White Paper

Understanding the Functional Agility of Modern Tape in Production and Protection Storage

By Mark Peters, Sr. Storage Analyst and Jason Buffington, Sr. Data Protection Analyst

September 2015

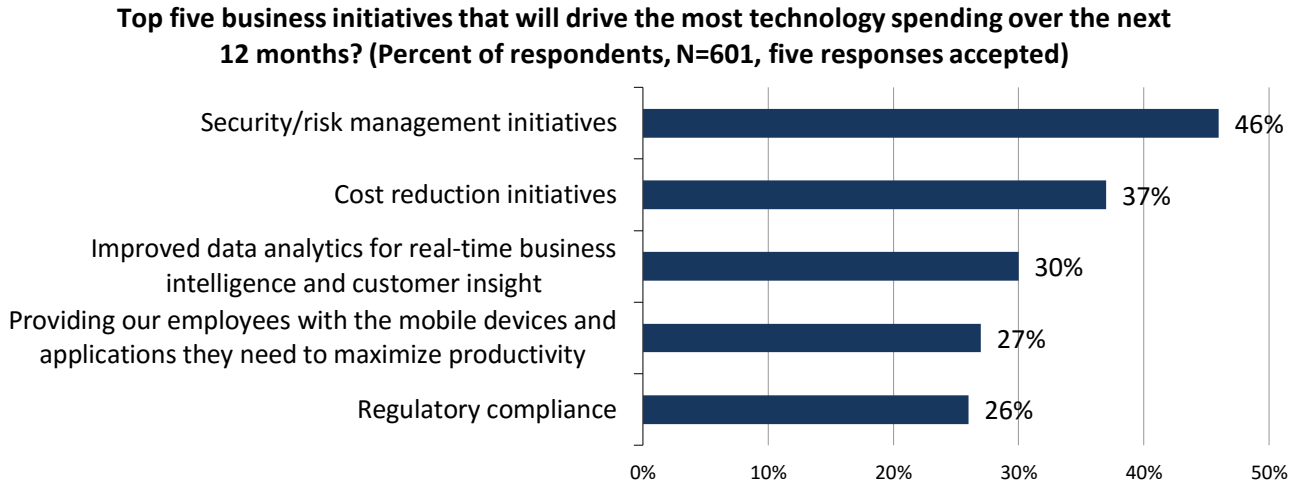
This ESG White Paper was commissioned by IBM and is distributed under license from ESG.

Contents

Introduction

IT continues to be pressured to *do more with less*; less budget and less staff, with less latency and less downtime, but with more data and higher SLA expectations than ever before. In almost all cases, the IT team's views on how the status quo should evolve doesn't drive IT strategy, but the business units' and leadership's directives do (see Figure 1).[1]

