



ACTIVE Archive

SPECIAL REPORT

November 2025

PREPARING FOR TOMORROW'S EXPANDING STORAGE CHALLENGE WITH ACTIVE ARCHIVE

The **Active Archive Alliance** serves as a vendor-neutral, trusted source for providing end users with technical expertise and guidance to design and implement modern active archive strategies that solve data growth challenges through intelligent data management.

TABLE OF CONTENTS

Table of Contents	2
Foreword	2
Introduction to Active Archive.	3
Active Archiving Reduces Data Center Energy Demand	4
Security Benefits of an Active Archive	5
Tape Enhances Active Archiving with Performance and Capacity Improvements	5
LTO-10 Arrives to Cost-Effectively Contain Soaring Secondary Storage and AI Demand	6
HDDs Are Key for Active Archives	6
Limited Metadata Challenges and Solutions for Unstructured Data	7
Growing Challenges for Infinite Archives Arrive—Technology Obsolescence, Cost, Excessive Conversion Time, and Data Migration	7
Active Archive Enhances Hybrid Cloud Storage	9
Applications for Active Archives.	9
The Data Lifecycle, Active Archives, and AI.	10
New Secondary Storage Tiers Arrive for a Data-Driven Future	11
■ Let's Think Ahead: Will Storage Move To Outer Space?	11
Framework of Requirements for Long-Term Archive	12
Conclusion	13

FOREWORD

This 2025 special report from the Active Archive Alliance highlights the growing importance of active archiving in managing the exponential increase in data volume and value. The report underscores the far-reaching impact of active archives, particularly in the context of long-term data retention and protection.

With the rise of AI, demands on IT infrastructure, including storage, power, and cooling, are rapidly intensifying. These demands create pressure to control costs, reduce energy consumption, and minimize carbon footprints. To meet these challenges, organizations must shift from short-term, reactive data projects to strategic, sustainable data architectures. Active archiving is a key solution for intelligent data management, offering benefits across industries.

The Alliance encourages collaboration and adoption of active archive solutions to build a sustainable, scalable, and efficient data storage environment for the future.

Foreword by:



Rich Gadomski

Director of Channel Sales and New Business Development, FUJIFILM North America Corp., Data Storage Solutions, Co-Chairperson, The Active Archive Alliance



Eric Polet

Director of Product Marketing, Arcitecta, Co-Chairperson, The Active Archive Alliance

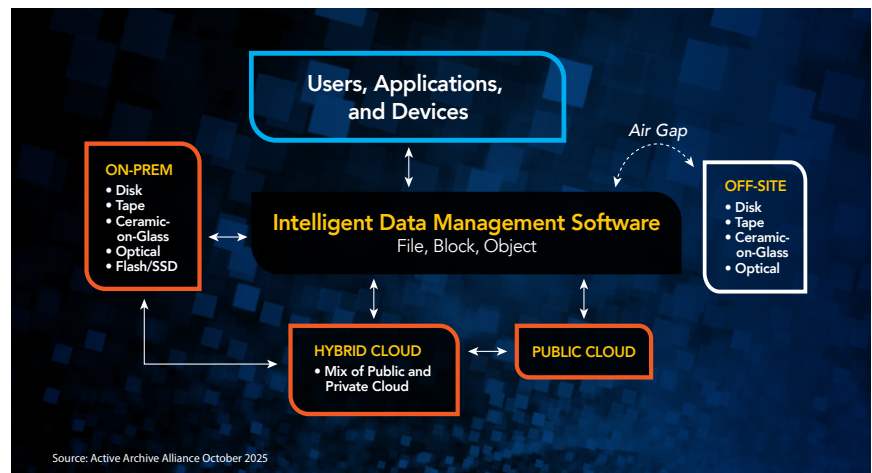
INTRODUCTION TO ACTIVE ARCHIVE

Traditionally, archival data referred to older, rarely accessed data retained for compliance or future reference, but that definition is rapidly changing. Today, archival data has evolved into a more dynamic, high-value resource. The exponential growth of data is overwhelming traditional storage topologies, making it difficult for organizations to manage and retrieve data efficiently. As organizations increasingly require fast access to vast amounts of historical data, the concept of an active archive has become a cornerstone of long-term and infinite data preservation strategies. Artificial Intelligence (AI), business intelligence, healthcare, and scientific research are examples of industries that mine archival data for insights previously overlooked. This shift is transforming data archives from passive repositories into active engines for innovation and discovery. As this trend accelerates, businesses will need to rethink their data management strategies and infrastructure to support cost-effective, scalable preservation solutions.

