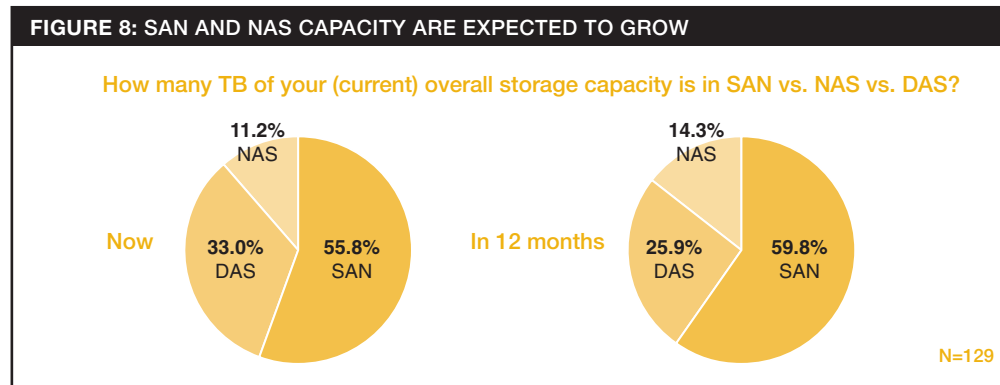
Data Storage Outlook finds that in the next 12 months, respondents project their total storage capacity to increase 26%. This finding is corroborated by a Gartner research survey, which finds 25% storage capacity growth in one year, and 41% growth in two years (refer to End Note 5). Total storage capacity varied considerably by company from less than 2 TB to more than 1 PB.

Key drivers for NAS and SAN growth include traders' needs for increased data; regulatory issues such as SEC, HIPAA, Sarbanes-Oxley (SOX), the U.S. Patriot Act, and others; ecommerce initiatives; implementation of digital repositories; corporate data retention policies; and support of email environments.

Data Storage Outlook is of the opinion that the increase in storage capacity coupled with a flat storage administrative staff is a significant factor driving interest in storage management software.

### CURRENT AND PROJECTED STORAGE PERCENTAGES

Overall, respondents state that the largest percentage of their current storage capacity is in SAN: 55.8% of their current storage capacity is in SAN vs. 33.0% in DAS vs. 11.2% in NAS (Figure 8).



Source: Data Storage Outlook—Corporate Technologies, Inc.

In the next 12 months, both their SAN and NAS capacity are expected to grow as a percentage of total storage, while the DAS share of storage is expected to significantly decline. Overall, these respondents expect to have 59.8% of their total capacity in SAN vs. 25.9% in DAS vs. 14.3% in NAS, in the next 12 months.

### CURRENT AND PROJECTED MEAN STORAGE CAPACITY

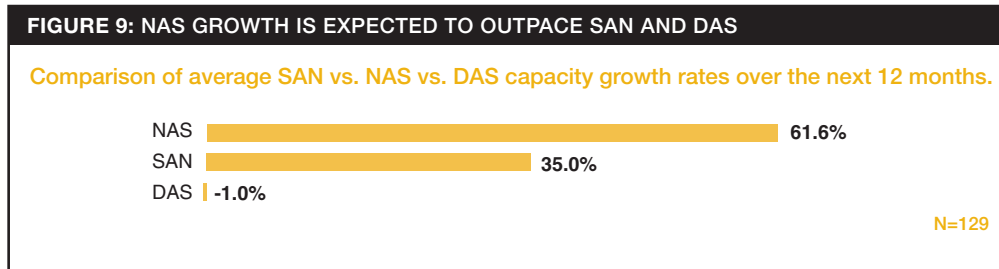
Current mean reported capacity varies by type of storage. Companies report a mean average of 25.27 TB for SAN, 5.10 TB for NAS, and 15.22 TB for DAS. A handful of companies have difficulty calculating their actual storage capacities. (Refer to the Appendix, Figure A, for mean capacities for key vertical markets such as financial services, manufacturing, and life sciences.)

Similarly, the study data shows that projected mean storage capacity varies by type of storage. Companies project a mean average of 34.12 TB for SAN, 8.24 TB for NAS, and 15.06 TB for DAS. (Refer to the Appendix, Figure B, for expected mean storage capacities by key vertical markets in the next 12 months.)

The Data Storage Outlook study reveals many IT decision-makers are in the midst of proactively slowing their DAS growth, while planning for significant growth of their NAS and SAN environments in the next 12 months.

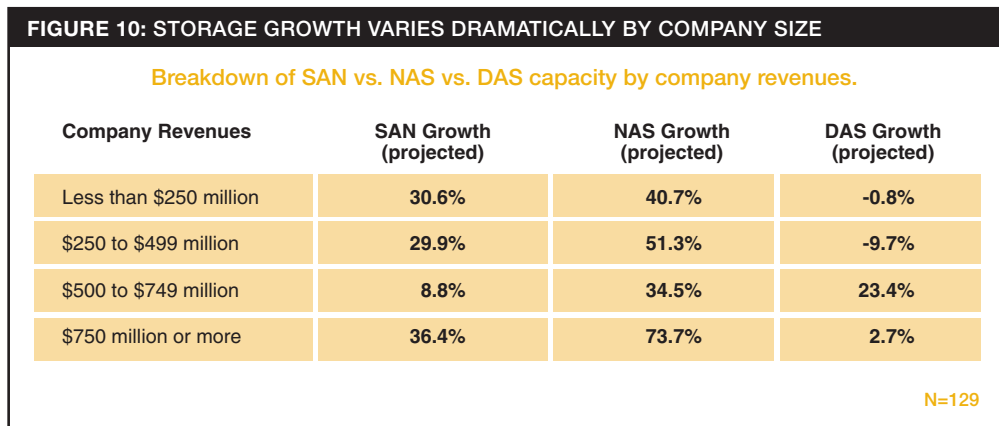
### STORAGE CAPACITY GROWTH RATES

The study shows that, overall, IT decision-makers expect average growth rates for their SAN, NAS, and DAS storage to differ dramatically for the next 12 months. On average, SAN capacity is expected to grow by 35.0%, compared to 61.6% growth for NAS and -1.0% growth for DAS (Figure 9).



Source: Data Storage Outlook—Corporate Technologies, Inc.

Noteworthy, respondents' expectations for their SAN vs. NAS vs. DAS growth vary dramatically by size of company, based on company revenues (Figure 10). Medium businesses (\$500 to \$749 million in revenues) expect significantly stronger DAS growth (23.4%) than companies with revenues higher or lower. NAS growth also appears to correlate with company size, as companies with \$750 million or more in revenues expect more aggressive growth (73.7%) than smaller companies. SAN growth is modestly stronger (36.4%) in companies with \$750 million or more in revenues than in companies smaller.



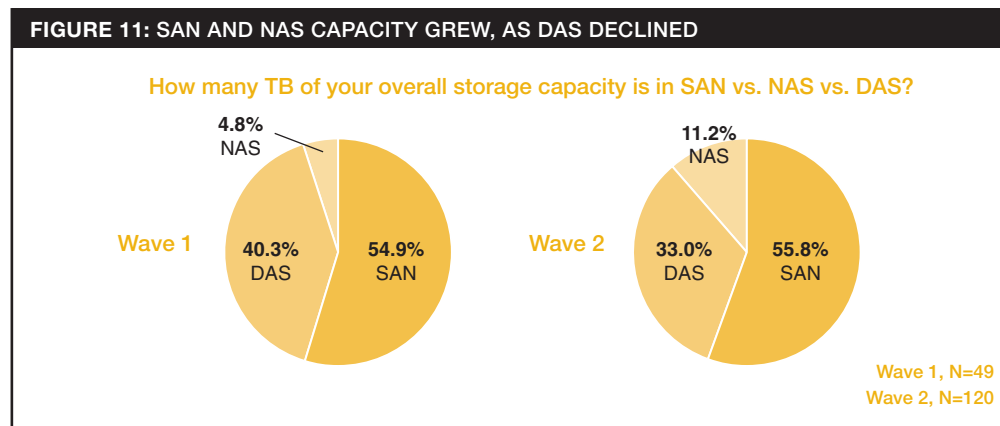
Source: Data Storage Outlook—Corporate Technologies, Inc.

During the course of the survey, Data Storage Outlook spoke to five organizations that are expecting 400%-plus SAN growth and four organizations that are expecting 400%-plus NAS growth. In all cases, the organizations reflected a wide range of both industries and company size.

Data Storage Outlook finds that storage growth rates for SAN vs. NAS vs. DAS vary significantly by key vertical markets, specifically for financial services, manufacturing, and life sciences. (Refer to the Appendix, Figure C, for specific growth rate information for the next 12 months.)

A number of organizations state they are moving away or are interested in moving away from DAS. Instead, they plan to use NAS and/or SAN, in order to reduce cost and improve management efficiencies. IT executives report that, in the past, DAS growth was often a result of the demands of internal groups/departments that wanted to keep control of their data, and made IT decisions on a project-by-project basis.

## TRENDS IN STORAGE PERCENTAGES AND MEANS



Source: Data Storage Outlook—Corporate Technologies, Inc.

Not surprisingly, Data Storage Outlook finds that SAN and NAS percentages of overall storage capacity have increased over the term of the research, while DAS has declined (Figure 11). SAN grew from 54.9% of overall storage capacity (Wave 1) to 55.8% (Wave 2), NAS grew from 4.8% of overall storage capacity (Wave 1) to 11.2% (Wave 2). Over the same term, DAS declined from 40.3% of overall storage capacity (Wave 1) to 33.0% (Wave 2).

### CONCLUSIONS: STORAGE CAPACITY IN SAN VS. NAS VS. DAS

*Overall storage capacity will continue to increase for the foreseeable future for several reasons. Not only is data increasing, but so also is the reluctance to delete data a reluctance that will continue until tools are available that will allow companies to determine data importance. Second, large amounts of data are required for the more robust Business Intelligence tools, which are increasingly being leveraged by companies to gain competitive advantage through superior revenue generation and operational efficiencies. Third, the purchase of storage is becoming more organized and proactive with the emergence of dedicated storage teams focused on addressing their enterprises overall storage strategy rather than on a per-project short-term basis.*

*Data Storage Outlook concludes, based on actual comparison data, that NAS storage capacity will continue to grow faster than SAN, but that SAN will continue to be the most-used storage technology for the foreseeable future. Both of these will increase at the expense of DAS, which is being phased out by the IT decision-makers surveyed. The reason that NAS storage, in particular, is experiencing such significant growth rates is its superiority in handling unstructured data. Data Storage Outlook finds that companies are increasingly taking data from their SAN storage and migrating it to NAS (often as part of a file server consolidation effort).*

**TUTORIAL: THE DATA DELETION CHALLENGE**

*Almost all computer facilities face challenges around data retention and deletion.*

*It takes more time and effort for a user to delete their files than for the user to leave the old files alone. Also, the penalty if the user is wrong about a deletion for example if they delete a file they end up needing later is quite large. Some files are not recoverable, while most require a restoration request and a long elapsed time before the file is again available. The net result, when multiplied by all users of a computing facility, is constant growth of disk space use.*

*Similar challenges are involved in disk space used by application data, and frequently this data dwarfs user data in terms of space used. One of the hardest problems is often simply figuring out who owns the data which becomes complicated when that person is no longer employed by the company. An error in deletion of old application data could result in the application not working or not producing correct results, and a large effort to recover from the problem. This problem is magnified by the propensity of application administrators to keep multiple copies and versions of the application data online. Typically these copies are used for fast restoration in case of problems, QA testing of changes, or development of new application features.*

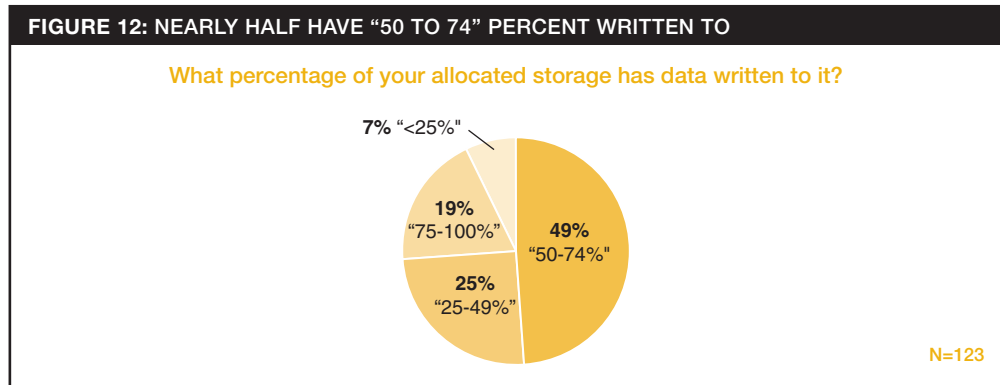
*Adding to the general malaise surrounding deletion is the confusion caused by regulations. Federal and State laws about data retention can be difficult to understand and implement. Company SLAs to their customers can place requirements on retention as well. Corporate legal teams must understand all of these requirements and advise IT as to what must be deleted, what must be saved, and the time periods of these activities per data type. For example a company might force deletion of email more than thirty days old while requiring accounting records to be kept for seven years (but deleted as quickly thereafter as possible).*