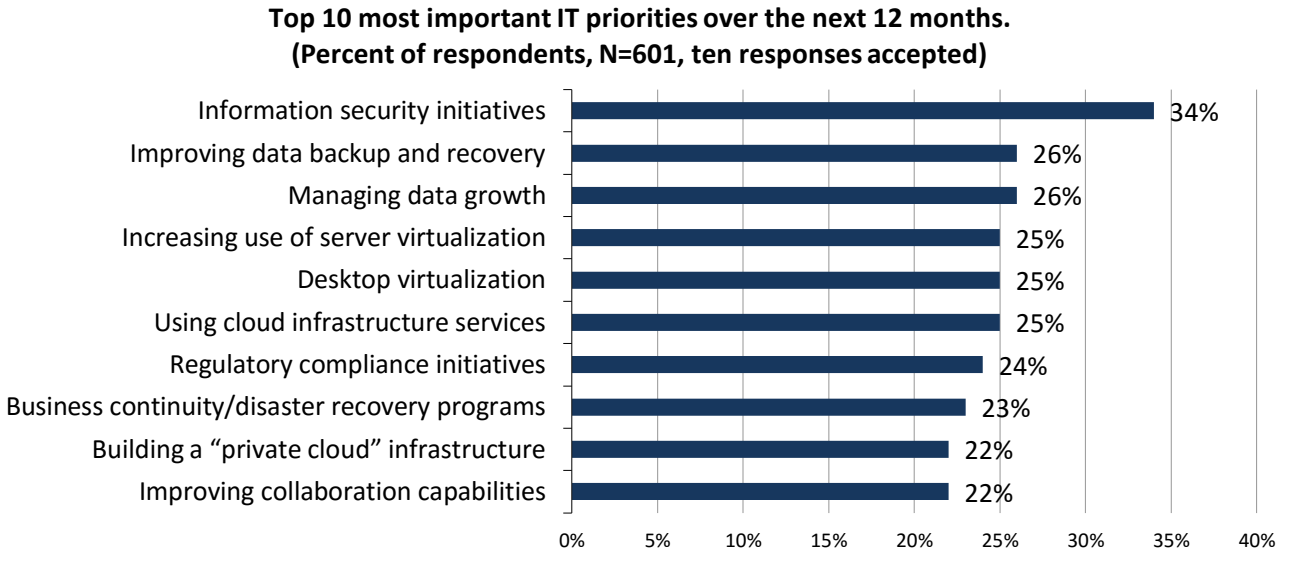
Figure 1. Top Five Business Initiatives Driving Technology Spending



Source: Enterprise Strategy Group, 2015.

Perhaps nowhere is IT truly more challenged to *do more with less* than in storage. This is especially true in production and protection storage, considering that those underlying infrastructures are growing at approximately 40% year-over-year[2] even though storage and data protection *budgets* are growing far less than that. While budgets aren't growing linearly to scale, Improving backups and managing data growth are among the IT spending priorities most-cited by ESG research respondents in 2015 (see Figure 2).[3]

Figure 2. Top Ten IT Spending Priorities in 2015



Source: Enterprise Strategy Group, 2015.

[1] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.
[2] Source: ESG Research Report, *Backup and Archivng Convergence Trends*, April 2014.
[3] Source: Ibid.

Said another way, organizations simply cannot keep scaling storage in the ways that many historically have. Instead, IT architects and storage/data protection experts need to rethink how they manage storage, particularly high-value data that may be less frequently accessed.

Understanding the Technical Challenges in Managing Dormant Data

In order to fully assess the challenges, it is important to consider the functional requirements and caveats in managing dormant/semi-dormant (yet still valuable) data in both primary storage and secondary/tertiary protection storage.

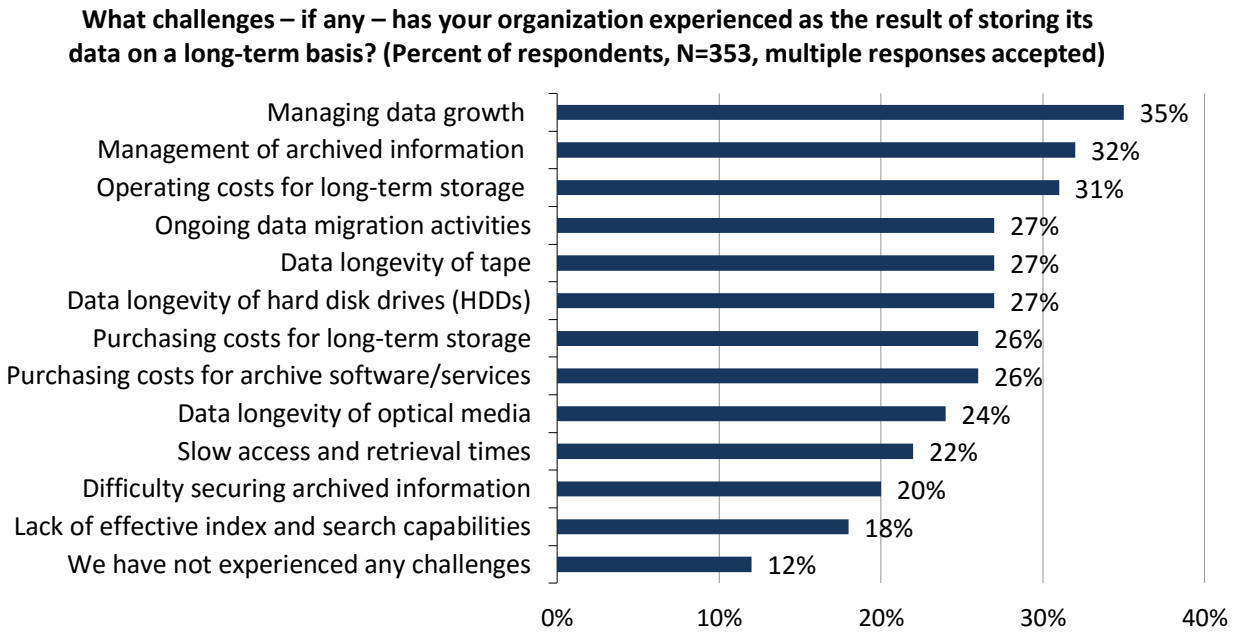
Technical Challenges in Managing Cold Production Data

