ESG

Enterprise Strategy Group | Getting to the bigger truth.™

Leveraging Tape for Active Archiving at Scale

How Modern Tape Solutions Accelerate Business

By Christophe Bertrand, ESG Senior Analyst
and Monya Keane, ESG Senior Research Analyst

May 2020

Contents

Market Landscape: This Is a Great Time for Tape Solutions

Nearly a third of senior IT decision makers surveyed by ESG (31%) for its annual technology spending intentions survey named cost reduction as one of the top business initiatives driving their technology spending this year.[1] Essentially, over the next 12 months, they plan to invest some money to avoid paying even more later. The priority is to find any technology that improves their infrastructure's economic efficiency.

That wasn't the only interesting finding uncovered by the survey. Twenty-nine percent of respondents said that solutions able to improve data analytics for real-time business intelligence and customer insight would be another major business initiative driving spending for their organizations. It is likely that those organizations will particularly favor analytics solutions that are available at a great price point. Additionally, these organizations were intent on strengthening their cybersecurity stance—mentioned by 40% of respondents, improving cybersecurity was the most-cited business initiative driving technology spending for 2020.

It is vital now to seek out solutions able to support all of those priorities. IT environments have become very complex, and the reason most frequently reported by ESG research respondents for that complexity is higher data volumes, cited by 37% of respondents. ESG believes that the rising complexity of IT is creating a domino effect. Data is increasing across all tiers, ultimately resulting in a need for more archiving storage capacity (see Figure 1). Notably, new data security and privacy regulations are also causing IT complexity to increase because they enforce required retention periods, which increases the size of organizations' data archives.

Figure 1. The Domino Effect of Ever-rising Data Volumes
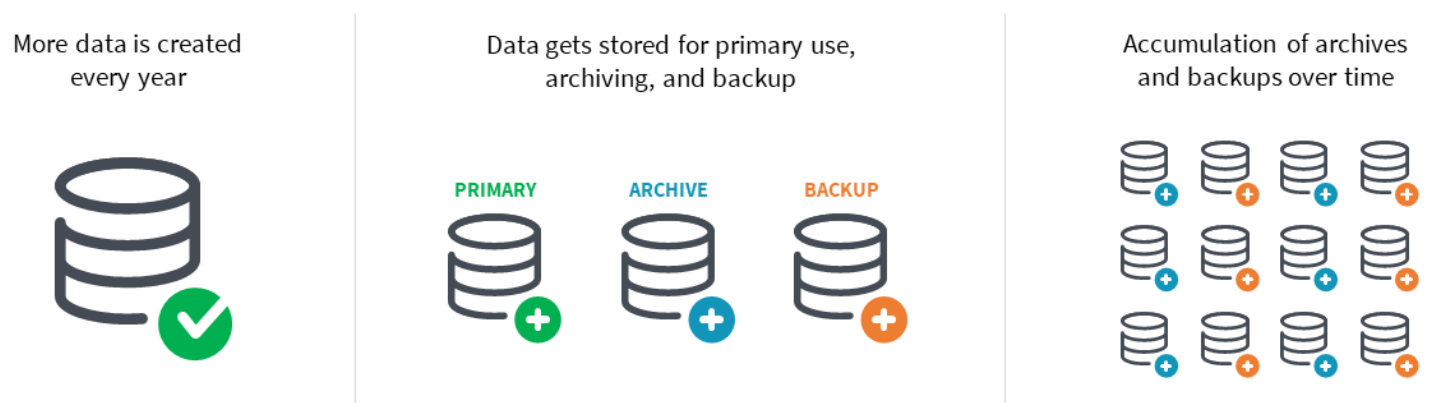
Source: Enterprise Strategy Group

Smart Organizations Are Focusing on Tape

Tape is one means for an organization to overcome the rising complexity and make retaining massive quantities of data easier. ESG found that 67% of the organizations it surveyed use tape for archiving, and 57% use tape for backup. Twenty-eight percent of the respondents told ESG they plan to continue to invest in tape and increase their current tape footprint. Sixty-two percent of the organizations using isolated recovery (i.e., isolated storage capacity) solutions said they'd definitely be able to recover from a cyberattack.[2] Tape is often a hallmark of that kind of isolated topology, as it is especially useful for air-gapping.