*Deciding which data to retain versus delete remains an inefficient process it is often less costly to simply add additional disk space than to figure out which data can safely be deleted.*

## Utilization of Storage

### ALLOCATED STORAGE CURRENTLY WRITTEN TO

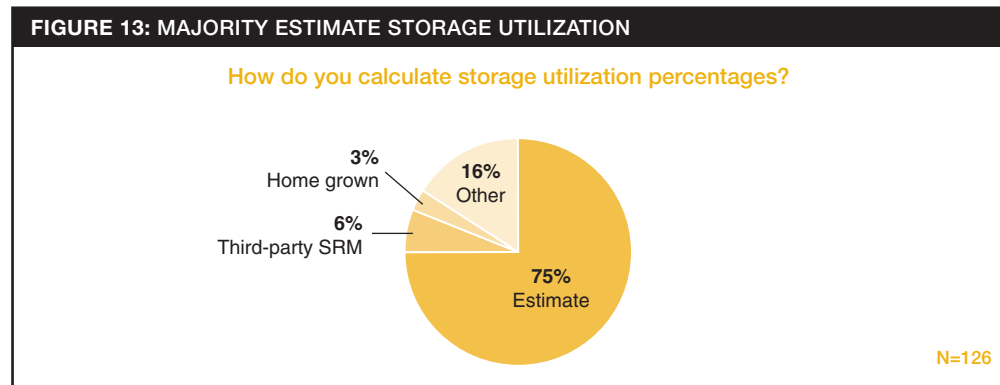
Nearly half of respondents 49% have 50 to 74 percent of their allocated storage already written to (Figure 12). In comparison, 19% of respondents estimate 75 to 100 percent of their allocated storage has been written to, and another 32% of respondents say 49 percent or less has been written to.



Source: Data Storage Outlook—Corporate Technologies, Inc.

### CURRENT CALCULATION OF STORAGE UTILIZATION

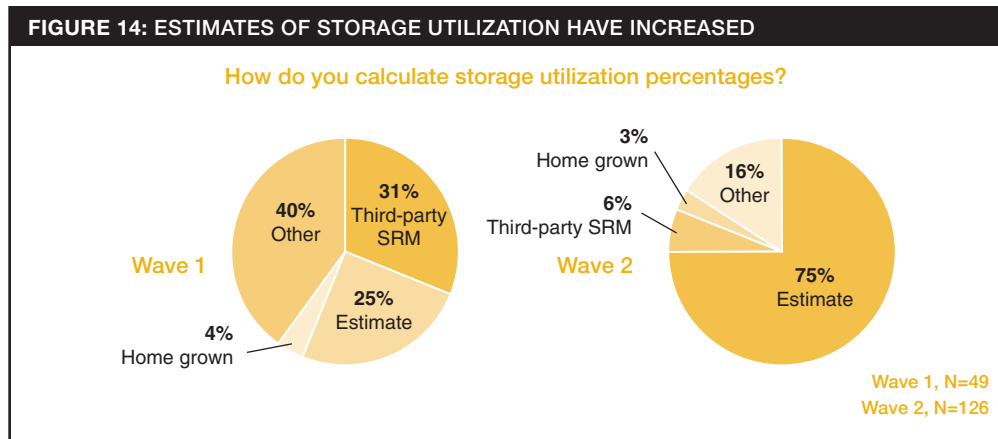
When asked how they calculate their storage utilization percentages, the overwhelming majority of respondents (75%) do so via estimates because they lack automated tools (Figure 13). Only 6% report that they use a third-party SRM tool, and 3% report using a home-grown tool. Another 16% report other, which includes tools from a variety of manufacturers most receiving few mentions. Many of these IT decision-makers are under the false impression that their existing storage products offer them access to these automated tools. A *Storage* magazine survey of purchase intentions finds that the most-wanted storage tools are SRM, performance management, operations management, and configuration management (refer to End Note 7).



Source: Data Storage Outlook—Corporate Technologies, Inc.

**TRENDS IN STORAGE UTILIZATION**

Data Storage Outlook finds that the mean percentage of storage written to has changed little across the two waves of research. In Wave 1, respondents reported that, on average, 59.9% of their allocated storage had been written to. In Wave 2, respondents report that 56.9% had been written to, a decrease of 5.0%.



Source: Data Storage Outlook—Corporate Technologies, Inc.

The percentage of respondents relying on estimates to determine storage utilization has increased substantially over the two waves of research (Figure 14). In Wave 1, only 25% of respondents relied on estimates. In Wave 2, 75% of those interviewed said that estimates were their primary means of determining utilization percentages. This data probably does not indicate that three times the number of individuals are currently relying on estimates than in the first wave of research. It is rather most likely an indication that with the growing focus on storage, senior management is increasingly aware that utilization estimates are just that estimates.

**CONCLUSIONS: UTILIZATION OF STORAGE**

*There are many opportunities for realizing significant cost savings by maximizing storage utilization. Adoption of third-party SRM tools, which most of the IT decision-makers surveyed acknowledge they do not have, would allow storage utilization to be measured accurately. Identifying storage capacity that may be lying idle and using floor space would significantly boost utilization. Identifying situations where companies may be purchasing more storage than they need because they mistakenly estimate 100% of their storage has been written to would yield significant cost savings. There is also an opportunity to address allocation issues with thin provisioning technology a strategy that helps organizations identify the optimal time to buy storage capacity and thus lowers the cost of shared storage. Refer also to the tutorial: Thin Provisioning Technology.*

## THIN PROVISIONING TECHNOLOGY

The IT executives surveyed in Data Storage Outlook say their customers IT end users are traditionally not accurate in estimating their future storage requirements. As a result, IT departments may order storage capacity before it is needed, and thus spend IT dollars now when it would be more cost-effective to delay such purchases. Thin provisioning is a technology for deferring storage costs, and taking advantage of the declining costs of storage. In addition, it is also a strategy to reduce costs by eliminating the purchase of unneeded disks. A Gartner survey predicts that newer technologies that help users manage storage capacity growth and increase utilization such as thin provisioning will continue to see healthy growth (refer to End Note 5).

### **TUTORIAL:** THIN PROVISIONING TECHNOLOGY

*Thin provisioning can be utilized in either a SAN- or NAS-based storage model. For example, there might be four users who each request 1 TB of disk capacity (for a total of 4 TB), but they only utilize a total of 1.8 TB. This means that 2.2 TB of capacity is allocated, but not utilized. With thin provisioning you allocate the 4 TB logically, but physical capacity is allocated only when it is written to. The host sees the full logically allocated capacity. This allows IT to purchase storage capacity as it is required, so storage purchasing is more efficient.*

*The Data Storage Outlook survey reveals that respondents typically use only a portion of their current storage capacity (e.g., 30-40%). This means that IT organizations often spend unnecessary dollars to spin empty disks and cool excess storage capacity in their data centers. With the thin provisioning technology, IT decision-makers can defer the cost of storage until the user needs the capacity. (Note: Most IT organizations interviewed say they want to run at 70%-plus utilization rates for storage, but want to avoid running at 100%.) Thin provisioning enables IT organizations to meet end users' current storage needs (e.g., .5 TB) and reserve IT budget for the other .5 TB of capacity until the user needs this capacity. The delayed storage purchase decision is transparent to the end users. With thin provisioning, the IT organization decides on the optimal time to buy the additional capacity and thus lowers the cost of shared storage.*

*The use of thin provisioning technology has grown significantly since we discussed it in the first wave of Data Storage Outlook. The technology is offered by several vendors with many others promising it is on the way. One key to successful use of thin provisioning in a production environment is system monitoring. The storage device should provide alerts to ensure you never run out of physical capacity. Nobody wants to explain to a user that the application is down because the disk space they paid for does not exist yet. This is an area where the vendor solutions are quite different. The thin provisioning monitoring options range from a log entry to an automatic phone home service event. Be sure you ask your vendor how their implementation ensures your production environment will run smoothly and without interruption.*

## Information Lifecycle Management (ILM)

### CURRENT ILM ADOPTION

The Data Storage Outlook study finds that the concept of Information Lifecycle Management (ILM) is not widely understood by IT professionals. When asked about current use of ILM technology, 86% of respondents say they do not currently use it. Among these storage professionals, slightly more are either evaluating ILM technology or plan to evaluate it within 12 months (54%), than those who are planning to not evaluate ILM technology (46%). Similarly, a recent Gartner survey among global SMBs, finds that almost half of the respondents have never heard of ILM and less than 30% plan to implement ILM (refer to End Note 9).

### TRENDS IN ILM ADOPTION

Data Storage Outlook finds that ILM usage and planned usage is increasing over the term of the research. Respondents using ILM rose from 5% (Wave 1) to 14% (Wave 2). However, respondents evaluating ILM currently, within six months, or within twelve months declined from 53% (Wave 1) to 46% (Wave 2), indicating that future usage of ILM will be likely to grow, but slowly.

### CONCLUSIONS: INFORMATION LIFECYCLE MANAGEMENT (ILM)

*Many of the IT decision-makers interviewed for Data Storage Outlook asked the question: At what capacity does Information Lifecycle Management make sense for my organization? This is a very subjective question. If your decision is between adding 1 TB of expensive fibre channel disk per year versus building and managing an ILM infrastructure, it will be more cost-effective from both an operational and capital perspective to buy the disk. At the other extreme, if you are adding 20 TB per year of data and it is only accessed frequently for the first three months, then it is very cost-effective to build the ILM infrastructure and move the older data off to less expensive disk or tape.*

*These examples leave a large grey area, as this is not an exact science. There are many issues to consider including: rate of data growth, frequency of access, and staff expertise. Each organization is different and these factors need to be weighed to determine the best solution for your organization. One area in which many managers find ILM to be a win is when it is application-specific. For example, email ILM is one rapidly growing area. See also the tutorial: Virtual Tape Library (VTL) Technology.*

**TUTORIAL:** INFORMATION LIFECYCLE MANAGEMENT (ILM)

*Information Lifecycle Management (ILM) refers to setting policies for which type of data should reside on which type of storage, and using tools (either manual or automatic) to move data between those storage types. Other names given to this methodology are HSM (Hierarchical Storage Management) and Data Storage Management (DSM).*

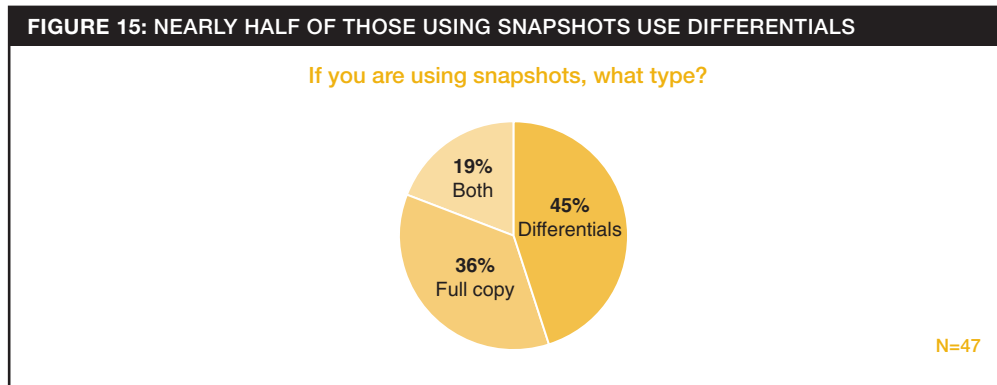
*Throughout its life, data requires different levels of performance, availability, protection, migration, retention, and disposal. At first, fast read and write speeds simplify editing and sharing. The data is also critical to daily operations and therefore demands high levels of availability and protection. Most data becomes less important in time, and is used less frequently. As these changes occur, data can be moved to different tiers of storage that provide appropriate levels of availability, capacity, cost, performance, and protection. These tiers frequently include a variety of technologies such as fibre channel disk, ATA, SATA, or FATA disk, and tape. The complexity comes in moving the data between the tiers transparently to the user community and integrating the ILM system into the existing disk and backup environments. There are several commercial solutions to this problem, but storage managers need to weigh the cost, complexity (in implementation, integration, management, and maintenance), and performance against the gain in resulting storage efficiency before determining whether ILM is a good fit in their environment.*

*The high cost of primary storage is a constant concern and ILM appears to be a way to help control this cost. However, the cost equation is more complex than it appears. First, the cost and complexity of deploying and managing an ILM solution is often significant. Also, the price delta between primary and secondary disk is shrinking. If adding secondary storage means adding a new storage array, the cost savings is reduced ever further. Lastly, the impact to the backup infrastructure needs to be measured. ILM is a very compelling technology solution, but often the capacity required to make the cost justification is more substantial than expected.*

## Availability Requirements

### CURRENT STORAGE ARRAYS: SNAPSHOT CAPABILITY

Figure 15 shows that among respondents saying their storage arrays currently offer snapshot capability, nearly half (45%) use differentials (also see *the tutorial: RAID, Snapshots, and Filesystems*). An additional 36% use full copy, and 19% report using both full copy and differentials.