It is beguiling to talk about "cold" production data as if it is something very easily defined, and therefore has technical challenges and attributes that must be served entirely differently from "warm" or "hot" production data. The truth is that categorizing and segmenting data—and hence how it is treated—will vary from user to user, and across applications within users. What about data that is "tepid" or "lukewarm"? Is it getting hotter or getting cooler? Hence, many users will say something along the lines of (to paraphrase the well-known legal quote) "Well, I don't know exactly what cold production data is, but I do know it when I see it!" What does all this mean? Simply that cold *production* data is technically no different in essence from any other kind of production data; it must invariably be easily and directly accessible, and at a speed and price that is in line with the application, user, and/or business need for which it is being used. For some users and uses, that can only mean server-based flash, while for others, it can perfectly easily (and sensibly) mean a sophisticated, integrated modern tape system.

Technical Challenges in Managing Long-term Protection or Preservation Storage

Regardless of whether it's for industry- or government-mandated regulatory compliance or simply for operational best practices, organizations of all sizes are storing more data for longer, even as new data continues to grow. When asked about the challenges related to long-term retention, the problems identified by research respondents are mostly summarized in managing the scale of the data (including both CapEx of the storage itself and OpEx of managing the info, infrastructure, and its maintenance), as seen in the top four responses shown in Figure 3.[4]

Figure 3. Challenges Experienced as a Result of Storing Data on a Long-term Basis



Source: Enterprise Strategy Group, 2015.

[4] Source: ESG Research Report, *Backup and Archiving Convergence Trends*, April 2014.

Plan for a Hybrid Approach to Storage, Particularly as It Relates to Data Retention

In considering the challenges seen in Figure 3, it becomes apparent that IT architects and operators have to rethink how they leverage storage in order to fulfill the long-term data retention requirements that their organizations have, while delivering higher performant storage for production usage. In many cases, this will result in a hybrid approach to storage, particularly as it relates to data retention over time. And for myriad reasons, many organizations have and will continue to utilize modern tape solutions as part of their modern data protection and management strategy.

How Modern Tape Addresses the Technical Needs of Modern Storage

The preceding sections have established that the value proposition of tape is well suited to addressing one of the foremost challenges of modern IT: balancing increasing data scale with increasing pressure to constrain costs. Furthermore, the value proposition of tape has particular suitability for long-term, low-activity, or dormant data and data protection uses. However, at some point, the *conceptual* aspects that create such a value proposition must be made credibly manifest by the underlying technical attributes of modern tape media and systems. Theory doesn't protect and store data…technology does. The purpose of this section is to establish some facts about tape, and to address the negative myths that all too often rule many users' perceptions.