---

[1] Source: ESG Master Survey Results, *2020 Technology Spending intentions Survey*, February 2020. All ESG research references and charts in this white paper have been taken from this master survey results set, unless otherwise noted.
[2] Source: ESG Master Survey Results, *2018 Data Protection Landscape Survey*, November 2018.

Separate ESG research found that organizations' most-cited preferred location for sending backup data offsite is a public hyperscale cloud provider (named by 37% of respondents).[3] The cold-storage option that those offsite cloud hyperscalers use for long-term, low-cost archiving is almost certainly going to be a tape-based solution.

As a whole, we are looking at a "perfect storm" for modern tape-based solutions to step up to alleviate the complexities and challenges of today's IT. After all, tape has an historically demonstrated excellent cost profile. It plays an integral role in supporting cybersecurity with isolated air-gapping. And onsite or offsite, it offers virtually endless capacity and scale.

Exploring Tape Technology's Ability to Support Active Archives

The best tape solutions go further, though—beyond bolstering cybersecurity, cost efficiency, and capacity availability. That is because those tape solutions are able to serve a role as *active* archives.

An active archive is a tiered storage topology/solution that gives IT systems or human end-users access to data through a common, unified file system that automatically retrieves and places that data on the appropriate storage tier. As a whole, active archives utilize several media types: SSDs/flash drives, HDDs, cloud storage, and importantly, magnetic tape.

The purpose of an active archive is to optimize placement of data on the most appropriate medium according to specific user-defined parameters (typically related to retention and recovery/restoration patterns), with cost optimization also typically in play. Less time-sensitive and infrequently accessed data is stored on less-expensive media. Active archives also eliminate the need for storage administrators to migrate data manually between storage systems.

According to the Active Archive Alliance,[4] active archives "provide organizations with a persistent view of the data in their archives and make it easy to access files whenever needed. Active archives take advantage of metadata to keep track of where primary, secondary, and tertiary copies of data reside within the system, in order to maintain online accessibility to any given file in a file system, regardless of the storage medium being utilized."

Figure 2 illustrates a typical active archive topology—in this case, a clustered architecture from vendor QStar Technologies. It shows the compute layer and the storage/archive components being driven by the Archive Manager software for writing and recalling of data on the active archive. Tape would be an obvious fit for such an architecture due to tape's inherent characteristics.

Figure 2. Example of an Active Archive's Clustered Architecture
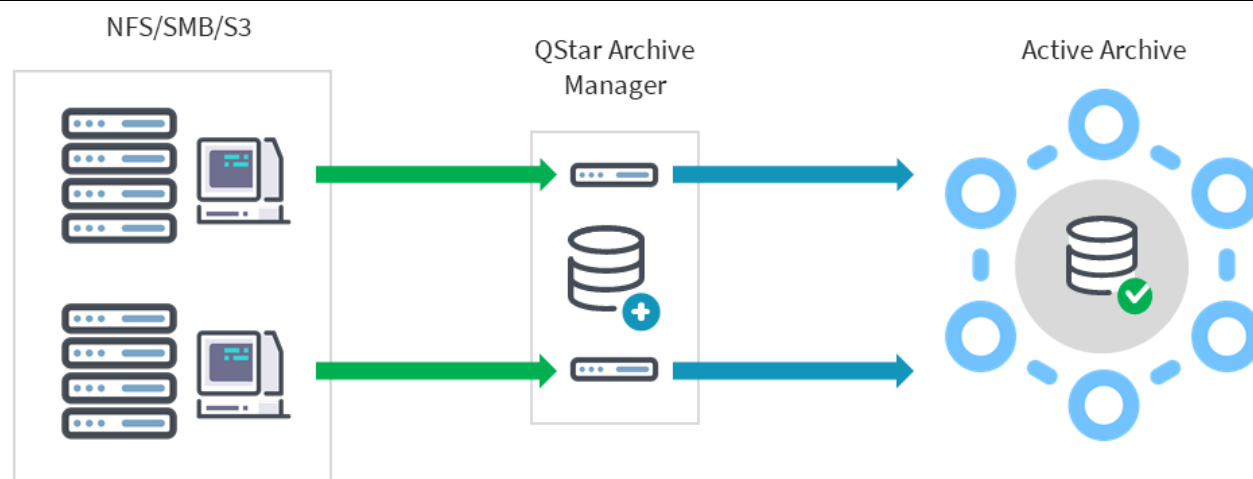
Source: QStar Technologies

---

[3] Source: ESG Master Survey Results, *Data Protection Cloud Strategies*, June 2019.
[4] https://en.wikipedia.org/wiki/Active_Archive_Alliance.