Source: Data Storage Outlook—Corporate Technologies, Inc.

However, it appears that many IT decision-makers are unclear about the difference between differentials and full copy.

Among those using snapshots, most-mentioned vendors are NetApp, EMC, Microsoft, and HP.

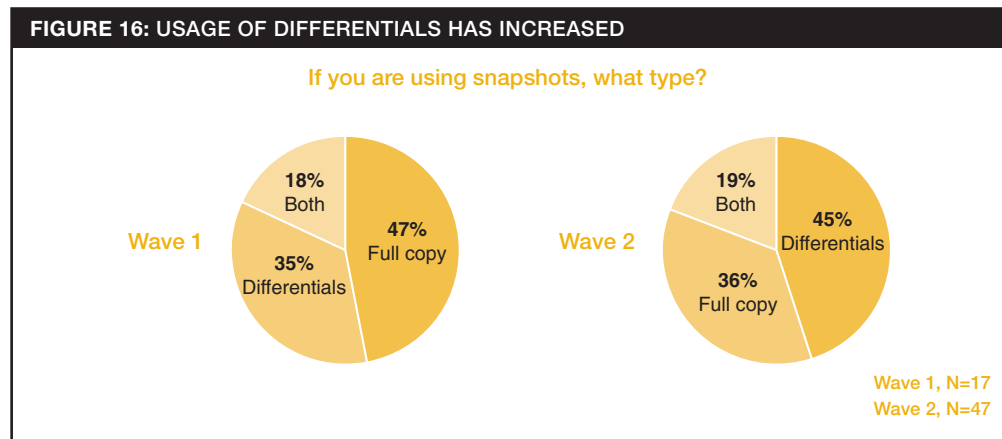
### CURRENT CONTINUOUS DATA PROTECTION (CDP)

Just over 6% of respondents indicate they are using CDP as an availability technology. The majority (63%) indicate they are using it for databases, while others (13%) indicate its use for databases and email.

No CDP vendor received more than one mention.

## TRENDS IN AVAILABILITY REQUIREMENTS

Snapshot preferences have changed over the term of the research, from full copy to differentials (Figure 16). Full copy mentions declined from 47% (Wave 1) to 36% (Wave 2), while differentials increased from 35% (Wave 1) to 45% (Wave 2). Overall, specific vendor technology reported remained relatively unchanged.



Source: Data Storage Outlook—Corporate Technologies, Inc.

CDP adoption remained extremely low across both waves of the research. However, Gartner predicts that CDP and remote replication solutions will continue to rise in their significance, as businesses move to protect their data (refer to End Note 5). In keeping with that prediction, a *Storage* magazine survey finds that 22% of respondents plan to implement CDP and 45% plan to evaluate it (refer to End Note 7).

**CONCLUSIONS: AVAILABILITY REQUIREMENTS**

*There are significant tradeoffs to be made when assessing if snapshots or Continuous Data Protection is a better storage solution. The huge advantage of CDP is that it reduces the Recovery Point Objective to zero, by offering infinite granularity of recovery if it is necessary to go back in time. Its current disadvantage is that CDP technology is still evolving and no single vendor has established a market-leading position making the introduction of such technology into one's core storage infrastructure high-risk. Most companies can tolerate a 15-minute RPO, which makes snapshots a superior alternative in terms of risk, ease-of-use, and cost-effectiveness. To maximize the benefits afforded by snapshots, companies should consider only technology that provides very efficient snapshots as an integral part of the technology rather than as an add-on.*

## DATA PROTECTION TECHNOLOGIES

RAID, snapshots, and filesystems are very different technologies, but they sometimes get grouped together. Unfortunately, their implementation also varies quite a bit. The following tutorial looks at several emerging data protection technologies, pointing out the implementation differences within each.

### **TUTORIAL:** RAID, SNAPSHOTS, AND FILESYSTEMS

*What is new in RAID, snapshots, and filesystems?*

*Often filesystems and RAID are thought of as solved problems areas of technology that do not need further development. In fact there are several emerging data protection technologies that were not generally available at the time of the first wave of DSO.*

*The first is RAID-6. There are several vendors shipping RAID-6 technology with many more promising it is on the way. While they differ slightly, there is one common theme. They use two parity disks in a RAID set to protect the data similar to the way RAID-5 uses a single disk. This allows the storage system to protect the data even in the scenario where two disks in the same RAID set are lost. Data would be lost if this RAID set were protected by RAID-5. As disks continue to increase capacity faster than they increase performance, rebuild times are increasing and this additional data protection is becoming more and more important.*

*Unfortunately, most RAID systems cannot protect data from some all-too-common scenarios. It is an unfortunate fact of life that disk firmware and RAID controllers have bugs. Let's examine a couple of scenarios. First, a disk drive receives a command to write a block of data, then it discards the data and acknowledges a successful write. When that location is read again, the old invalid data will be returned and the read will be considered a success. This is affectionately referred to as *Don't write. Don't tell.* What happens when a block is read from one half of the mirror and it does not match the same block on the other half of the mirror? Most RAID solutions cannot figure out which data is good and the data block is considered corrupt. Solutions are starting to emerge to address these challenges. They are able to determine which half of the mirror contains the valid data and know that an old data block is old.*

*Mirroring (technically, RAID-1) makes one or more copies of a filesystem (or volume, etc.) and keeps it synchronized with the original copy. While most technologies support more than one copy, most sites run with just one copy. The initialization of mirroring requires that a full copy be made (so mirroring requires at least 2X the amount of the original disk). When an update is made to the original, the update is mirrored to the copies. The benefit of a mirror is that if the original copy fails, the data is instantly, automatically and fully available on the mirror. Once the fault is resolved, the good mirror can be copied to the original storage and the full mirror functionally is then restored, all with zero data loss.*

**(CONTINUED)****TUTORIAL: RAID, SNAPSHOTS, AND FILESYSTEMS**

*Mirrors have other uses aside from just data protection. For example, a mirror could be split, causing the automatic updates to be terminated or paused. At that point two identical copies of the data exist. One copy could be used for an upgrade, or as the source of a backup, while the other continues in production use. Once the split-mirror work is done the mirrors can be resynchronized for full RAID protection.*

*Snapshots (called various names by various companies including block differential and virtual copy) differ in that they are not a RAID level. They do not offer protection from source disk failure. Rather, they are a periodic log of the blocks that change on the original storage. In effect, if a snapshot of a filesystem is taken, it is an image of that filesystem at the time of the snapshot. As time passes after that snapshot, the original storage will be updated via writes, but the snapshot will not change, continuing to show the storage as it was at the time the snap was taken. The implementation of snapshots varies wildly, from very efficient (metadata updates) to less efficient (copy-on-write of each original block before it is updated at the source). Some systems support read/write snapshots (the snapshot is modifiable) while most only allow read-only access. Typically, many snapshots can be taken and stored (say, hourly) while using only some small percentage of disk more than the original storage (10-20% is common). Thus it is efficient in disk use. Each snapshot can be used to restore data that was recently improperly modified, restore files that were deleted, undo changes, and even reset entire filesystems to known-good states (say, before an unfortunate upgrade was attempted).*

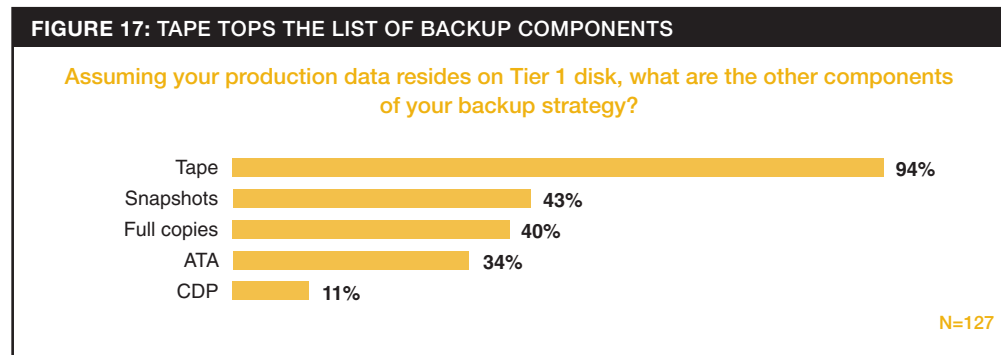
*Why are snapshots useful? There are many uses for it:*

- *By having the snapshot copy mounted, a user could move back in time by accessing the snapshot copy rather than the original filesystem. For example, if the system created a snapshot of a filesystem every midnight, a user could look at the system as it was yesterday, before he deleted that important file. He could then copy the file from yesterday to the current filesystem to restore its use.*
- *By allowing multiple snapshots of the same filesystem, it is possible to have views of the filesystem as it was every day in the past week, or every four hours in the past day, and so on.*
- *By quiescing a database or other important application, taking a snapshot, and continuing the application, a reliable, consistent backup could be made of data that is constantly changing. The snapshot copy is a view of the world at a time when the data was on disk and not changing. The application can continue running, users can continue their work, and the backup can be done without interfering with their activities.*

## Backup Requirements and Processes

### CURRENT COMPONENTS OF BACKUP STRATEGY

Respondents were asked what components comprise their backup strategy, assuming their production data resides on Tier 1 disk. The survey reveals that 94% of respondents mention tape (Figure 17). This finding is consistent with a *Storage* magazine survey, which finds tape still working alongside disk and still a key to backup and DR (refer to End Note 7).

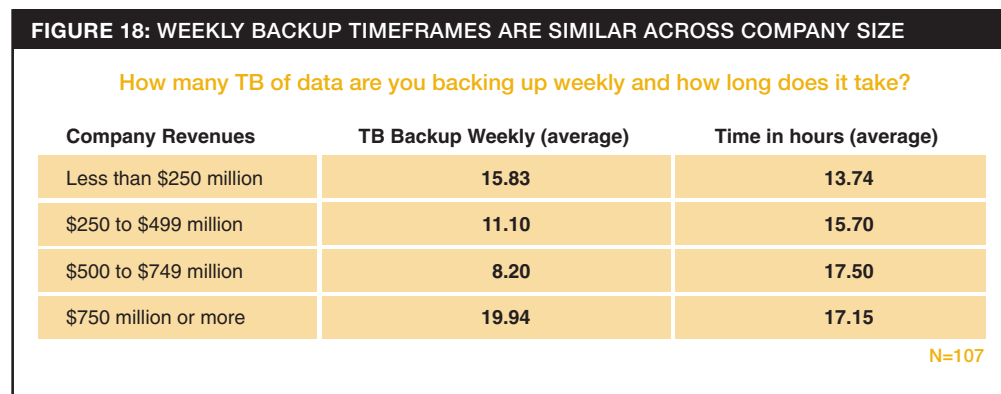


Source: Data Storage Outlook—Corporate Technologies, Inc.

Data Storage Outlook participants also mention, on average, one to two other technologies, suggesting they are finding additional ways to protect data. Some 43% of respondents are using point-in-time copies (snapshots), 40% are using full copies, and 34% are using 2nd tier disk (ATA). Only 11% report they are using Continuous Data Protection (CDP).

### CURRENT NUMBER OF TB BACKED UP

Data Storage Outlook finds the amount of data being backed up weekly varies by company size (Figure 18). The largest companies, those with \$750 million or more in revenues, have the largest amount of data to backup weekly: on average, 19.94 TB of data.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Overall, the majority of respondents (57%) report backing up less than 10 TB of data weekly. An additional 25% report backing up 11 to 25 TB, 9% report backing up 26 to 50 TB, and 9% report backing up more than 50 TB of data weekly.

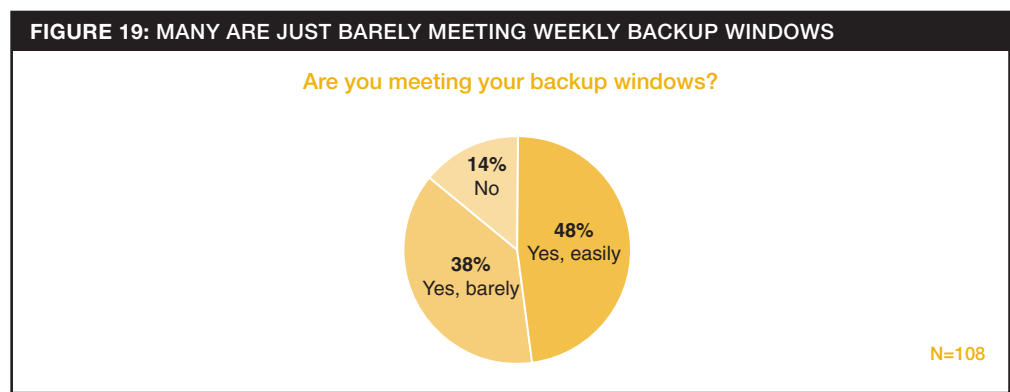
### CURRENT TIME REQUIRED TO BACKUP

In general, it is typical for companies to report performing full backups over the weekend. Just over one-third of respondents report the number of hours it takes them to do weekly backups as either less than 12 hours (36%) or more than 20 hours (36%).