For many organizations, at least 80% of digital data falls into the low-activity or inactive category — making it the most significant and fastest-growing class of stored data. Historically, this data has gone largely untapped, but AI is revealing its hidden value. Companies aiming to stay competitive must recognize the strategic role that archival data can play in their long-term success. To meet future demands, the archival storage paradigm must reinvent itself—moving from static storage to intelligent, accessible, and monetizable data ecosystems.

An active archive addresses the limitations of traditional physical archives, providing fast access times to secondary (archival) storage systems. Data in an active archive is always online and easily accessible, allowing for quick retrieval and analysis. When the data becomes inactive, it returns to archival status. Fueled by the rapid growth of AI, previously low-activity or cold data can suddenly become active for prolonged periods of time, enabling the building and training of AI algorithms. An active archive intelligently manages data throughout its lifecycle, allowing organizations to balance the need for immediate access to archival data.

Today's active archive typically integrates two or more storage technologies (SSD or HDD, tape libraries, or optical disc and public and private cloud storage) behind Intelligent Data Management Software (IDMS) operating in the background to provide a seamless means to manage archive data in a single virtualized storage pool. IDMS refers to platforms or systems that use advanced technologies—like AI, machine learning (ML), and automation—to manage, organize, and optimize data throughout its lifecycle. These tools go beyond traditional data management by adding intelligence to how data is stored, accessed, protected, and utilized. For example, smart storage optimization moves data between storage tiers (e.g., hot, warm, cold) based on access frequency and business value.



Intelligent data management software is the key component for an active archive.

Note that high-capacity nearline HDDs are used as an active archive without tape; however, this approach can become costly as archives increase in size, requiring more HDDs and, consequently, significantly more energy. Using metadata and global namespaces, the data management layer of an active archive enables data to be readily searchable, accessible, and retrievable, regardless of the storage platform or media on which it resides. See the Active Archive conceptual view above.

ACTIVE ARCHIVING REDUCES DATA CENTER ENERGY DEMAND

In the United States, power consumption by data centers is on course to account for almost half of the growth in electricity demand between 2025 and 2030. Primarily driven by AI and cryptocurrency use, the US economy is expected to consume more electricity in 2030 for processing data than for manufacturing all energy-intensive goods combined, including aluminum, steel, cement, and chemicals. Electricity demand from AI-optimized data centers is projected to more than quadruple by 2030, reaching 945 terawatt-hours (TWh), which is slightly more than Japan's entire electricity consumption today.

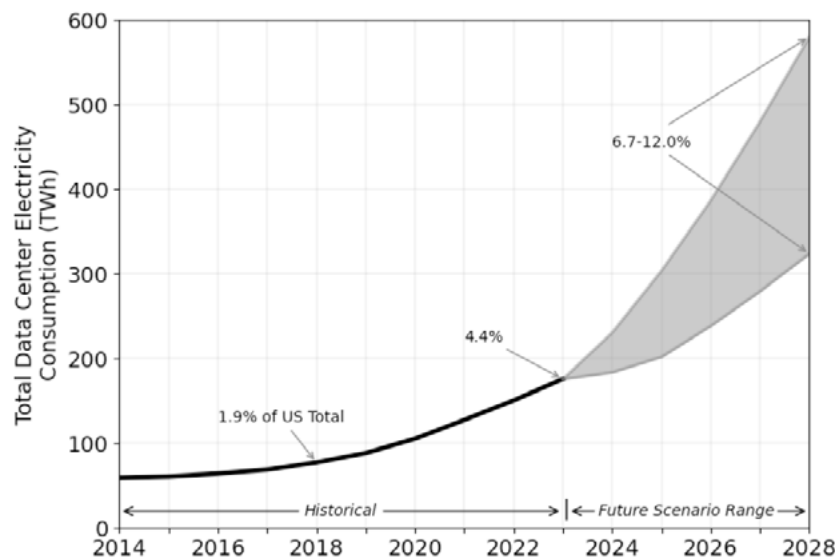


figure 1 Provides an estimate of total U.S. data center electricity use.

Source: Berkeley Lab, 2024 U.S. Data Center Energy Usage Report, Dec. 2024

“Driven by the explosive growth of AI, global data center power demand is now increasing five times faster than grid capacity (IEA, 2025). At the same time, data center operators are rediscovering tape for its unmatched ability to store exabytes of data with minimal energy use. Recent tape technology breakthroughs have firmly reestablished tape as an essential tier in both on-premises and cloud environments—delivering maximum efficiency for workloads such as infinite archives, regulatory compliance, backup, and disaster recovery.”

Ted Oade

Director of Product Marketing, Spectra Logic

- Data centers consumed about 4.4% of total U.S. electricity in 2023.
- Domestic Energy Usage from Data Centers Expected to Double or Triple by 2028.
- Projected to consume approximately 6.7 to 12% of total U.S. electricity by 2028.
- Servers and HDDs are the biggest data center energy consumers.