Modern Tape Media: Capacity and Reliability

Tape media must be highly reliable, portable, and rugged enough to be moved without impacting reliability, while having a high capacity and very long life for archival applications. Modern tape drives include enterprise-class products such as IBM's TS11xx family, and a number of products built around LTO.[5] The latest media is based upon a Fujifilm developed technology called Barium Ferrite (BaFe), although large amounts of the preceding generation, Metal Particulate (MP) media, are still sold and will be in use for many years.

- **Capacity:** The amount of data that can be placed onto a given area of media is the areal density. Compared with a standard 3.5" HDD—which has a typical areal density approaching 600 Gb/in^2—the tape areal densities in use today are not nearly as "crowded," which, in colloquial terms, simply means that BaFe capacity has a lot of room to grow. Some useful current comparisons are shown in Table 1.

Table 1. Comparing the Key Specifications of Tape Media and a 3.5" SAS Hard Disk Drive

Drive Type and Media	Capacity (native)	Data Transfer Rate	Areal Density
LTO-5 MP	1.5 TB	140 MB/sec	1.2 Gb/in^2
LTO-6 MP, BaFe	2.5 TB	160 MB/sec	2.2 Gb/in^2
TS1150 BaFe	10.0 TB	360 MB/sec	8.0 Gb/in^2
3.5" HDD SAS 7200 RPM	4.0 TB	175 MB/sec (sustained)	578 Gb/in^2

Source: Enterprise Strategy Group, constructed from vendors' published product specifications, 2015.

- **Future Capacity Outlook:** Projections from vendor-neutral groups such as INSIC suggest that the annual tape areal density growth rates will either maintain or exceed their traditional 30% values, while annual HDD areal density growth rates will not maintain their traditional 40% values using vertical recording and will likely slow toward just 10% per annum.[6] IBM's research and development efforts are in line with this upbeat future for tape:

[5] LTO refers to the technologies developed by the LTO Technology Provider Companies HP, IBM, and Quantum and supported by an ecosystem that produces compatible media and associated hardware and software products.
[6] Source: Information Storage Industry Consortium, *Magnetic Tape Storage Roadmap,* October 2013.

- In January 2010, scientists at IBM Research in Zurich recorded data onto the (then) new BaFe tape media at a density of 29.5 billion bits per square inch, which equated to the potential to yield a native (uncompressed) capacity of 35 TB on a single cartridge.
- Four years later, in May 2014, IBM revealed a proof of concept for 154 TB of uncompressed data on a single cartridge. Just one year later, that number rose to 220 TB.[7]

This shows that the current uncompressed cartridge capacity (10 TB for the TS1150) has enormous room to grow, meaning that tape can remain highly cost-competitive and attractive for years to come.

- **Raw Reliability:** Since tape is often glibly dismissed by some IT users as "unreliable," the contemporary situation is worthy of in-depth consideration…and might come as a surprise to many. The better electro-chemical attributes of BaFe media (such as high coercivity and Signal-to-Noise ratios) lead to the simple fact that the *raw reliability* of the media for today's enterprise and midrange tape systems—as measured by the published uncorrectable bit error rates from vendors—ranges *from one to three orders of magnitude higher* than that of even the best (Fibre Channel or SAS) disk media (see Table 2).

Table 2. Comparing the Raw Reliability of Tape and Disk Media

Media Type[8]	Technical: Uncorrectable Bit Error Rate (UBER) Range	Practical: What This Means in Plain English
Enterprise Tape	1 x $10E^{20}$ bits	Enterprise tape media (BaFe on IBM TS1150 drives) has raw reliability levels that are *4 orders of magnitude higher* than enterprise HDDs.
Midrange Tape	Generally quoted at 1 x $10E^{17}$ bits	Mainstream midrange tape media (MP) has raw reliability that is *2 orders of magnitude higher* than enterprise SATA HDDs.
Disk Drives (HDD)	1 x $10E^{14}$ bits for desktop SATA 1 x $10E^{15}$ bits for enterprise SATA 1 x $10E^{16}$ bits for enterprise class FC/SAS	The underlying reliability of the media used in all standard contemporary disk drives ranges from *1-6 orders of magnitude less* than contemporary tape drive media.