The average timeframe to conduct weekly backups does not vary dramatically by company revenue size (Figure 18). Data Storage Outlook finds that it takes companies with revenues in the \$500 to \$749 million range 17.50 hours (on average) to backup 8.20 TB of data, as compared to the largest companies that take 17.15 hours (on average) to backup more than two times as much data (19.94 TB).

### CURRENT BACKUP WINDOWS MET

In total, 86% of the IT decision-makers say they are meeting their weekly backup windows (Figure 19). However, 52% of the respondents are barely meeting this backup window, or not meeting it at all. There is clearly an opportunity to implement some proactive strategies, such as adding servers, adding new tape drives and libraries, and optimizing and tuning the backup infrastructure and to do so in a sufficiently decisive manner so the backup treadmill will not reappear a few months later in the face of dramatically increasing data loads.



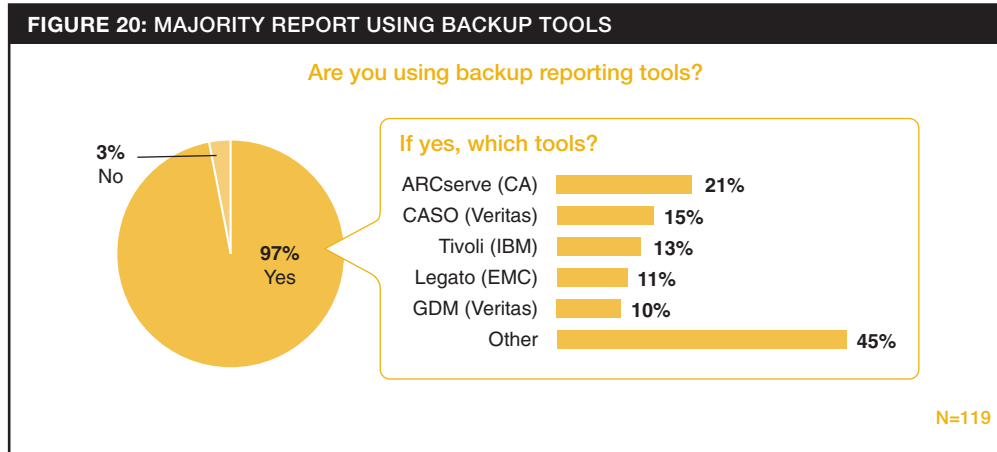
Source: Data Storage Outlook—Corporate Technologies, Inc.

### CURRENT PERCENT OF BACKUPS FAILING

Data Storage Outlook finds that, on average, just over 4% of backups fail 73% of respondents report that 1 to 5 percent of backups fail. An additional 10% of respondents report that 0 percent of backups fail, and 2% report that more than 20 percent of backups fail.

**CURRENT BACKUP REPORTING TOOLS USED**

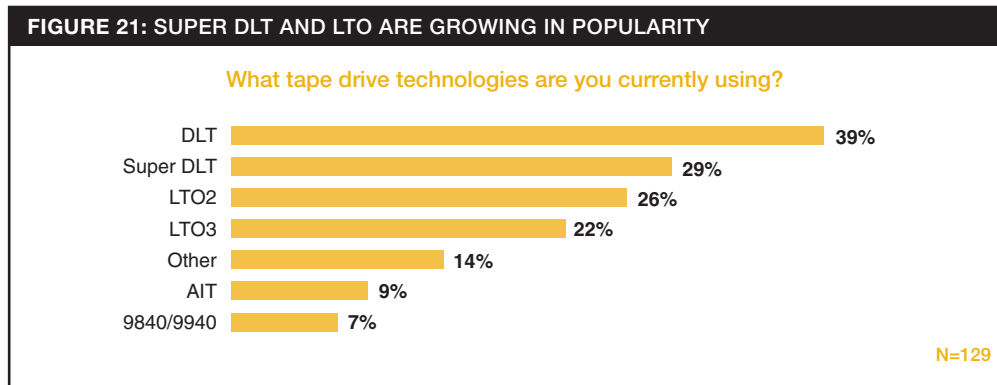
The survey indicates that the great majority of respondents are using backup reporting tools 97% report such tools are in place, while only 3% say they do not use these tools (Figure 20). Tools most mentioned are ARCserve (Computer Associates), Centralized Administrative Server Option (Veritas), Tivoli Storage Manager (IBM), Legato NetWorker (EMC), and Global Data Manager (Veritas). Other tools were reported by 45% of respondents, but most were single-mentions.



Source: Data Storage Outlook—Corporate Technologies, Inc.

**CURRENT TAPE DRIVE TECHNOLOGIES USED**

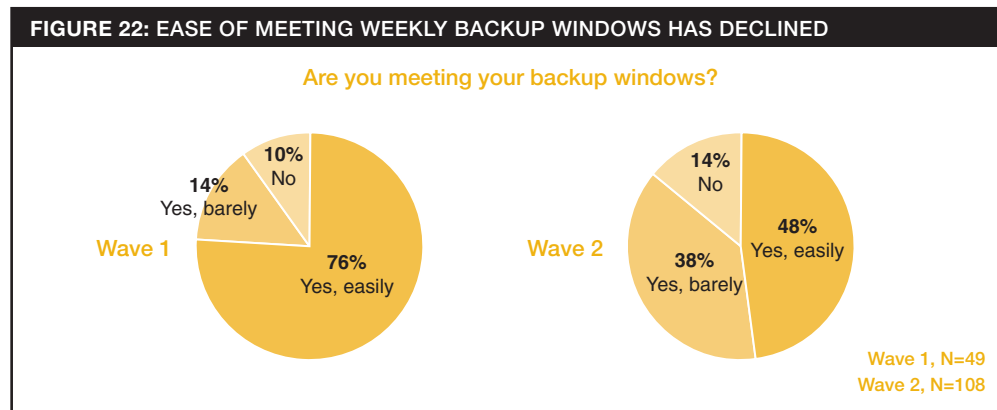
Most companies report use of multiple tape drive technologies (Figure 21). Overall, 39% of respondents currently use DLT tape drives. It is clear that companies have been expanding their use of LTO tape drives, and in many cases this means phasing out DLT. The survey also finds that 29% of respondents currently use Super DLT tape drives, 26% use LTO2 tape drives, and 22% use LTO3 tape drives. Only 9% of respondents use AIT and 7% use 9840/9940. The other category includes drive technologies such as 4 mm, DAT, and 9-track.



Source: Data Storage Outlook—Corporate Technologies, Inc.

**TRENDS IN BACKUP REQUIREMENTS AND PROCESSES**

The ease with which backup windows are being met has declined across the two waves of the research (Figure 22). The percentage of respondents reporting they were barely meeting their weekly backup window, or not meeting it at all, increased from 24% (Wave 1) to 52% (Wave 2). Conversely, the percentage of respondents reporting they were easily meeting their weekly backup windows decreased from 76% (Wave 1) to 48% (Wave 2).



Source: Data Storage Outlook—Corporate Technologies, Inc.

Respondents have shifted their mix of tape drive technologies across the two waves of research. Between Wave 1 and Wave 2, usage of DLT, Super DLT, LTO2 and AIT declined by 17.1% to 36.6%. Only LTO3 increased, from 14% of respondents (Wave 1) to 22% of respondents (Wave 2) an increase of 57.1%. These IT decision-makers continue to rely on multiple tape drive technologies, although the number of technologies employed appears to be lessening: 1.63 mean mentions were reported in Wave 1 versus 1.46 in Wave 2.

**CONCLUSIONS: BACKUP REQUIREMENTS AND PROCESSES**

*Companies will always run on the edge of meeting their backup windows because of the backup treadmill dilemma, where escalating data requires escalating storage, which is never sufficient for more escalating data. Clearly the overall trend in backup strategy is towards disk as the first primary restore, and tape as the write-once read-never component for DR. Data Storage Outlook finds it interesting that only 94% of respondents report tape usage, and hopes that the remaining 6% are replicating to a remote DR facility rather than not backing up. Data Storage Outlook is also of the opinion that respondents indicating usage of backup reporting tools are, in fact, using backup tools with minimal reporting capability. Given the tremendous insight afforded by competent reporting tools, it is recommended that companies acquire such tools or obtain reporting modules for their existing backup tools.*

## TAPES, BACKUPS, AND RESTORES

Backup and restores are crucial functions to the integrity of the data at a site. Yet once an implementation is in place, these functions tend to be ignored. Backups are certainly not glamorous, and no one likes to fuss with them. The risk/reward ratio on the backup and restore process is a deterrent as well. If, for example, someone tweaks the backups to make them run faster or smoother, or makes the restores easier or faster, then the rewards tend to be small. On the other hand, there could be significant negative impact to the company if that tweaking caused restores to fail.

The following tutorial provides a discussion of the state of backups and restores (B&R), and a discussion of some current technologies that may make the reward better than the risk for updating the B&R at your site.

### **TUTORIAL:** THE TRUTH ABOUT TAPES, BACKUPS, AND RESTORES

*Here are a few data points to have handy when making backup decisions at your site:*

- *What tape technology do you use?*
- *What is the capacity of the tape, both raw and compressed (if you use compression)?  
What compression rate do you reach?*
- *Do you use disk as a staging storage for tape (i.e., nearline storage or VTL)?*
- *How many tapes do you buy a year, and at what cost?*
- *If you use a library, how many full and incremental tape sets do you keep loaded?*
- *How many tape sets do you send off-site?*
- *Do you encrypt the data on disk, or tape, or both?*

*The next step in evaluating your backups and restores (B&R) methodology is to use a backup calculator. Some vendors provide them if you are using or evaluating their technology. This type of calculator is often flexible about accepting quite a lot of input data, and allows what-if scenarios to be evaluated.*

*The truth about tape: upgrading your tape drives can save money.*

*Tape media costs money, and older media costs about the same as new media but stores far fewer bits. For instance, if you have DLT8000 drives, enter the appropriate values in a calculator to see how many LTO3 tapes and drives you would need. For the most complete picture, determine how many tapes you are using in a time period (three years, for example), and how many you buy in that time period. Then determine how many tapes you would need using a newer tape technology. Multiply by the cost of the tapes, and you*

**(CONTINUED)**

**TUTORIAL:** THE TRUTH ABOUT TAPES, BACKUPS, AND RESTORES

*may find that over that time period, even including the cost of the new drives and possibly even a new robot your costs would be lower than sticking with the older, lower-density tapes.*

*The truth about tape: compression rates vary wildly depending on the data being moved to tape.*

*Unfortunately, compression rates are hazardous to consider in capacity and performance calculations because they are highly affected by the data being compressed. Some data is non-compressible (consider audio recordings that are already compressed), while some are highly compressible (such as clear text). The safest bet, if you have mixed data or do not know the data compressibility, is to assume zero compression. This is the worst case, so any compression of the data would be good news, and would add extra capacity and performance to your solution.*

*The truth about tape: encryption has a cost.*

*In fact it has multiple costs. When done in software, encryption can cause backup throughput to decrease by 50%. Some backup encryption solutions are hardware-based and do not share this problem. Both hardware and software encryption can undo the good work of compression and increase the amount of data that needs to be written to tape.*

*The truth about tape: leading-edge firms are reducing the use of tape.*

*In truth, no one likes tape. Many companies are using disk-to-disk-to-tape backups.*

*D-to-D-to-T solutions are those in which incremental backups go to local disk (but not the same disks as it is backing up!). Some are going further and doing disk-to-disk-to-disk-to-tape backups, in which a production site does disk-to-disk backups, data is replicated to a disaster recovery site, and finally the data is written to tape at that DR site. The basic model is that if tape is not being taken off-site (say, incrementals) then just write it to disk.*

*Only use tape as an off-site transport, write-once and read-hopefully-never (during a disaster).*

*In conclusion, backups and restores are complicated, fraught with risk, and difficult to get excited about. However, there are rewards for evaluating the current solution at your site and considering upgrades of various kinds. Backup calculators can help guide your selection, but there are truths about B&R technology that should also be considered. Avoiding tape except for off-site and disaster recovery solves many problems and is a leading-edge solution.*

## Virtual Tape Libraries (VTL)

### CURRENT VTL ADOPTION

Data Storage Outlook finds that VTL adoption in the surveyed storage environments is low. Overall, 80% of respondents report that VTL is not incorporated. Of those, 26% say that they plan to implement VTL, while 74% of them report that it is not planned to be incorporated in the foreseeable future. A recent Gartner survey finds a lower percentage planning to implement VTL in the future 16% in the next two years (refer to End Note 5).