“Active archives based on object storage with tape are becoming increasingly important in the context of data-intensive AI workloads, especially for data sets that must meet compliance and legal requirements. They also make a significant contribution to reducing storage and energy costs.”

Thomas Thalmann

CEO, PoINT Software & Systems GmbH

In today's data centers, every watt and square foot counts—and an active archive can make a significant impact. The growing demands of AI computing are pushing data centers and energy grids to their limits. From a storage hardware perspective, hard disk drives (HDDs) consume a substantial amount of electricity to operate and cool. In contrast, magnetic tape is the most energy-efficient storage medium for data centers available today. By migrating low-activity or inactive data from HDDs to tape, organizations can reduce carbon emissions by up to 97% compared to a strategy that uses only HDDs. This approach is supported by the research report “Improving Information Technology Sustainability with Modern Tape Storage” by Brad John's Consulting, which compares three scenarios: an all-HDD solution, an all-tape solution, and an active archive strategy that moves 60% of low-activity HDD data to tape. Over a 10-year period, this active archive approach resulted in a 58% reduction in carbon emissions! This is quite impressive, but will it be enough?

The report also evaluates CO₂e emissions from energy consumption, raw material acquisition, manufacturing, and end-of-life disposal of storage media. As secondary storage demand continues to grow, a tape-based active archive is essential for achieving data center sustainability goals. Shifting low-activity data from HDDs to an active archive not only supports environmental initiatives but also delivers substantial financial benefits.

SECURITY BENEFITS OF AN ACTIVE ARCHIVE

An active archive strengthens the security of archival data by incorporating advanced features such as encryption, data immutability, and an air gap, along with access and authentication controls. Ransomware has been around for years and remains the number one IT security concern. Ransomware in 2024 took things to a new level, as 58% of businesses had to shut down operations after an attack for an average of 12 hours. This was accompanied by 5,289 reported ransomware incidents worldwide, a 15% increase from 2023 and more than double the total in 2022. The average ransom demand in 2024 was \$2.73 million, almost an increase of \$1 million from 2023. The true scale of the crisis may be obscured, as many victim companies reportedly refuse to disclose when or if they pay attackers. A successful ransomware attack can result in data loss, business interruption, revenue loss, fines, goodwill impairment, and legal fees.

Data encryption is available on SSD, HDD, and tape products and converts data from a readable format into an unreadable, encoded format called ciphertext. Users and processes can only read and process encrypted data after it is decrypted with a special key. The decryption key is secret, so it must be protected using Access Control Lists (ACLs) against unauthorized access. ACLs are made up of rules that either allow or deny access to a computer or network environment. Because archival data is typically static and unchanging, administrators can use WORM (Write–Once, Read–Many) storage devices for immutability to prevent data from being deleted or overwritten, and to safeguard data integrity, availability, and confidentiality. Proper archiving ensures that data remains unchanged and traceable over time.

“Tiering is necessary to group large datasets and assign them levels of importance and priority. An active archive serves this purpose well, as it allows data to be relegated to a lower tier while still being available rapidly should it be needed by the AI engine.”

Eric Polet

Director of Product Marketing, Arcitecta

“Tape libraries offer a highly reliable and cost-effective solution for securely storing or archiving large volumes of data over the long-term. Modern tape storage systems combine exceptional data protection with significant cost advantages compared to other archival technologies.”

Marc Steinhilber

CEO, BDT Media Automation GmbH

TAPE ENHANCES ACTIVE ARCHIVING WITH PERFORMANCE AND CAPACITY IMPROVEMENTS

The benefits of active archives are far-reaching and can significantly affect businesses of all sizes. However, typically, the larger the archive, the more benefits are realized with an active archive. AI storage demand is soaring and generates valuable datasets that often become less frequently used after initial processing. The accessibility of archived data is critical to more quickly build and refine models and maximize the value of AI investments without the inefficiencies and high costs associated with traditional storage methods.

HDDs and SSDs provide the “active layer” of an active archive and have much faster access times than tape to the first byte of data. The average access time of a typical SSD is <.1 ms., and an HDD is roughly 8.5 ms. For large files, tape systems have faster access times to the last byte of data. Previously available for IBM enterprise tape drives, oRAO (Open Recommended Access Order) arrived with LTO-9 full-high tape drives. oRAO reduces initial file access time, serving as an archival data retrieval accelerator that enables applications to retrieve non-consecutive (random) files from tape by optimizing physical seek times between files. oRAO can improve random access time to data segments on tape by as much as 73% when compared with linear retrievals of the same data segments while significantly reducing physical tape movement and drive wear by creating an optimally ordered list of files on a cartridge.