Source: Enterprise Strategy Group, constructed from vendors' published product specifications, 2015.

These media reliability differences are not at all marginal—they are significant. As one might expect, the latest enterprise tape media is considerably more reliable than the mainstream midrange media, but both offer multiple orders of magnitude better raw data reliability than disk drives. These differences matter because of the workloads for which tape is used. Of course, in daily operations, many variables can affect "your mileage." Such variables include degradation, the tweaks that individual vendors claim to be able to make, and the slight differences in calculation methods that they certainly do make all the time! *However,* all such variance and debate is essentially irrelevant when one considers the ability of tape media to reduce the potential for hard errors by so much compared with disk media. The variances between vendors and their assumptions and models cannot overshadow the key point: *All types of modern tape media have better raw reliability than any type of modern disk media.*

[7] Announced April 2015.

[8] For this media comparison, Enterprise Tape refers to Barium Ferrite tape media used in enterprise tape drives such as the IBM TS1150 (the specs used here), and Midrange Tape refers to the traditional MP formulation of LTO tape media.

Comparative Reliability: A casual glance at the numbers in Table 2 might make them seem disingenuous—after all, most disk systems *seem* fairly reliable in practice, don't they? That's true, but this perception applies to *systems*, rather than just the media. When evaluating such systems, the reliability (both perceived and real) comes in large part from the application of such tools as Error Correction Code and various levels of RAID. Since these tools can be—and indeed sometimes are—also used on tape systems, it is categorically reasonable to compare the underlying raw media reliabilities for two main reasons:

- The fact that tape media reliability exceeds that of disk media runs counter to much of the "received wisdom" in the IT industry; this can lead to IT professionals making suboptimal choices in their media decisions.
- Raw media reliability is clearly a vital "suitability attribute" in terms of the workloads for which tape is deployed; most tape is deployed for archive and backup purposes, which makes the maximum reliability of the underlying media (which, from an operational perspective, can be translated as guaranteeing data access) of paramount importance.

Modern Tape Systems: Additional Reliability and Value

The raw reliability of tape media is so often questioned that addressing the facts in depth is important, but other aspects of the overall tape system should also be mentioned. Hardware systems, such as the drives and automated tape libraries from IBM, are themselves extremely reliable these days. Furthermore, the interaction between the drives and the media (in other words, head wear and other media impact) is optimized by improvements on both sides: sophisticated tape tracking and handling on the one hand, and more capable media with a better Signal-to-Noise ratio on the other hand. This means not only fewer errors, but it also precludes most performance degradation as heads and media age.

The upshot of good drives, libraries, and media is not only ultra-reliable data access, but also the advantage of extremely fast streaming transfer speeds when tape-based data is actually needed.

Modern Tape Systems: Management

IBM can provide a couple of other contemporary notable tape capabilities, which contribute to an improving overall technical value proposition for tape in IT operations:

- **Simplified Data Access and Interchange with LTFS:** To improve the access and interchange capabilities of tape, a file system specification for LTO called linear tape file system (LTFS) became available in 2010 and has changed the rules for tape access. IBM's implementation of LTFS—much of LTFS's value comes from it being a standard across vendors—is branded as Spectrum Archive, which is part of the broad (and expanding) IBM Spectrum Storage family. At its simplest explanation, LTFS makes tape self-describing, portable, and accessible like disk—cartridges can be thought of like enormous USB drives!

  IBM's Spectrum Archive implementation of LTFS provides an easy way to archive data to tape without the need for other backup software. The metadata of each cartridge, once mounted, is cached in server memory. Metadata operations, such as browser directory and filename search, do not require physical tape movement. LTFS allows tape data to be accessed in a manner similar to disk or other removable media, and provides the ability to drag, drop, and share data without regard to platform.