Of the 20% of respondents saying that VTL is a part of their storage environment, the most-mentioned vendors are EMC, followed by Quantum. Although *Storage* magazine finds that a higher percentage has implemented VTL 30% no usage trend is as yet emerging (refer to End Note 7).

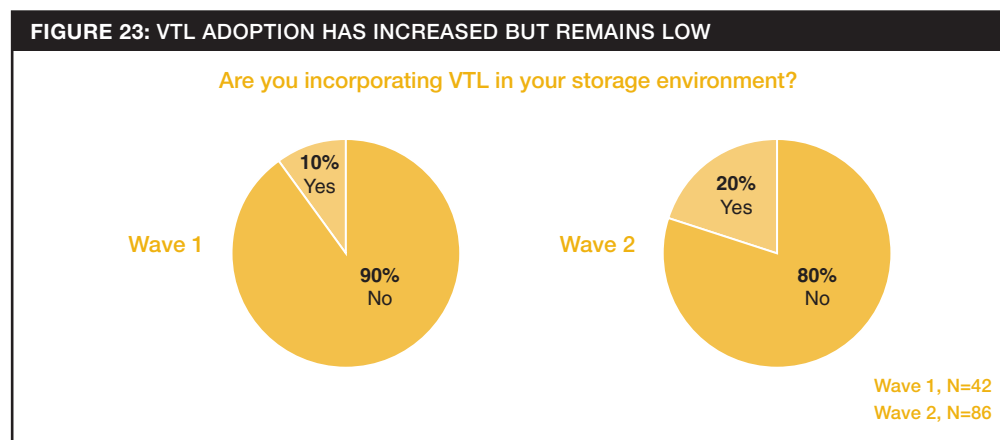
Storage capacity dedicated to VTL averages 10.54 TB, but tends to be higher in larger companies 15.00 TB (\$500 to \$749 million) and 16.60 TB (over \$750 million).

Among respondents using VTL, the great majority (82%) report that they are not using deduplication technology.

*See also the tutorial: Virtual Tape Library (VTL) Technology.*

### TRENDS IN VTL ADOPTION

Data Storage Outlook finds that the adoption of VTL has increased over the term of the research (Figure 23). In Wave 1, 10% of respondents reported that they were incorporating VTL in their storage environment versus 20% in Wave 2.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Although the percentage of those planning to adopt VTL has increased, the percentages remain low. In Wave 1, 22% said they planned to adopt VTL, while in Wave 2, 26% said adoption would be likely. The number reporting that VTL adoption would be not in the foreseeable future declined slightly from 78% (Wave 1) to 74% (Wave 2).

### **CONCLUSIONS: VIRTUAL TAPE LIBRARIES (VTL)**

*VTL today is primarily deployed in small companies, or as a point solution in large core data centers. Its appeal lies in the approach VTL devices use to address the backup treadmill problem a tape interface to a disk-based system. The biggest challenge to widespread adoption of VTL technology is that there is as yet no mainstream VTL vendor with market-proven product. A second challenge to widespread adoption has to do with the fact that many VTL devices currently have limited performance and use large amounts of disk space. However, VTL is a rapidly evolving technology and significant improvements are probable in the near future.*

### **TUTORIAL: VIRTUAL TAPE LIBRARY (VTL) TECHNOLOGY**

*Backups are a never-ending IT challenge. As long as data footprints continue to expand, the issue will not go away. Get the backups under control this quarter and by next year, the growth in storage will again be pushing the backup windows to the limit. Virtual Tape Library (VTL) technology offers a potential solution to this challenge. VTL devices offer a tape interface to a disk-based system, which allows for easy deployment into an existing backup infrastructure.*

*The VTL vendors would have us believe it is the Holy Grail of backup solutions. Deploy a VTL and your backup problems will be solved forever. But depending on what your backup issues are, this may not be the case. One danger is that the VTL can act as a Band-Aid and leave the actual problem unsolved. The key to a long-term solution is to address the root problem.*

*VTL systems present their own set of challenges:*

- How does your backup software vendor license VTL? It may be by capacity, tape count, or something else. Given the evolving nature of VTL, these pricing models could also be different the next time you go to make a change to your environment.*
- How does your data get to tape? In some solutions, the VTL device writes it to tape directly and in others the backup server duplicates the virtual tape to a physical tape. If the VTL writes the tape, then the backup software catalog has an incomplete view of the tapes. If the backup software duplicates the tape, then additional workload is placed on the backup server.*
- How full will the physical tape be when it is written? If the VTL does not use the same compression algorithm as the tape drive, then it cannot guarantee a tape copy will fill the physical tape. The cost of physical tapes that are never more than 75% full can add up to many thousands of dollars.*

(CONTINUED)

**TUTORIAL:** VIRTUAL TAPE LIBRARY (VTL) TECHNOLOGY

- *Will the VTL improve backup performance? VTL devices offer wildly varying performance levels. And the features such as compression, data deduplication, tape duplication, and others can have a huge impact on the performance of the system in a production environment.*

*The backup software vendors are somewhat apprehensive about VTL in the backup infrastructure. Anything that takes control of a portion of the environment is a threat to them. This is likely why they offer the disk pooling features at a lower cost or even include it as a base feature. Disk pooling allows the backup software to leverage a portion of disk storage for its own cache. Many of the benefits of VTL can be obtained by leveraging the disk pooling feature available in many backup products.*

*Tape is never more painful than when the time comes to do a restore. Is the tape on-site? Is there an available tape drive? Is the tape readable? Was the backup multiplexed or multistreamed? This is an area where VTL shines. As long as the data is on the VTL, the restore will be fast and efficient. Because there is no physical tape or physical tape drive, nearly all of the traditional tape restore challenges are gone.*

*We are seeing VTL systems sold as a solution to everything from slow tape drives to a slow network. If you have tape drives that are five years old, it will likely be easier and more cost-effective to upgrade your tape infrastructure. Obviously, a VTL will not cure a bottleneck in your network, but it will provide a fast destination for your backups. In some cases replicating that to another VTL in a remote location may eliminate the need for tape entirely.*

### **TUTORIAL:** DEDUPLICATION TECHNOLOGY

*Deduplication is a very compelling technology. Who doesn't like the idea of automatically reducing the amount of data being stored by removing the redundant bits? Deduplication is not the same as compression. Although, some vendors are attempting to blur the line between the two technologies. Deduplication searches through the storage looking for redundant sequences of data. This data is then factored out to reduce the total storage footprint. The general idea is to only store these sequences once.*

*There are a number of products bringing this technology to market. At this time there are only a couple of vendors offering deduplication in a general-purpose block-based or NAS storage array. Instead, the majority of deduplication products in the marketplace are task-specific (Virtual Tape Libraries VTL for example). Each one functions a little differently, so be sure it will solve the problem for your specific data.*

*Several of the deduplication technologies have block-alignment dependencies. This can have a significant effect on the ability to reduce a data footprint. Let's use the example of an 8 MB document file stored on a device that uses 8 KB blocks. If the file is opened, a single kilobyte is inserted at the start of the file, and it is then saved. The net effect is a new file where the data in each block is shifted over by 1 KB. If the deduplication engine only understands block alignment, then when the blocks of the new file are compared to the original file, there will be no redundant data. If the engine has the ability to examine data without regard for block boundaries, then only 1 KB of new data will be written.*

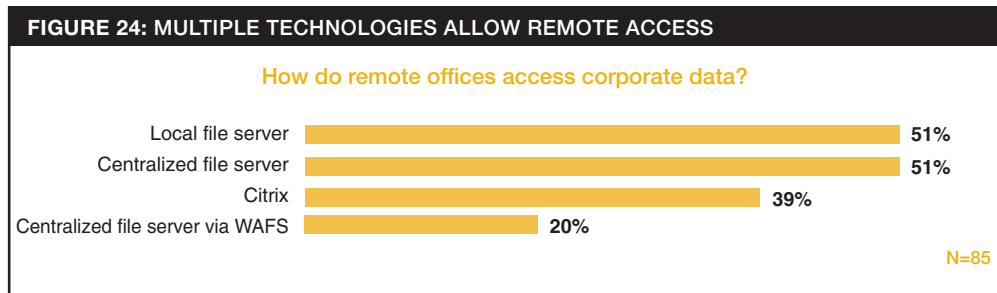
*The type of data being stored will have a significant impact on whether the block-boundary issue impacts your environment. If the goal is to save capacity each time a file is modified, then the block-boundary issue could cause a problem. However, if you are using the device to store backup data, then it is likely a non-issue. The backup device would be storing an exact copy of most of the data each time a full backup is done and a block-boundary algorithm would provide a substantial reduction in data footprint. It is exactly this type of repetitive backup scenario that vendors are referring to when they reference the 20-40X reduction in data footprint.*

*Watch for this technology to continue to grow in the marketplace. There is no debate that most environments have a tremendous amount of redundant data and everyone would like to reduce their storage footprint. Once the perceived risks in data deduplication technologies start to diminish, the economic argument should become compelling.*

## Remote and Mobile Work Force

### CURRENT ADOPTION OF REMOTE ACCESS TECHNOLOGY

Data Storage Outlook finds that storage professionals rely on multiple technologies to allow remote offices access to corporate data mean remote technology mentions per respondent average 2.12. Figure 24 shows that the majority of those interviewed mention local file server (51%) or centralized file server (51%). Others report Citrix (39%) or centralized file server via Wide Area File Services (WAFS) (20%).



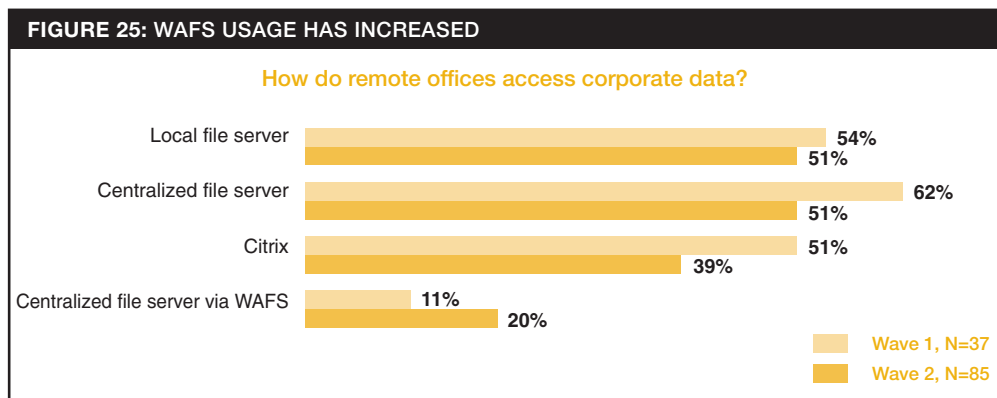
Source: Data Storage Outlook—Corporate Technologies, Inc.

Among respondents reporting WAFS, most-mentioned vendors are Cisco, followed by F5 and Juniper Networks.

More respondents said they did not backup desktops, laptops, and isolated servers (59%), than those who did such backups (41%). More respondents said they would not consider a hosted solution for backing up desktops, laptops, and isolated servers (57%), than those who would consider such a solution (43%).

### TRENDS IN ADOPTION OF REMOTE ACCESS TECHNOLOGY

Data Storage Outlook finds that technology preferences for remote access of corporate data have changed across the two research surveys fielded (Figure 25). Use of a centralized file server, local



Source: Data Storage Outlook—Corporate Technologies, Inc.

file server, or Citrix declined from Wave 1 to Wave 2. The use of a centralized file server via Wide Area File Services (WAFS) increased 81.8% from 11% (Wave 1) to 20% (Wave 2) although its use remains far lower than the other three technologies. The growing popularity of WAFS is corroborated by a *Storage* magazine survey, which finds that 22% of respondents report WAFS implementations in 2006. In 2007, an additional 12% say they will implement it, while 30% say they will evaluate it (refer to End Note 7).

Respondents participating in the Data Storage Outlook survey continue to rely on multiple technologies to deliver remote access, with mean mentions per respondent of 2.11 (Wave 1) and 2.12 (Wave 2).

The percentage of respondents reporting that they did or did not backup desktops, laptops, and isolated servers has remained virtually unchanged across the two waves of research, suggesting that the behavior is unlikely to change going forward.

Conversely, those reporting that they would consider a hosted solution for backing up desktops, laptops, and isolated servers have dramatically increased across the research survey period. Those saying they would consider a hosted solution has increased from 32% (Wave 1) to 43% (Wave 2), while those saying they would not consider such a solution have declined from 68% to 57%. While respondents saying that they would not consider a hosted solution are still in the majority, they may not be so in the near future.