LTO-10 ARRIVES TO COST-EFFECTIVELY CONTAIN SOARING SECONDARY STORAGE AND AI DEMAND

The latest tape industry report reveals a record 176.5EB of total tape capacity (compressed) shipped in 2024, representing a 15.4% growth over 2023. This growth is a result of evolving infrastructure requirements fueled by the continued implementation of modern technologies, such as AI and ML, which have contributed to significant growth in unstructured data and shifts toward lower-cost hybrid cloud environments. The Linear Tape-Open (LTO) Consortium continues to advance tape technology in capacity and performance. In 2025, it launched LTO-10, the tenth generation of the industry's most widely adopted tape storage format. The LTO program's most recent LTO technology roadmap extends the LTO Ultrium standard through 14 generations. The roadmap calls for tape capacities to double with each new generation, with LTO-14 delivering up to 1,440 TB or 1.44 PB compressed per cartridge.

HDDs and tape most often combine to host an active archive, offering features to boost its performance. HDDs and SSDs have significantly faster access times than tape to the first byte of data and are most often used as the cache buffer.

RAIT, like RAID for HDDs, serves as a tape data rate multiplier enabling high parallel data transfer rates from an array of tape drives. Faster access time and throughput capabilities position LTO for AI and ML applications that are increasing access to archival data, creating a dynamic active archive.

"The resurgence of LTO tape technology is driven by AI's insatiable appetite for training data storage, hyperscale operators seeking cost-effective cold storage solutions, and enterprises implementing active archive strategies that balance accessibility with long-term retention economics."

Pete Paisley
Owner, MagStor

"In today's era of AI and exponential data growth, HDDs are more critical than ever - delivering the durability, capacity, and economic advantage needed to store, scale, and manage massive data volumes with confidence. These attributes make HDDs foundational to active archive strategies, where long-term data remains readily accessible for ongoing analysis, innovation, and operational agility."

Scott Hamilton

Senior Director, Product Management,
Marketing & CX, Western Digital

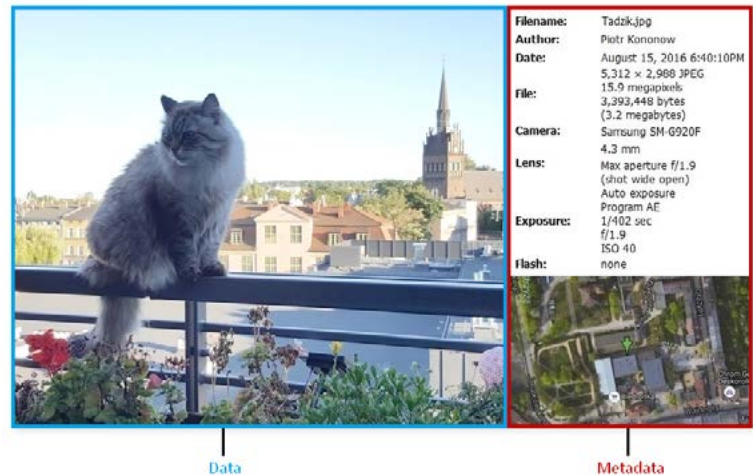
HDDs ARE KEY FOR ACTIVE ARCHIVES

HDDs have been the workhorse of the storage industry, currently representing over 80% of installed data center storage capacity. Nearline HDDs are ideal for storing data that is accessed infrequently but must remain readily available and represent the "active" component in an active archive. HDDs are used heavily for storing massive datasets used in AI training and inference. HDDs continue to increase capacity innovation with HAMR (Heat-Assisted Magnetic Recording), potentially enabling capacities to reach 50 TB per drive, SMR (Shingled Magnetic Recording), which overlaps data tracks like roof shingles to increase density and power efficiency, and MAMR (microwave-assisted magnetic recording) using microwave fields to assist data writing without heating the disk.



LIMITED METADATA CHALLENGES AND SOLUTIONS FOR UNSTRUCTURED DATA

Organizations face significant challenges in extracting insights from archival data, most of which is unstructured data (limited or no metadata) and scattered across multiple diverse storage systems. Metadata is used to describe, organize, find, and manage data effectively. A metadata file contains descriptive information about other data. Data classification, identification, and tagging are procedures describing data based on its relevance to the enterprise, making it easily searchable and trackable. Metadata can be assigned manually; however, the ideal data classification software generates metadata upon ingest, creating a detailed inventory of data assets to facilitate quick storage and retrieval.