- **Intelligent Data Management Software and Appliances Boost Tape Efficiency:** IBM offers analytic software to improve the management insight and performance of its complete tape environment. This capability can optimize cartridge utilization and improve media lifecycle management while also offering proactive tape management to track, manage, and report status on tape library usage and media health from file creation to end of life. These tools identify any potential tape media-related errors and media degradation, eliminating unscheduled downtime and increasing the reliability and availability of tape operations. Moreover, IBM's overall tape environments can now be managed as part of IBM's Spectrum Storage

"ecosystem"—the all-encompassing, global, and flexible software management and integration of users' overall storage.

Tape Systems' Security

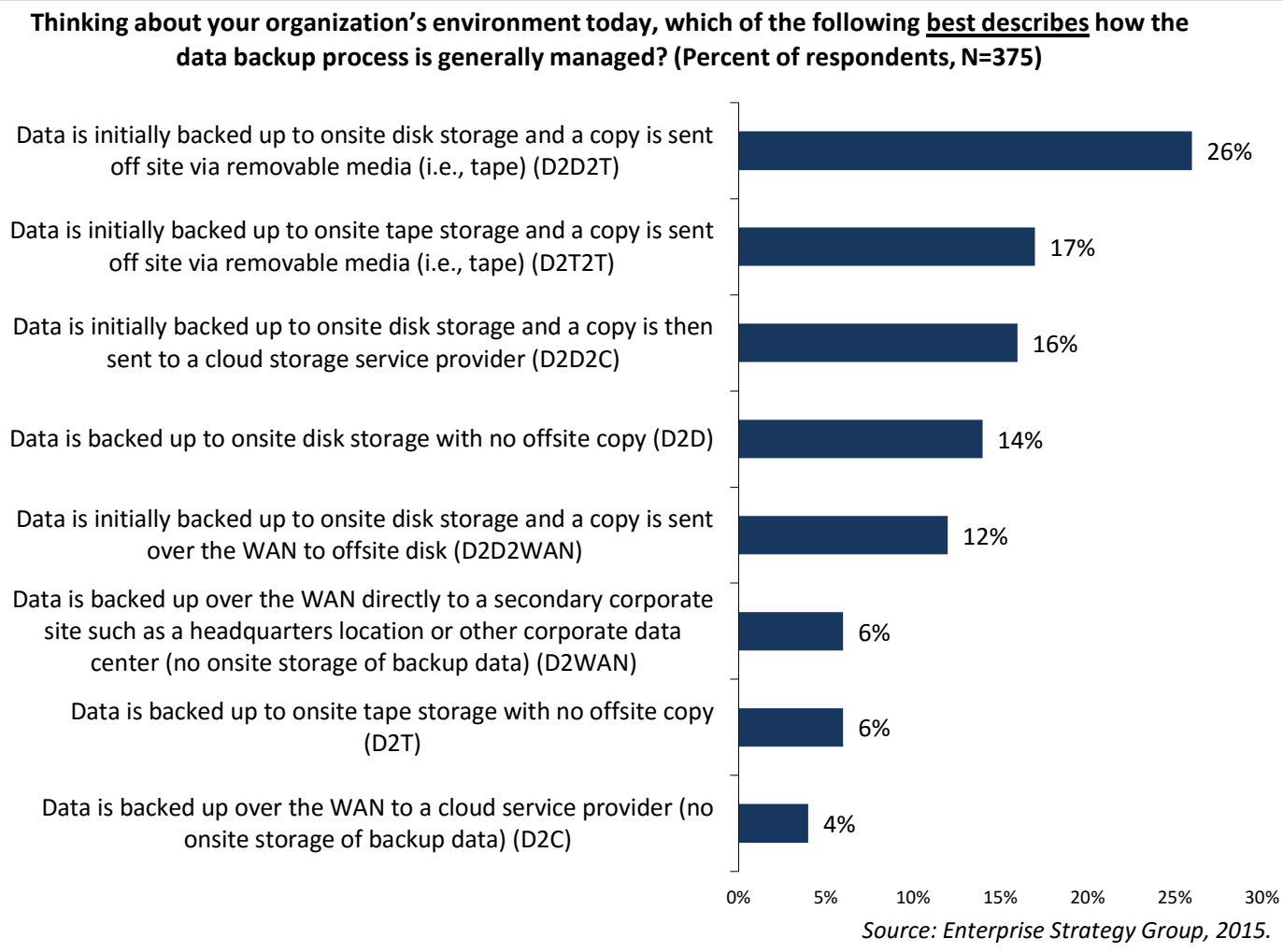
One of the most compelling attributes of tape from a technical perspective is that it is excellent in terms of security across multiple aspects. With data security in general being a top concern for most users, tape represents a multi-pronged defense:

- **Physical:** The combination of tape's fast data rates (up to 360 MB/sec native for the IBM TS1150) and physical mobility is ideal for placing data in highly physically-secure locations.
- **Software:** Tape systems are essentially immune to propagating online or software system glitches.
- **System:** Tape remains the ultimate long-term backup system, with the highest levels of media reliability. It delivers the most straightforward vehicle for DR, while its fast streaming rates make for quicker recoveries when needed.

How Modern Tape Is Utilized for Modern Data Protection and Preservation

Data protection and preservation are synonymous with tape for good reasons; and modern data protection still requires modern tape to this day. As seen in Figure 4, tape is utilized as part of the primary backup processes for nearly half (49%) of all surveyed IT organizations, a number that is higher among enterprise organizations.

Figure 4. Primary Backup Topologies by Media Choice(s)



Source: Enterprise Strategy Group, 2015.

In fact, according to Figure 4, tape is the exclusive media for nearly one in four (23%) organizations' primary backup topology. But it is important to recognize that "tape" in this context is modern tape, with modern performance characteristics that align with modern IT service level expectations for data retention and retrieval.

Tape is the exclusive media for nearly one in four (23%) organizations' primary backup topology.

How Modern Tape Satisfies Modern SLAs for Data Retention and Retrieval

Ignoring presumptive terms like "backup" or "archive," when asked about the behavioral characteristics of their long-term data retention solution, ESG research respondents cited the following solution criteria:[9]

- More than half (58%) of environments add data to their long-term storage solution on at least a daily basis, with another 30% adding data at least weekly.
- Similarly, retrievals occur at least daily in 30% of environments, with another 32% retrieving at least weekly.
- 85% of organizations report that the average age of data IT is asked to retrieve is less than two years old.
- Data retrievals tend to be smaller than many presume, with 35% being less than 1 GB and 75% being less than 100 GB.

Simply put, those kinds of storage behaviors cannot be met with the tapes of the past, but are easily met by today's modern cartridges and drives, while still yielding operational and economic benefits.

The Bigger Truth

The most important realization about "modern tape" is that it is in fact "modern."

From a purely cost-per-TB storage consumption perspective, tape makes sense—and dollars. But it can only do that because the performance and reliability of modern tape formats have greatly surpassed the poor reputations of the tapes of previous decades. With the continually growing capacities of modern tape cartridges, the actual number of media cartridges can decline as the storage efficiency per cartridge increases, thereby further reducing operational management costs for the cartridges themselves. Of course, high-capacity libraries and auto-loaders further reduce the manual handling costs of tape, so performance, portability, and pristineness become the real decision makers.

While disk systems don't have longevity, cloud services are not yet able to ensure the authenticity of the original data without forcing subscribers into a nonnegotiable vendor lock-in to preserve the defensibility of the restoration of data over the entire length of retention that is mandated, regardless of the duration. It is for these reasons, among others, that so many organizations continue to embrace tape as a strategic aspect of their storage management and data preservation strategy.

[9] Source: ESG Research Report, *Backup and Archiving Convergence Trends*, April 2014.

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

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