#### **CONCLUSIONS: REMOTE AND MOBILE WORK FORCE**

*Legal concerns over the security of data are driving corporate backup strategies. Enterprises can either choose to limit data on desktops, laptops, and isolated servers, backup that data locally, or centralize the backup and allow remote access. The accelerating adoption of WAFS, expected to gain even more momentum in coming years, suggests that the overall trend is towards remote access which avoids the security risk of having data spread throughout the enterprise. WAFS adoption and the dramatic increase in those considering a hosted solution represent acceptance of a connected world with service on the network rather than in the data center. In Wave 1 of the research, IT professionals were far more hesitant to backup sensitive data on someone else's network. Today, with earlier bandwidth and security issues addressed, many companies have come to the conclusion that they cannot afford to not use service providers. The trend to remote access, however, does not imply lesser security concerns for data on laptops and desktops. If anything, security will become more important on these personal computing devices, but will be addressed as encryption technology becomes tightly integrated with operating systems.*

## WIDE AREA NETWORK (WAN) ACCELERATION

The following is a tutorial on Wide Area Network (WAN) Acceleration, which explains how companies can both increase bandwidth between locations and provide fast file access. *Storage* magazine reports that 19% of respondents reported use of WAN accelerators in 2006, with 67% reporting that DR is driving wide area purchases (refer to End Note 7).

### **TUTORIAL:** WIDE AREA NETWORK (WAN) ACCELERATION

*WAN acceleration provides LAN-like speed across the Wide Area Network (e.g. Point-to-Point, Frame Relay, MPLS, VPNs). Up until this point, companies have had little choice but to increase bandwidth, allowing for more concurrent data to flow between locations, but doing nothing to speed up that data. WAN accelerators remove unnecessary data from the WAN, optimize TCP communication, compress traffic, speed-up application access, drastically lower latency, and provide fast file access.*

*LATENCY THE BANDWIDTH KILLER Application and file access were not, and are not, designed for the wide area; they rely on TCP, for the most part, to transfer data. TCP is a reliable transport, so every transaction is acknowledged before sending the next transaction. On LANs, where round-trip time (latency) is only a millisecond, these applications are fast, and become faster with increased bandwidth. On a WAN, however, adding bandwidth only allows for more data to be accessed at the same time, it does nothing to increase the speed of the data, so application and file access remain slow; this is akin to sending back-and-forth information through the postal service you can send bigger packages, but they take just as long to get there and longer for the response Got it, please send the next package. WAN accelerators provide TCP, application, and bandwidth optimization.*

*TCP OPTIMIZATION Communication across TCP/IP networks requires a series of acknowledgements in order to transfer data. Operating system functions and applications are chatty, filling the wide area pipe with a series of back-and-forth communication, adding a lot of overhead, taking away from available bandwidth and decreasing application responsiveness.*

*WAN accelerators form a channel across the wide area circuit ensuring every communication is acknowledged on both sides. It is important to note that these appliances do not act as the remote service communication is established and maintained from client to server. As a session is formed, the overhead of transferring back-*

**(CONTINUED)**

**TUTORIAL:** WIDE AREA NETWORK (WAN) ACCELERATION

*and-forth acknowledgements across the WAN is eliminated, as these communications are kept local, thus lowering the response time significantly. Another benefit of optimization is filling the pipe, as TCP window sizes can be controlled and utilized, eliminating the inherent jigsaw effect of TCP contention control.*

*APPLICATION OPTIMIZATION Prominent applications (e.g. CIFS, MAPI, HTTP, NFS, FTP, and SQL) are accelerated by utilizing the predictable instructions intrinsic to them. Windows file sharing, for instance, is slow across the wide area due a large number of acknowledgements necessary to perform basic operations (e.g., copying, browsing, or opening a folder). With application acceleration, CIFS traffic is many times faster, allowing companies to consolidate file shares. Optimization of the other protocols listed above yields similar results.*

*BANDWIDTH OPTIMIZATION Data moving across a WAN is often data that has moved across the circuit before, in one form or another. As data (files, email, print jobs, backups) moves from one WAN accelerator to another, a segment-level cache is created on each accelerator. As bit patterns are repeated, reference points to those patterns are sent to the remote accelerator instead of the actual data; this reduces the traffic that moves across the WAN greatly, and also provides for a much faster transfer as the remote accelerator sends the referenced bits at LAN speeds.*

*When data that previously traversed the WAN is seen again, then a simple reference point is sent and the data moves at close to LAN speed. If the data has changed slightly, such as someone updating a spreadsheet, then a few reference points are moved, plus the new data, resulting in LAN-like speed. Forms, company logos, formatting within various document types contain redundant information that is removed from the WAN, even if it's the first pass. As the file moves from one accelerator to the next, repeatable information is identified, and then reference points are sent to the far end, significantly reducing how much data goes between the accelerators in relation to how large the file is. In some of the products, the segment-level cache is protocol-independent, so a print job of a file populates the accelerator's cache with the file's bit pattern, resulting in high-speed access to the file.*

*MAKING THE RIGHT SELECTION Every company uses their networks differently, so it's important to consider many issues what applications are used across the WAN, file characteristics (sizes, types, uniqueness), trending, number of sessions, and so on before selecting a solution. WAN accelerator manufacturers approach the WAN optimization differently; as a result, there's no one solution for every scenario.*

**(CONTINUED)**

**TUTORIAL:** WIDE AREA NETWORK (WAN) ACCELERATION

*REASONS TO CONSIDER WAN OPTIMIZATION:*

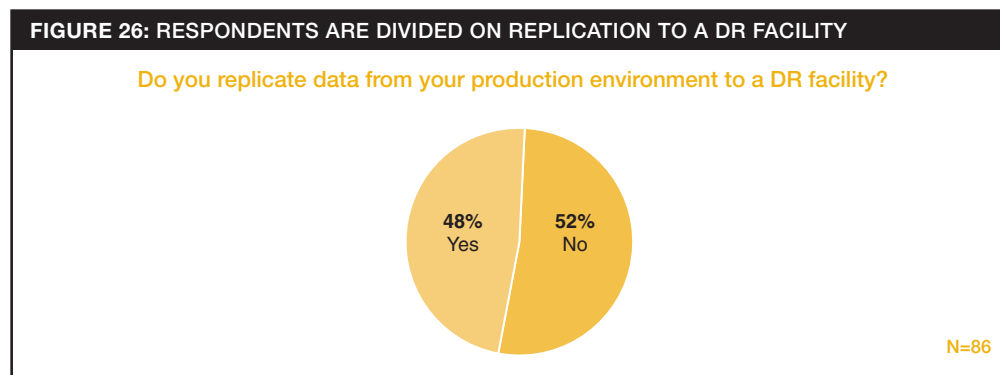
- *Increase application performance*
- *Server consolidation*
- *Client to server email across the WAN*
- *Backups*
- *Freeing up bandwidth for other applications*
- *Lowering latency*
- *International circuits*
- *Slow VPN performance*
- *Multiple locations*
- *Data replication*

## Storage Disaster Recovery (DR) Requirements

### CURRENT DATA REPLICATION TO A DR FACILITY

Respondents are equally divided on replicating data from their production environment to a DR facility (Figure 26). Data Storage Outlook finds that 48% replicate versus 52% who report they do not.

Data Storage Outlook finds that among respondents replicating data to a DR facility, 53.8% of production data, on average, is being replicated. This average varies considerably by company. Overall, 28% report less than 25 percent of production data is replicated, while 21% report 100 percent of production data is replicated.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Among respondents replicating data to a DR facility, the majority (51%) employ host-based replication, 39% employ array-based replication, and 10% employ both host- and array-based.

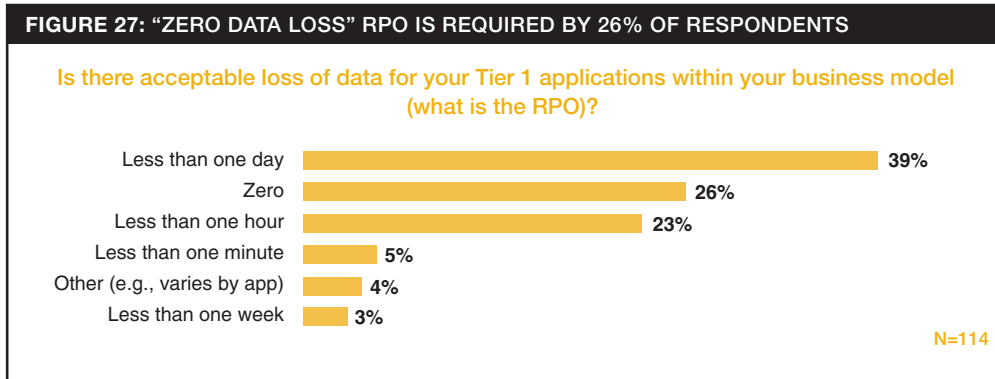
Respondents not currently replicating data to a DR facility are split as to whether they will do so in the future. The majority (54%) plan on such replication 21% within 6 months and 33% after 6 months. Slightly fewer (46%) view such replication as not in the foreseeable future.

### ACCEPTABLE DATA LOSS (RECOVERY POINT OBJECTIVE)

Figure 27 shows that 26% of the IT decision-makers have a Recovery Point Objective (RPO) of zero data loss that is acceptable for their critical business applications (e.g., email and ERP). In comparison, 23% report an RPO of less than one hour and 39% report an RPO of less than one day. An additional 5% report an RPO of less than a minute and 3% report less than a week. The remaining 4% say their needs vary based on applications. Overall, respondents Recovery Point Objective is moving towards zero.

A slim majority of respondents (53%) report they have the IT infrastructure to support their RPO for critical business applications (e.g., email and ERP). Although 40% of respondents report that they are easily meeting this objective, 13% say they are barely meeting the objective. The survey finds that 47% report they are not meeting their data loss requirements.

This data loss issue is of particular concern to life sciences companies since this data loss can delay their FDA filings and cost these companies millions in lost revenues.



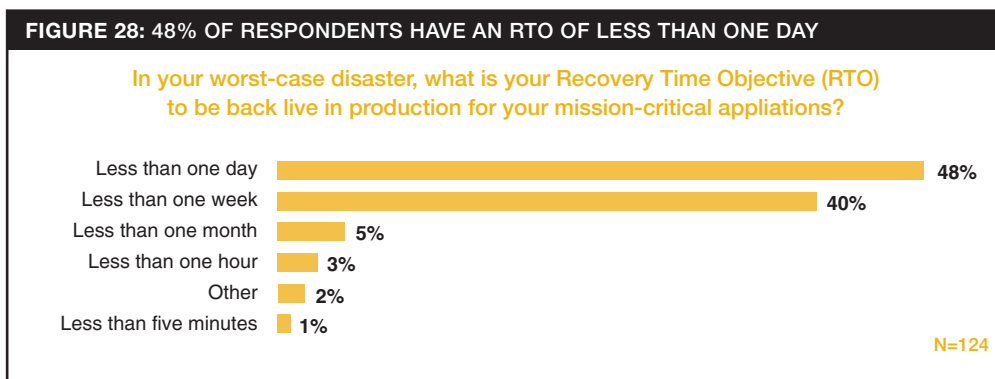
Source: Data Storage Outlook—Corporate Technologies, Inc.

**TIME TO BE BACK LIVE (RECOVERY TIME OBJECTIVE)**

The largest percentage of respondents 48% have a Recovery Time Objective (RTO) of being back live for mission-critical applications in less than one day (Figure 28). Another 40% of respondents have less stringent requirements less than one week.

Only 3% have the very tough RTO of being back live in less than one hour (1% actually report they were required to be back live in less than 5 minutes), in comparison to 5% who can wait up to one month. The remaining 2% report other. These respondents have varying requirements that depend on the nature of the mission-critical application or staff availability.

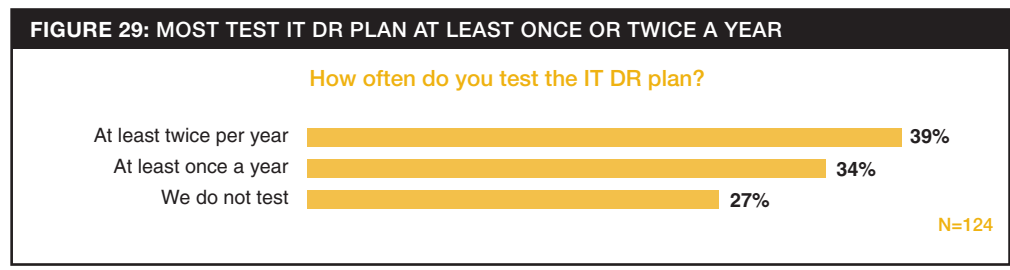
When asked if they were currently meeting their requirements to be back live, 85% said yes, although split in terms of how easily 43% said just barely and 42% said easily. An additional 15% of respondents reported that they were not currently meeting their Recovery Time Objective.



Source: Data Storage Outlook—Corporate Technologies, Inc.

### TESTING OF THE IT DR PLAN

It is important to make testing of the restore and recovery process a part of the IT test plan although many respondents do not (Figure 29). Surprisingly, 27% of all respondents admit they do not have a test plan. In comparison, most respondents have a testing schedule in place: 34% of respondents report they test at least once a year, while 39% test at least twice a year.



Source: Data Storage Outlook—Corporate Technologies, Inc.