This conceptual illustration of metadata above shows how a digital file or image is surrounded by descriptive tags, such as author, date, file type, camera used, and more. Metadata helps both the original researcher and others to interpret, cite, and build upon the data.

Massive amounts of content are required – ideally tagged with metadata – for AI to function effectively. By enriching data with descriptive metadata, AI systems can process, categorize, and analyze information more swiftly and accurately. This approach enhances the capabilities of AI-driven applications to locate data to build and refine reference models. Common metadata examples include file creation date, author, file type, and location. Metadata standards exist to ensure consistency. As organizations increasingly rely on data for strategic decision-making, the ability to migrate, preserve, and access archival data efficiently will become a defining factor in long-term success emphasizing the role of metadata.

GROWING CHALLENGES FOR INFINITE ARCHIVES ARRIVE—TECHNOLOGY OBSOLESCENCE, COST, EXCESSIVE CONVERSION TIME, AND DATA MIGRATION

The dawn of the infinite archive is here, making long-term data retention a necessary discipline. More applications are requiring permanent (forever) retention times (see chart on right), redefining the scope and magnitude of data archiving. Every application in this chart is a candidate for an active archive. Unlike traditional cold archives, which are often static and difficult to access, active archives offer dynamic options that ensure data remains accessible and secure. Movies, financial records, patents, news, sports events, legal, medical, and pharmaceutical data, as well as government and national security data, are a few examples of data that are being retained in a digital format forever.

As organizations increasingly rely on electronic file storage, a significant challenge arises in ensuring that data remains accessible and searchable over decades — sometimes even centuries. Metadata plays a crucial role in this process, enabling faster retrieval of archived content regardless of its age.

One of the most overlooked risks in long-term archiving is the risk of format obsolescence. A document saved in 2025 as a Word file may not be readable in 2075 due to changes

Record Types	Retention Period
Accounts payable ledger	7 years
Accounts receivable ledger	7 years
Audit reports of accountants	Permanently
Bank statements	7 years
Capital stock and bond records	Permanently
Charts of accounts	Permanently
Contracts and leases	Permanently
Correspondence (legal)	Permanently
Deeds, mortgages, bill of sale	Permanently
Employee payroll records	Permanently
Employment applications	3 years
Inventory records (products)	7 years
Insurance records	Permanently
Invoices to customers	5 years
Invoices from vendors	5 years
Patents	Permanently
Payroll records and tax returns	7 years
Purchase orders	5 years
Safety records	6 years
Time cards and daily reports	7 years
Training manuals	Permanently
Union agreements	Permanently

in software ecosystems, vendor discontinuation, or lack of backward compatibility. This reality underscores the importance of selecting open, standardized, and well-documented formats for archival purposes. Storage hardware also has limited lifespans, requiring migration to new, more efficient technologies. For example, hard disk drives (HDDs) typically last 4–6 years before replacement or failure. Data retained for 100 years would require at least 10 migrations to newer media. Each migration is time-consuming, and introduces risks of data loss, corruption, and increased operational costs.

As the amount of data steadily grows, each migration takes more resources to complete. To mitigate these risks, many enterprises adopt a rolling migration strategy, updating a portion of their infrastructure annually. This approach spreads the workload and reduces the risk of large-scale failures during mass migrations. Data often outlives the systems designed to store it. A sustainable long-term archiving strategy must account for metadata integrity, format compatibility, hardware refresh cycles, and proactive migration planning.

“Organizations will continue to use hybrid models to balance cost, performance, and security, storing less sensitive content in the cloud while keeping high-value assets with strict compliance requirements in more highly secured environments, including on-premises.”

Paul Luppino

Director, Global Digital Solutions, Iron Mountain



“AI training data, models, governance, and generated content need somewhere to live. For AI to thrive, data needs to be accessible and retained in a sustainable and affordable data storage solution.”

Christian Pflaum

CEO and Co-Founder, Cerabyte

Sometimes, data can go missing during the migration process. This mishap can be due to format incompatibilities, automatic truncation, unknown validation settings, and network interference, among other factors. A recent study by Experian revealed that 64% of the data migration projects they analyzed went over budget, and only 46% of projects were delivered on time. Less than 70% of projects were deemed successful. Key challenges and risks to unlocking the value of archival data are:

- 1 making archival data accessible at ingest (by assigning metadata classification, index, and catalogs)
- 2 managing the long-term archival storage infrastructure
- 3 ensuring that only the potentially needed archive data is stored
- 4 ensuring the security and availability of archival data.