Can Tape Be Used for Big Data?

A perception exists that the only way to deliver deep-learning analytics is through a sophisticated, high-end, real-time data warehouse. While that type of solution certainly does provide real-time capabilities, it also requires a significant investment in hardware resources—an investment that may be too cost-prohibitive for many IT organizations.

Fortunately, leveraging modern tape systems with data management software supports many archiving-related use cases for a much more attractive price. Tape fits nicely into a well-architected solution in which the processes being served don't depend on a very short "time to first byte" (i.e., processes not centered on real-time analysis or parallel and distributed data processing. Those processes do not tolerate much latency, making then inherently dependent on network and system bandwidth.)

Other use cases are more tolerant and well-suited to tape for active archiving at scale. In these cases, random data access is not the key performance factor, so they can be cost-optimized by leveraging tape technology with just minor effects on quality of service. Situations in which a tape-based active archive makes sense include:

- Data quality validation—the process of ensuring data has undergone cleansing to ensure it is both correct and useful.

- Data cleansing—the process of detecting and either correcting or removing corrupt or inaccurate records from a record set, table, or database. The process involves identifying the incomplete, incorrect, or irrelevant data, and then replacing, modifying, or deleting it. Data cleansing may be performed through scripted batch processing .

- Data sampling—a statistical analysis technique used to select, manipulate, and analyze a representative subset of data points to identify patterns and trends in the larger data set being examined.

- Other use cases include analyses of high-dimensionality data, data reduction projects, and active archiving of distributed data sources.

Real-world Example 1: CERN, 'The Best of Both Worlds'

CERN, established in 1954, is a European research organization that operates the largest high-energy particle physics laboratory in the world. The experiments performed by CERN scientists generate a deluge of data that must be efficiently archived for later retrieval and analysis.

The CERN Tape Archive (CTA) is the tape storage system holding the custodial copy of this data. It leverages a tape back-end connected to a disk system that handles disk-cache functions. CTA replaced a previous system that had reached its scalability and performance ceiling.

The star of CERN's mission is the Large Hadron Collider, built between 1998 and 2008 in collaboration with more than 10,000 scientists, hundreds of universities and laboratories, and more than 100 countries. It lies in a tunnel that is 17 miles in circumference and as deep as 574 feet beneath the France-Switzerland border near Geneva. In the collider, particles are accelerated almost to light speed and smashed together as hard as possible. The goal is to see if any new, unknown particles are created. The amount of data generated from this work is staggering. According to CERN, one billion particle collisions per second will generate one petabyte of data … *per second*.

CTA supports CERN's massive data-collection and analysis effort. CERN deployed the active archive platform aiming to provide the best of both worlds by combining disk and tape. A representative of the agency reports that the main goal of CTA "is to make more efficient use of the tape drives to handle the higher data rates anticipated during Run–3 and Run–4 of the Large Hadron Collider."

But as Figure 3 shows, CERN faces some challenges in regard to archival storage over the next decade:

- The exponential growth of the total volume of data due to technology improvements that will allow CERN to collect more data.

- Computing power and disk-capacity constraints that will change the way in which archival storage is used in the experiments.