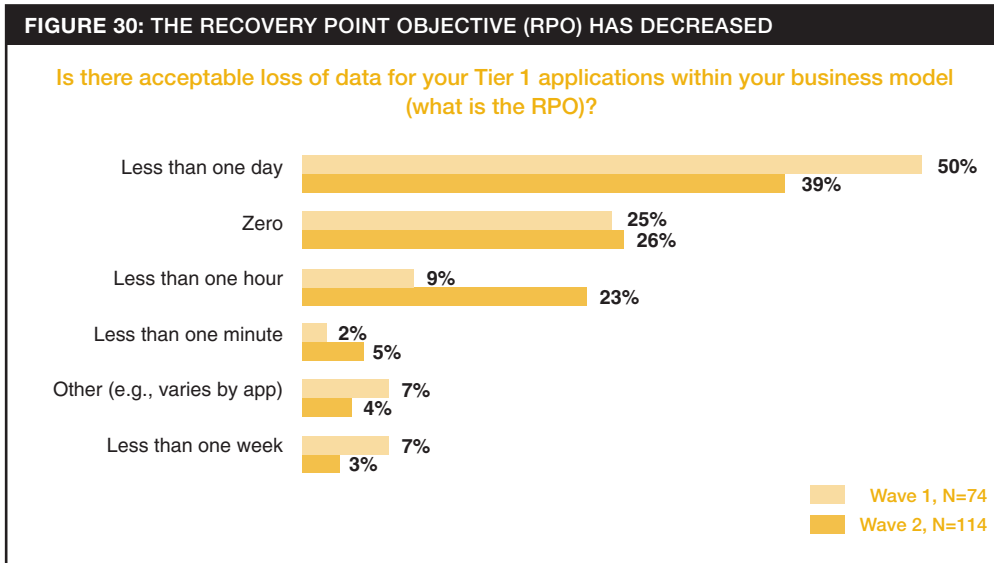
Among respondents testing the IT DR plan, the majority (72%) test mission-critical items 62% test both mission-critical applications and data, while 10% test mission-critical applications only.

### TRENDS IN STORAGE DISASTER RECOVERY (DR) REQUIREMENTS

Data Storage Outlook finds differences in replication behavior and plans across the two waves of the research. The percentage of those replicating data to a DR facility increased from 38% (Wave 1) to 48% (Wave 2). If this trend continues, those replicating may be in the majority in the near future. Corroborating this outlook is the fact that the segment planning to replicate within six months is also increasing, from 8% (Wave 1) to 21% (Wave 2).

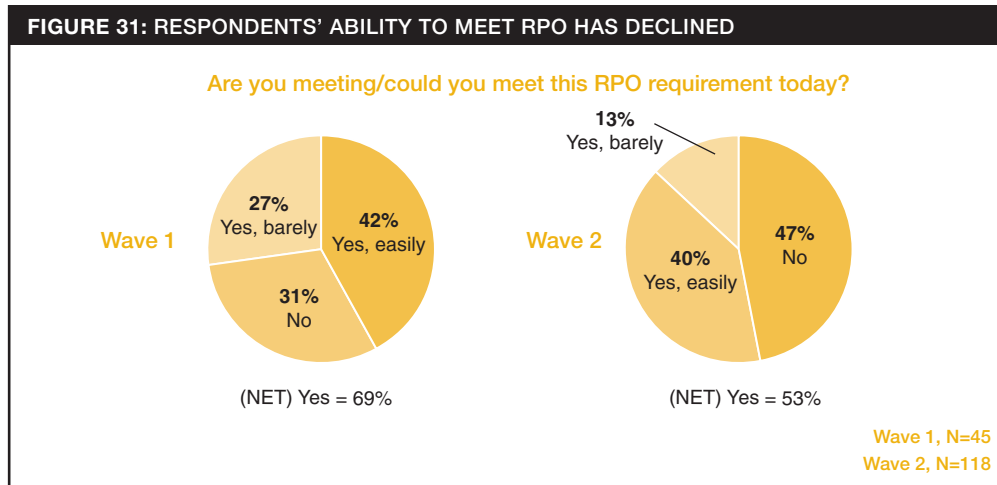
The mean percentage of production data replicated has also increased, from 47.3% (Wave 1) to 53.8% (Wave 2) a 13.5% increment. Those employing host-based replication have modestly increased from 47% (Wave 1) to 51% (Wave 2) while those employing host-based and array-based have decreased from 13% (Wave 1) to 10% (Wave 2).

Data Storage Outlook finds that the Recovery Point Objective for Tier 1 applications has changed over the course of the two surveys, and is getting smaller (Figure 30). Importantly, those reporting less than a day have decreased from 50% (Wave 1) to 39% (Wave 2) and those reporting less than an hour have increased from 9% (Wave 1) to 23% (Wave 2). Interestingly, those claiming zero remained relatively unchanged at about one-quarter of the sample.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Respondents' ability to meet their RPO has diminished across the term of the study (Figure 31). Those reporting yes either easily or barely have decreased from 69% (Wave 1) to 53% (Wave 2), while those reporting no have increased from 31% (Wave 1) to 47% (Wave 2). If this unfortunate trend continues, those not meeting their data loss requirements could outnumber those meeting such requirements in the near future.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Data Storage Outlook finds that the Recovery Time Objective for mission-critical applications is getting shorter over the course of the research. Importantly, those reporting less than one week have decreased from 53% (Wave 1) to 40% (Wave 2) and those reporting less than one day have increased from 28% (Wave 1) to 48% (Wave 2).

Respondents' ability to meet RTO remains relatively unchanged across the two waves of the research, with over three-quarters of the sample saying they are meeting their requirements, in spite of the fact that the time allowed has shrunk.

Data Storage Outlook finds that the number of respondents testing the IT DR plan increased over the term of the research, as did the frequency of testing. Those testing the plan have increased from 55% (Wave 1) to 73% (Wave 2), with the greatest increment being among those who report testing at least twice a year, which has gone from 21% (Wave 1) to 39% (Wave 2). Those reporting that they do not test their disaster recovery plans declined in the same time period from 45% to 27%.

Data Storage Outlook also finds that the items tested changed over the term of the research. There is a clear trend towards testing of both mission-critical applications and data, which has increased from 44% (Wave 1) to 62% (Wave 2). Correspondingly, those who report that they test mission-critical applications have slightly increased from 7% (Wave 1) to 10% (Wave 2).

#### **CONCLUSIONS: STORAGE DISASTER RECOVERY (DR) REQUIREMENTS**

*Data Storage Outlook finds it interesting that 85% of respondents report they are meeting their RTO requirements, while only 53% report they are meeting their RPO requirements. Clearly, it is easier to be back live than it is to be back with acceptable data loss. It is sobering to learn that almost half of respondents (47%) report they are not meeting RPO requirements. An enterprise may require a more stringent RPO than funding is available for. Given the extreme importance placed on RTO and RPO, it is critical that management be kept informed of how well those objectives are being met and if they are considered business-critical, that appropriate funding levels are made available.*

*DR via replication is becoming increasingly common for three reasons. First, companies are mandating smaller RPO and RTO requirements. Second, the cost of replication is dramatically decreasing as a result of replication functionality being baked into other technology, lower TCO, and lower communication line costs. Third, newer technologies have become available such as WAFS that enable lower costs. Although the testing of the DR plan increased over the term of the research, still 27% report they are not testing, which certainly is not considered best-practice. The availability of newer technology that allows online DR testing without disrupting replication or production should drive the percentage of those not testing to zero over time.*

## Storing Employee Email

The Data Storage Outlook interviews reveal that the term email retention is being interpreted differently by different companies. The scenarios include one or more of the following:

- Companies use an Information Lifecycle Management (ILM) solution for users' mailboxes. As email ages, the IT organization puts this email into a less costly place that is still accessible by the end-user.
- IT organizations choose to save a copy of every inbound and outbound email message.
- Companies provide their employees with a set storage capacity, (e.g., 100 MB) and let people choose which email to retain. Companies acknowledge that some departments may petition the IT department for more storage capacity because of the nature of their work, (e.g., marketing departments may need more storage capacity because they often deal with large image files).
- Companies ignore the problem and neither backup nor restore email.

### LENGTH OF TIME STORED

Figure 32 shows that while 35% of respondents store their employee email indefinitely, 43% report having policies that define the email retention timeframe: 3 months or less (13%), 3-12 months (3%), 1-2 years (9%), and 2-7 years (18%). In comparison, 22% report no formal email retention policy.



Source: Data Storage Outlook—Corporate Technologies, Inc.

In some cases email retention is also tied into regulatory compliance. Many companies acknowledge spam as a daily annoyance and project that spam is impacting their storage requirements since some regulations require that companies store email indefinitely. Removing the spam before storing email will have a significant reduction on storage requirements.

**CAPACITY DEDICATED TO STORING EMPLOYEE EMAIL**

On average, respondents report dedicating 2.24 TB to storing employee email although capacity varies considerably by company. Overall, 42% of respondents report less than 1 TB of storage dedicated to storing employee email, while 17% report more than 5 TB for email storage.

Actual responses vary from .02 TB to 100 TB, depending on company revenue size.

(Also, refer to Appendix, Figure D, for an overview of how storage capacity dedicated to employee email varies by key vertical markets, including financial services, manufacturing, and life sciences.)

**TRENDS IN STORING EMPLOYEE EMAIL**

Data Storage Outlook finds that companies are storing email longer the term of email storage is lengthening from the first to the second wave of research (Figure 33). Respondents reporting indefinitely increased from 24% (Wave 1) to 35% (Wave 2), while those reporting less than three months dropped from 20% to 13% in the same timeframe.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Storage dedicated to employee emails has increased dramatically across the term of the research, from 1.42 TB, on average (Wave 1), to 2.24 TB (Wave 2), a 58% increment. A recent Gartner survey among global Small and Midsized Businesses (SMBs) finds that email is now one of the four applications consuming the largest amount of storage capacity behind databases and ahead of files (e.g., Microsoft Word) and ERP (refer to End Note 5).

Email storage capacity increased in all vertical industries covered except computer manufacturing, with life sciences reporting the largest percentage increase. Companies of all sizes reported increased email capacity.

**CONCLUSIONS: STORING EMPLOYEE EMAIL**

*Email is trending towards being stored indefinitely not only because storage is relatively inexpensive, but also because increasing storage is less problematic than figuring out what email can be safely deleted. Such storage is not without its challenges, however. For compliance and other reasons, companies need to be able to find specific email content, which is cumbersome, at best, with tools currently available.*

*Most email retention policies today are being driven by legal concerns, and the problem is getting larger every year. Compliance tools are being purchased for email management purposes as much as for legal compliance. These tools can reduce the footprint on the email datastore online by extracting significant quantities (over 90%) from the Exchange environment out to tier 2 storage.*

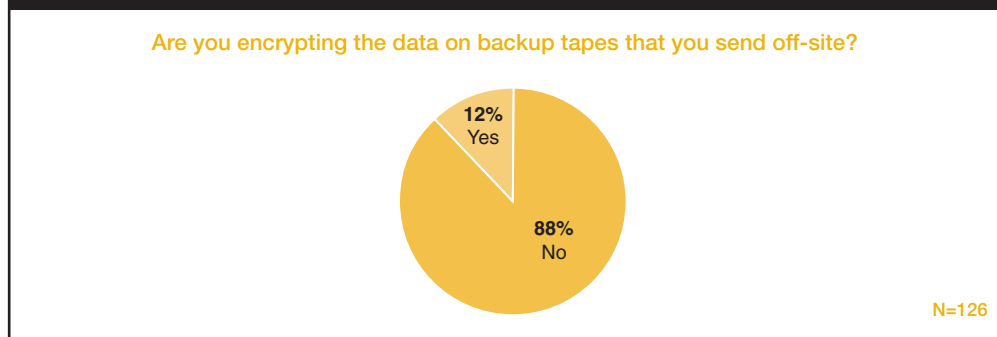
## Storage Security Requirements

### CURRENT DATA ENCRYPTION ON BACKUP TAPES

Data Storage Outlook is surprised to find only 12% of respondents currently encrypt the data on the backup tapes they send off-site (Figure 34). This issue was raised in the study because there are some regulated industries, such as pharmaceuticals/life sciences, where companies need to ensure accuracy of data. Therefore, these companies are concerned with whether they can restore data off a tape that is not encrypted, and still be compliant with FDA and other regulations.

Overall, 88% of respondents believe it is unnecessary to encrypt their backup tapes. Most feel that they pay third-party vaulting companies, so they do not need to worry about such issues. Currently, it appears that data encryption is still not a big concern for IT decision-makers.

**FIGURE 34: ONLY 12% ENCRYPT DATA ON BACKUP TAPES**



Source: Data Storage Outlook—Corporate Technologies, Inc.