ACTIVE ARCHIVE ENHANCES HYBRID CLOUD STORAGE



HYBRID CLOUD

- Combination of both public and private cloud
- Shared security responsibility
- Helps maintain tighter controls over sensitive data and processes

Active archiving enables organizations to harness the full potential of large datasets stored across multiple storage systems and multiple locations. Using an active archive for hybrid clouds is a modern data management solution designed to bridge on-premises and cloud environments, enabling seamless, long-term storage and access to archival data. Unlike traditional archives, which are static and difficult to retrieve, active archives

are dynamic, searchable, and optimized for performance and cost efficiency across hybrid infrastructures. Data is enriched with metadata at the time of ingestion, enabling fast search, classification, and retrieval regardless of where the data physically resides, on-premise or in the cloud.

Frequently accessed data can remain on-premises or in high-performance cloud tiers. In contrast, colder, less-accessed data is automatically moved to low-cost cloud storage. A unified or global namespace provides a single, logical view of data across both on-prem and cloud locations, allowing users and applications to access files without needing to know where they are physically stored. Hybrid cloud architectures provide built-in redundancy and disaster recovery capabilities, ensuring data durability and availability even in the event of hardware or cloud service failures.

"According to the 2025 Wasabi Cloud Storage Index report, although many organizations think they will never access data stored in low-cost, deep archive tiers, the reality is most organizations (84%) end up accessing this data on a weekly or monthly basis. And with fees for access and API calls, customers spend more on fees than on actual per GB storage costs."

George Hamilton

Director of Product Marketing, Wasabi

APPLICATIONS FOR ACTIVE ARCHIVES

Active archives are becoming increasingly vital for organizations that manage large volumes of data, as they need to be accessible for analytics, compliance, or operational purposes. Below is a breakdown of applications and industries where active archives shine, along with leading platforms:

- **Compliance & Legal Discovery**
Retaining emails, contracts, and communications for regulatory audits and legal holds.
- **AI/ML & Big Data Analytics**
Leveraging historical data for training models or longitudinal analysis.
- **Healthcare Imaging & Genomics**
Long-term storage of medical images, genomic data, and patient records.
- **Video Surveillance & Media**
Storing and retrieving high-resolution video and multimedia content efficiently.
- **Medical Research**
The ability to identify potential drug targets, predict drug interactions, and optimize clinical trials helps to identify disease markers earlier, develop new therapies
- **IoT & Sensor Data**
Managing massive streams of time-series data from devices and sensors.
- **HPC (High Performance Computing) Environments**
- **Universities**
- **National Laboratories**

"We have seen many customers move towards a hybrid active archive, especially for large media archives in the US. On-premises storage provides high performance and cost-effectiveness and, when combined with cloud object storage, the solution provides a high level of data protection."

Phil Storey

CEO, XenData

THE DATA LIFECYCLE, ACTIVE ARCHIVES, AND AI

From Cold - to Hot - to Cold

Primary Storage
Active Online Data
~20% of Data

Secondary Storage
Cold, Archival, Persistent Data
~80% of Data

Active
Hourly, Daily

SSD, HDD

Active Archive

Tape

Cold Data That Becomes Hot for a Brief Time

Amount of Stored Data

Colder-->

Warmer-->

Dynamic Archive Data Lake

Deep Archive

Dark Data

Inactive
Monthly, Annually, Never...

Probability Of Access $P(a)$

100%

75%

50%

25%

0-5%

Days Since Last Access

0

1 week

30 days

90-120 days

1+ years

10+ years

100+ years

Hot

Cold

Life Cycle Profile Factors
Probability of access $P(a)$ declines as data ages.

Most data typically becomes **cold/archival** in ~90 - 120 days.

Cold/archival data can become **hot** at any time and stay hot for variable time periods. (Dynamic Archive).

Archival retention can be >100 years to ∞

Source: Horison Information Strategies, LLC

Storing less frequently accessed AI datasets in an active archive ensures that even as data ages, it remains readily available for future processes such as model retraining, new model development, audits, and compliance checks. As data volumes continue to grow, the active archive becomes a foundational component of modern data infrastructure —empowering AI-driven enterprises to unlock the full potential of their data.

Business Line Executive, IBM DRI Tape Solutions



NEW SECONDARY STORAGE TIERS ARRIVE FOR A DATA-DRIVEN FUTURE

The storage technology hierarchy has traditionally been illustrated as a pyramid, with the fastest and most expensive technologies—such as SSDs—at the top, and higher-capacity, lower-cost solutions—like tape—at the bottom. Today, approximately 20% of the world's data is always active, best served by primary storage technologies such as SSDs and HDDs. The remaining 80% consists of lower-activity, archival and cold data, which is more efficiently and cost-effectively stored using secondary solutions, such as tape and, in some cases, nearline HDDs.