Figure 3. A Deluge of Data—Predicted Tape Archival Storage Needs for CERN
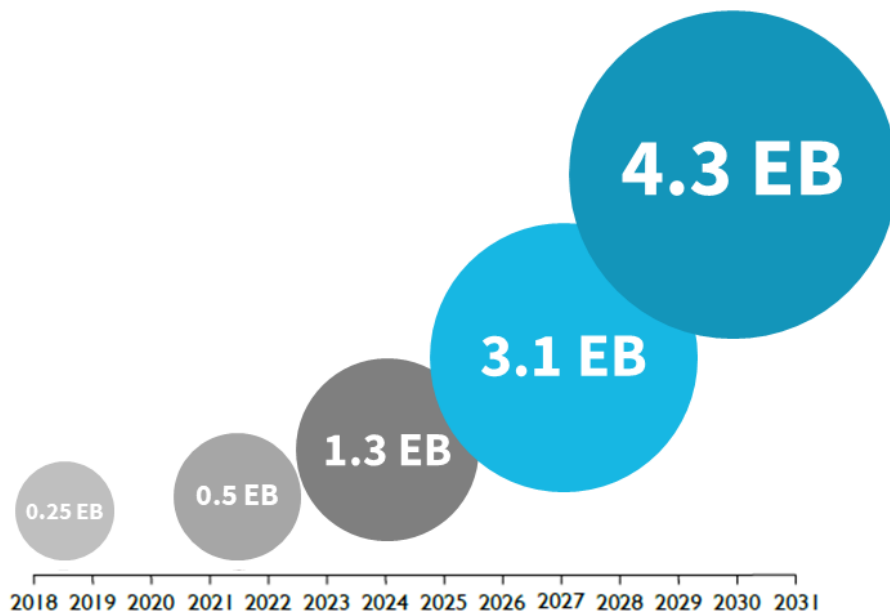
Source: CERN Tape Archive, from development to production deployment, EPJ Web of Conferences 214, 04015 (2019)

To mitigate the impact of this tsunami of data and ease the pressure on storage and budgets, leveraging tape is now CERN's focus. CERN understands that tape is a much more economical form of storage than disk. The cost savings from the increased use of tape will allow the experimenters to close the gap between resource needs and budget.

CERN is leveraging the best of both worlds (the worlds of disk and tape) by wisely using the right underlying technologies to optimize that combination and ultimately ensure scalability, performance, and cost mitigation are maintained.

Real-world Example 2: A Major Educational Institution

This institution is one of the most competitive and prestigious universities in the United States. ESG spoke with one of the university's IT executives to get a sense of its approach and the key benefits it is getting from its tape-based active archive solution.

According to the IT leader, the university built an active archive with tape technology because it wanted a "known entity"—something that would be easy to implement and easy to move to accommodate future unknowns.

In terms of benefits the institution has experienced, the IT leader reports that the IT organization did not have to buy more expensive disk to support "the unquenchable appetite of creatives for being always online. Enough is never enough."

For the university, tape was a particularly good option because it was averse to using cloud storage. The IT manager reported that the university believes "the cloud is still not ready for nearline storage due to the open checkbook it would require."

When offering a recommendation to peer organizations looking to implement a tape-based active archive, the IT manager said, "It depends on what they are trying to accomplish. We are using tape for nearline storage, not disaster recovery. If 'active archive' equals 'nearline storage,' then that is what I recommend."

The Benefits of a Tape-based Active Archive

Clearly, a tape-based active archive provides numerous benefits: scalability, good performance, the ability to support a broad archiving ecosystem, and operational efficiency.

Of course, the economics of tape are especially appealing. ESG research and financial modeling conducted for Linear Tape-Open (LTO)-based solutions shows dramatic savings in hardware, media, staff, and maintenance costs can be achieved by leveraging tape over disk. In fact, ESG's analysis of a typical large-scale data retention use case showed nearly $13.5M in estimated cost savings over a ten-year time horizon, with an additional $400,000 in incremental user benefit delivered over and above what is expected with a disk-based alternative. The result was an impressive 577% return on investment over ten years.[5]

But perhaps reliability is a tape-based active archive's most compelling characteristic. LTO tapes have an archival life exceeding 30 years. They can support a million "passes" and 20,000 write cycles per tape, and they have a Mean Time Between Failure (MTBF) rating of 250,000 hours at a 100% duty cycle. Additionally, data integrity technology is built in with block-level checksums.

The Bigger Truth

Tape is not dead. In recent years, actually, it has made significant leaps and bounds in terms of performance. Its cost profile and scale cannot be beaten by alternative technologies. And ESG's research shows that, if current data growth patterns continue, tape *must* be considered for nearline storage and for active archives.

It's not the answer to everything, but tape can help organizations that are struggling with massive data storage, analysis, and recovery requirements. It's about accessibility. And accessibility is not synonymous only with cloud or disk storage.

That's why every vendor in the backup and recovery space is working with tape vendors to bolster the long-term archive layer of their solutions. Tape is not only here to stay; it can perhaps offer an even bigger positive impact than ever before.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

www.esg-global.com   contact@esg-global.com   508.482.0188

[5] Source: ESG White Paper, *Analyzing the Economic Value of LTO Tape for Long-term Data Retention*, February 2016.