It also appears that encryption will become more of a priority in the future. Among respondents not currently encrypting data on backup tapes, the majority (68%) said such encryption was not in the foreseeable future. However, 32% report that they plan to encrypt data on backup tapes

31% within 6 months and 1% after 6 months. This finding is consistent with a *Storage* magazine survey which reveals that 20% plan to implement data encryption for security in the coming year, and 42% plan to evaluate it (refer to End Note 7).

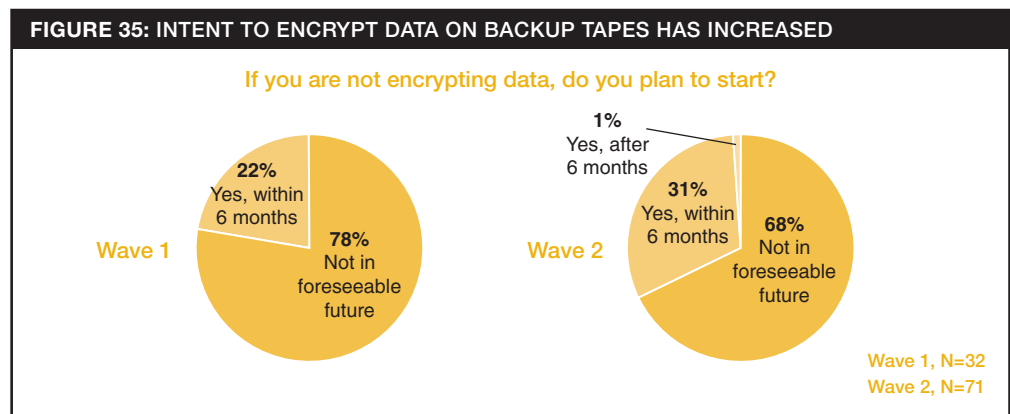
### USE OF THIRD-PARTY TO HOUSE BACKUP TAPES

The majority of respondents, 77%, currently use a third-party to house their backup tapes, which Data Storage Outlook considers a good business practice. The third-party company cited most frequently by respondents is Iron Mountain.

The majority of respondents state that these third-party companies are meeting their specifications (98%), and that it is very easy to retrieve their stored backup tapes from these third-party companies (70%).

### TRENDS IN DATA ENCRYPTION ON BACKUP TAPES

Data Storage Outlook finds an increasing intention to encrypt data across the term of the research (Figure 35). Respondents saying they planned to encrypt data within six months rose from 22% (Wave 1) to 31% (Wave 2), Conversely, those probably not planning to encrypt in the foreseeable future dropped from 78% to 68% in the same timeframe.



Source: Data Storage Outlook—Corporate Technologies, Inc.

Data Storage Outlook finds that use of a third-party to backup tapes has increased over the term of the research. Respondents reporting use of a third-party rose from 69% (Wave 1) to 77% (Wave 2), while those not using a third-party declined from 31% to 23% in the same timeframe.

Satisfaction with third-party houses and ease of retrieving stored assets have remained relatively constant over the term of the research.

**CONCLUSIONS: STORAGE SECURITY REQUIREMENTS**

*Data encryption on backup tapes is an urgent problem that needs to be immediately addressed. A recent Gartner survey estimates the cost to companies of allowing customer information to be exposed through a compromised business process at \$90 per exposed account reflecting legal expenses, professional fees, and communications to affected customers, plus the cost of cleanup and recovery, and provision for after-the-fact security. At 145,000 compromised records, ChoicePoint reported \$11.4 million in charges directly related to its compromise incident and saw its valuation drop by \$720 million directly after the exposure. Other potential financial impacts noted by Gartner for companies experiencing large compromises include liability lawsuits, and customer and supplier desertion, which could lead to financial failure (refer to End Note 10). Some exposures are a result of backup tape loss, and in those cases data encryption on backup tapes prevents such compromise and the enormous costs associated with it.*

*It is difficult to understand why only 12% of respondents report they are encrypting the data on backup tapes sent off-site. Is it because they are possessed with a it won't happen here mentality? A fear of not being able to restore the data? Or could it be that the individual actually running the backups wants nothing to do with the perceived additional complexity associated with encryption? The fact is that there is no good reason to avoid encryption on backup tapes any longer. Although complexity of encryption varies with specific technologies, most have minimal negative impact to existing environments. And, at a cost of \$6 (data encryption) to \$16 (data encryption, host-based intrusion prevention, and strong security audits) per customer account, Gartner analyst Avivah Litan makes the case that protecting customer records is a magnitude less than paying for a compromise (refer to End Note 11).*

### **TUTORIAL:** STORAGE AND SECURITY

*Regulations and corporate disasters are pressuring companies to increase their security across the board. Fines and customer unhappiness are just some of the penalties being paid when data gets lost or revealed. There are many aspects to storage security and many new technologies that can be brought to bear to decrease unauthorized access and increase logging of all access to data.*

*The most common security improvement sites are making is the encryption of data as it moves from disk to tape. By encrypting data on tape, the vulnerability of that data as it is moved off-site (typically by non-company employees) is vastly improved. The two methods being used are the addition of a device in-line between the backup server and the tape drives, and the use of tape drives that themselves provide encryption. The in-line devices have two major advantages: they can dedicate hardware to encryption to provide better performance (in some cases line-speed) and they can provide infrastructure for key management. Tape-drive-based encryption can be convenient but frequently lacks in those two areas.*

*Another storage security aspect under consideration at many sites is encryption of data on disk. Frequently applications provide this encryption as part of the design and implementation of a computer service (i.e., storing credit card numbers in a column in a database) but sites are now seeking more broad-ranged data encrypt everything or encrypt data on specific volumes strategies. Some in-line encryption devices can provide this kind of security, but in general this is a rapidly evolving component of storage security.*

*A new area of concern and technical innovation is monitoring and auditing of data access. Typically the applications that are providing the access (i.e., databases) provide logging of that access via text files. This type of information is becoming less desirable as more rapid and distilled knowledge is needed. New applications are stepping into the breach and providing summary as well as detailed information about who has access privileges to what data and who actually accessed that data.*

## End Notes

- (1) [http://labs.google.com/papers/disk\\_failures.pdf](http://labs.google.com/papers/disk_failures.pdf)
- (2) <http://www.cs.cmu.edu/~bianca/fast07.pdf>
- (3) NAS includes CIFS and NFS
- (4) SAN includes iSCSI and file servers attached to SAN
- (5) Gartner research report entitled User Survey Analysis: Top Storage Demands From Global Small and Midsize Businesses, Pushan Rinnen, October 13, 2006
- (6) Forrester Research report entitled 2006 IT Spending in the SMB Sector, Michael Speyer, April 11, 2006, Forrester's Business Technologies North America.
- (7) *Storage* magazine report entitled Fall 2006 Purchasing Intentions Survey, Mark Schlack and Rich Castagna, August 2006
- (8) Gartner research report entitled Management Update: Budget Surveys Show Spending Increases and Shifting Priorities for IT, Barbara Gomolski, Robert De Souza, and Dean Blackmore, October 26, 2005
- (9) Gartner research report entitled Key Issues for IT Providers Targeting Small and Midsize Businesses, James A. Browning, March 30, 2007
- (10) Gartner research report entitled Management Update: Data Protection is Less Costly Than Data Breaches, John Pescatore and Avivah Litan, September 28, 2005
- (11) *Information Week* article entitled Security Cleanup Costs Much More Than Encryption: Gartner, Gregg Keizer, June 6, 2006

### NOTES

---

---

---

---

---

---

---

---

---

---

## Appendix: Key Vertical Market Data

**FIGURE A:** CURRENT STORAGE CAPACITY (SAN VS. NAS VS. DAS) BY KEY VERTICAL MARKETS

Vertical Market	SAN (average)	NAS (average)	DAS (average)
Financial Services	15.61 TB	2.39 TB	23.13 TB
Manufacturing	9.05 TB	1.72 TB	17.00 TB
Life Sciences	27.88 TB	4.33 TB	1.67 TB

N=Financial Services=23

N=Manufacturing=26

N=Life Sciences=4

**FIGURE B:** EXPECTED STORAGE CAPACITY (NEXT 12 MONTHS) BY KEY VERTICAL MARKETS

Vertical Market	SAN (average)	NAS (average)	DAS (average)
Financial Services	21.75 TB	2.83 TB	20.94 TB
Manufacturing	12.83 TB	1.92 TB	16.98 TB
Life Sciences	34.00 TB	6.00 TB	1.67 TB

N=Financial Services=23

N=Manufacturing=26

N=Life Sciences=4

**FIGURE C:** STORAGE GROWTH RATES (CURRENT VS. NEXT 12 MONTHS) BY KEY VERTICAL MARKETS

Vertical Market	SAN (average)	NAS (average)	DAS (average)
Financial Services	39.3%	18.2%	-5.4%
Manufacturing	41.7%	11.5%	-0.1%
Life Sciences	22.0%	38.5%	0.0%

N=Financial Services=23

N=Manufacturing=26

N=Life Sciences=4

**FIGURE D:** STORAGE CAPACITY DEDICATED TO STORING EMPLOYEE EMAIL BY KEY VERTICAL MARKETS

Vertical Market	Current Capacity (average)	Prior Capacity (average)	Percent Change
Financial Services	2.36 TB	1.97 TB	20%
Manufacturing	2.93 TB	0.50 TB	486%
Life Sciences	4.50 TB	0.75 TB	500%

N=28

## Report Contributors

Jesse St. Laurent

### DIRECTOR, RESEARCH AND DEVELOPMENT

Jesse St. Laurent has been with Corporate Technologies since 1996. As Director of R&D, Jesse is responsible for evaluating new technologies for inclusion in our solution architecture. He leads our proprietary New Product Adoption process and manages the Technology Interoperability Lab. Jesse is also very involved in architecting and implementing customer solutions and defining best-practices in our Storage/Backup practice.

Before joining Corporate Technologies, Jesse was a systems consultant for the Computer Science Department at Brown University. He has been working with Internet technologies since 1992, when he played a key role in launching an ISP operation. Jesse holds a BS in Computer Science from Brown University.

Peter Baer Galvin

### CHIEF TECHNOLOGIST

Peter Baer Galvin has been the Chief Technologist for Corporate Technologies since 1995. In this role he has been responsible for architecting and managing the implementation of many complex enterprise computing solutions for hundreds of our clients. Peter is also responsible for evaluating emerging systems and security technologies for potential inclusion in our solution set. He is a frequent speaker at conferences worldwide on the topics of systems management, security, and performance.

Peter is a regular columnist for *SysAdmin Magazine* and also the co-author of *Operating Systems Concepts*, a textbook used at many universities. Previously, Peter was an independent consultant and Systems Manager for Brown University's Computer Science Department. He holds a BA in Mathematics from Wesleyan University and a MS in Computer Science from the University of Texas at Austin.

Richard Roy

### PRINCIPAL NETWORK SECURITY ENGINEER

Richard Roy has 15 years experience implementing and designing complex corporate LAN/WAN environments, SAN attached Windows Cluster implementations, plus firewall, VPN, intrusion prevention, high availability, and load balancing technologies.

His experience prior to Corporate Technologies includes state and federal network/systems integration as a Senior Field Engineer, as well as being an IT Director. Richard has a BS in Computer Science from MCLA. Certifications include Sun Premier; CCNA, CCNP, CNE, CCSA, CCSE, MCP, JNCIA (-FWV, -IDP, -SSL); and F5 Big IP.

## **Corporate Technologies, Inc.**

For over a decade, Corporate Technologies has been helping companies tap the power of technology to drive the success of business. We are experts in every aspect of IT and we pride ourselves on solving your toughest IT challenges. We deliver a full spectrum of cost-effective, reliable infrastructure hardware, software and professional services in the areas of storage, networking, security, systems, support, and data integration. And we draw on our technical expertise and business knowledge to design, build, and maintain high performance, efficient, scalable, and secure enterprise IT solutions that work together.

Our clients benefit from our hands-on experience in production environments, as well as our demanding test practices. Our New Product Evaluation (NPE) process involves rigorous technical analysis to benchmark—in our independent Labs—performance, reliability, and interoperability metrics on promising new hardware and software technologies. Furthermore, we publish test results in Lab Reports and deliver them continuously to our customers to facilitate their technology evaluations.

This expertise enables Corporate Technologies to simplify our clients' storage infrastructure, streamline storage management, protect critical data, optimize resources, and ensure business continuity.

## **Data Storage Outlook™**

Corporate Technologies developed Data Storage Outlook™ in order to ensure that our product and services offerings meet the needs of our clients and the marketplace. It has become part of our ongoing market research initiatives. The 2008 Summary Report is designed to help IT decision-makers benchmark their data storage requirements and drive their strategic planning.

For more information, visit [www.StorageOutlook.com](http://www.StorageOutlook.com).

If you have questions about this report, contact [info@StorageOutlook.com](mailto:info@StorageOutlook.com) or call 781.273.4100.

To be notified about future industry outlook studies and updates on storage, security/networking, operating systems/servers, data integration, and other technology topics, contact [info@StorageOutlook.com](mailto:info@StorageOutlook.com) or call 781.273.4100.