As the volume of secondary storage data continues to soar, two new secondary storage tiers have emerged to complement the traditional archive tier, as advanced learning applications utilizing AI, ML, and Big Data analytics have reinvigorated the archives. These applications require faster access to analyze archival data, but for shorter periods of time, as the data returns to archival status after analysis. The secondary storage tiers include the Active Archive, Traditional Archive, and Deep Archive, each designed to meet specific performance, accessibility, and cost-efficiency requirements.

Archive Tier	A New Secondary Storage Model Emerges	Primary Mode
Active	<i>Write Once Read Many</i> - Online access to dynamic "active archival data"	WORM
Archive	<i>Write Once Read Seldom if Ever</i> - Traditional primary tier, lower activity, big data and data lakes, cold archives	WORSE
Deep	<i>Write Once Read Never</i> - Emerging technology tier for permanent, rarely accessed, dark data, often unclassified or untagged, the "golden or master copy", time capsule copy	WORN

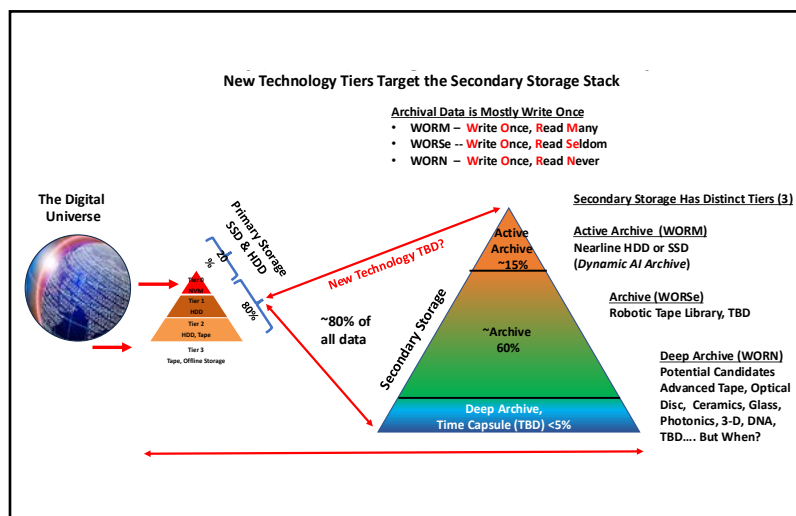


figure 3 Secondary (Archival) Storage Model—2025 and Beyond

Source: Horison Information Strategies, LLC



LET'S THINK AHEAD Will Storage Move to Outer Space?

As the space economy rapidly expands, a bold new frontier is emerging in data management: active archives in orbit. There is growing demand for systems capable of processing, storing, and archiving data directly on satellites, effectively extending the cloud into outer space. Today, more than 12,000 satellites orbit Earth, supporting a wide range of functions including communication, navigation, remote sensing, and scientific research.

To meet the demands of space-based data storage, non-volatile technologies such as MRAM, FRAM, and advanced flash memory are being developed to withstand cosmic radiation and extreme temperatures. Satellites equipped with onboard AI can classify, compress, and tag data with metadata in real-time, creating active archives that are searchable and usable without requiring Earth-based intervention. As satellite technology, edge computing, and AI converge, we may soon witness the emergence of a new tier in the cloud ecosystem — a celestial archive layer designed for the digital age.

Typically, about 15% of archived data may be active at any given time, supporting rapid analysis and decision-making. This figure could change with AI. Once analysis is complete, unmodified data is typically deleted from the active archive, while any modified data is rewritten to traditional archive storage. All three tiers are gaining momentum as a new secondary storage model strives to meet insatiable demands. The emerging Deep Archive and Time Capsule tiers will likely be addressed by new technology, targeting data to be stored forever (a golden, immutable copy) but which may be accessed rarely, if ever.

Meanwhile, emerging technologies, including advanced tape formats, higher capacity HDDs, ceramic-based storage, optical discs, and synthetic DNA-based storage, are in various stages of laboratory development. These secondary storage innovations may represent the future of ultra-dense, long-term data preservation, although they are not yet commercially viable. Ideally, about 80% of all data is most effectively stored in secondary storage.

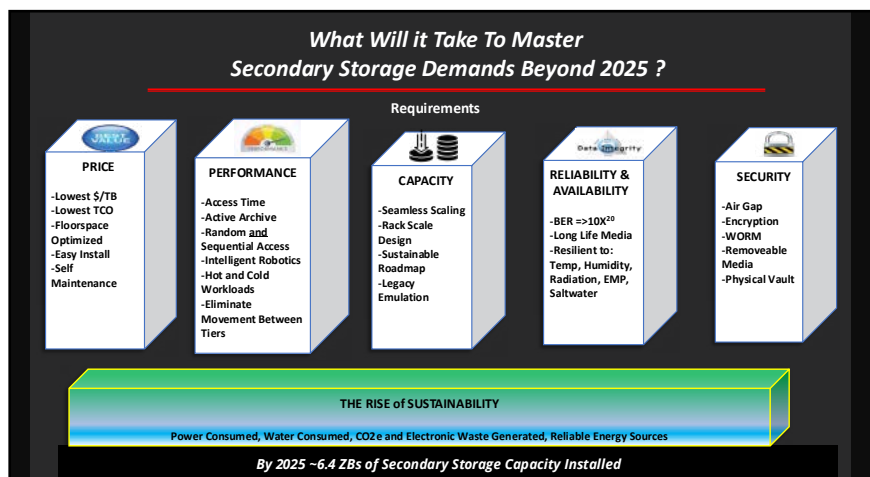
FRAMEWORK OF REQUIREMENTS FOR LONG-TERM ARCHIVE

For decades, HDDs and tape have combined to share the burden of secondary (archival) storage capacity, with tape remaining the most economical and environmentally sustainable option. However, neither technology alone is ideally positioned to meet the enormous future demands, as it takes both HDDs and tape to create the optimal active archive.

The need for a transformative secondary storage model to keep data actionable for decades to centuries is building. As a result, an advanced secondary storage model is expected to evolve supporting massive storage and archival requirements while controlling costs related to systems, power, cooling, and minimizing environmental impact. The rapid growth of digital data, further fueled by AI, is driving the need for more advanced petascale and exascale-capable secondary storage systems.

	Random Access	Fastest Access Time	Fastest Data Rate	Lowest Energy Consumption	Air Gap	Lowest TCO
HDDs	Yes	Yes	No	No	No	No
Tape	No	No	Yes	Yes	Yes	Yes

Future storage demand has signaled requirements for advanced secondary storage and active archive solutions that can deliver the required performance, scalability, and efficiency as depicted below.



“As global data volumes approach 180 zettabytes, traditional storage architectures are becoming economically and environmentally unsustainable. Active archive solutions deliver significant cost reductions and operational efficiency improvements while providing the immutable, energy-efficient foundation necessary for AI development at scale. The question is not whether to adopt active archive technology, but how quickly organizations can transform data storage from a growing liability into a competitive differentiator rooted in environmental responsibility.”

Rick Bump

Chief Executive Officer and Co-Founder,
SAVARTUS.

CONCLUSION

The days of cold, inactive archives are quickly passing as demand for enterprise storage capacity accelerates rapidly, driven by massive AI-fueled growth. This surge underscores the pressing need for robust data management strategies that encompass the edge, core data centers, and the cloud. At the heart of active archiving success lies the ability to efficiently manage vast quantities of data. For AI initiatives to deliver meaningful and productive outcomes, they must process, analyze, correlate, and draw insights from enormous datasets. Once data volumes exceed a few petabytes, an active archive becomes essential, offering the right balance of accessibility, performance, energy efficiency, and cost-effectiveness. It's realistic to think that nearly every archive will become an active archive when it grows up!

Organizations must build the secondary storage infrastructure of the future on a foundation of well-planned data storage solutions, understanding activity patterns, retention requirements, and workflows. Without this, they risk increased costs, compromised data security, reduced cyber resiliency, legal and compliance issues, degraded customer experiences, poor decision-making, excessive energy consumption, and even damage to brand reputation. We are squarely in the AI era, making effective data management a core competency for digital transformation. Clearly the active archive plays a pivotal role, and the role will only increase, enabling modern enterprises to unlock the full value of their data while maintaining operational efficiency and sustainability for the foreseeable future.

For more information on the **ACTIVE ARCHIVE ALLIANCE** or to learn more about an industry-specific active archive solution to meet your needs, visit:

activearchive.com



THE ACTIVE ARCHIVE ALLIANCE

The Active Archive Alliance is a vendor-neutral, trusted source for providing end users with technical expertise and guidance to design and implement modern active archive strategies that solve data growth challenges through intelligent data management. Active archives enable reliable, online, and cost-effective access to data throughout its life and are compatible with flash, disk, tape, optical, cloud, and new emerging technologies as well as file, block or object storage systems. They help move data to the appropriate storage tiers to minimize cost while maintaining ease of user accessibility.

ACTIVE ARCHIVE ALLIANCE MEMBERS & SPONSORS

FUJIFILM

ARCITECTA®
OPERATING SYSTEMS FOR META+DATA

SPECTRA

IRON MOUNTAIN®

wasabi
hot cloud storage

IBM

POINT
software & systems

cerabyte

MagStor

Western Digital

SAVARTUS

BDT bewegt

XenData