



ACTIVE Archive

Active Archive and the State of the Industry 2020

FOREWORD

"Exponential data growth is driving fundamental change in computing, storage, and analytics. Fast processors and storage density keep pace with velocity and volume, but value is harder. Sophisticated analytics like AI are key to extracting value from long-term massive data sets. And active archiving enables organizations to operate analytics using high-performance and cost-effective archival storage."

— Peter Faulhaber, Chairman of the Board, Active Archive Alliance and President and CEO of FUJIFILM Recording Media U.S.A., Inc.

"Active archives today are less about actively moving data up and down tiers of storage, and more about actively moving the data into workflows to make data readily available to application owners and processing engines in the cloud. An active archive can bring life to that data as it becomes more and more important to businesses around the globe."

— Molly Presley, Founder of AAA and Head of Global Marketing at Qumulo

A New Age Dawns for Digital Archives

Genomics.

Global Pandemics.

Autonomous Vehicles.

Real-Time Business Intelligence.

The world is changing fast, and data drives our response.

Organizations around the world are amassing data at unprecedented rates. Analytics are key to unlocking value, but exponential data growth impacts storage management and costs. This enormous pileup of digital archives is highly diversified, spanning multiple verticals and applications:

- Healthcare, telecommunications, media and entertainment, utilities, financial, security, life sciences.
- Compliance, the IoT, high-res media, rich data streaming, AI and ML, data lakes, surveillance.

Not only is data creation accelerating, data is the most highly valued asset across the organizational board. As a result of fast-increasing value, today's data demands an intelligent archive solution that leverages the advanced capabilities of data movement technology and scale-out hardware.

"Archival data volumes are rapidly growing as organizations are beginning to see the value of analyzing previously untapped data."

— Deirdre Evens, Executive VP & GM, Records & Information Management North America, Iron Mountain

How Much Data Are We Archiving?

IDC's recent DataAge 2025 report projects that all data created globally will grow approximately 30% annually and by 2025 could generate as much as 175 ZB (zettabytes = 1x10²¹ bytes). Most of this data will be transient and not result in actual storage demand. The amount of data actually stored is still projected to be a massive 7.5 ZB in 2025, up from ~1.1 ZB in 2019.

Some of this data is production-level, but most of this data – at least 60% of it – does not need to be held on higher cost storage; it can be stored on lower performance tiers (archived).

Data reaches archival status when it displays 90 to 120 days of low or no activity, is typically unchanging, and is rarely overwritten. Not all this data should be archived. But much of it should be, thanks to strong retention policies and the newfound ability to gain value from archived data.

Archiving applies retention policies. Active archiving enables organizations to quickly gain value from retained data.

“Data and content owners want “active access” to their content when and where they want it. IT and business owners want an “archive” solution that stores their content cost efficiently. An effective active archive provides active access to data, while storing and protecting it as cost efficiently as possible.”

— Mark Pastor, Director Product Management, Quantum

Backup, Archive, and Active Archive Are Different

The Active Archive Alliance's 2020 survey indicated 66% of respondents were still using backup systems for storing archive data. But backup and archive are not the same and have different objectives.

When IT uses backup copies as archives, and repeatedly backs up unchanging archive data, they lengthen the backup window and waste time and money on storage resources. The failure to distinguish between backup and archive becomes increasingly expensive as data centers grow. Operational inefficiencies and poor resource utilization are additional consequences.

- Backup creates a copy of data and restores it in case of data loss or corruption. Hard disk drives (HDDs) are a primary target for backup, and many organizations use tape for secondary, immutable, and air-gapped backup targets.
- Archiving frees up expensive capacity by moving less frequently used data to more cost-effective storage locations.
- Active archiving uses highly scalable archive systems with intelligent software. The software uses rich metadata, indexes, directories, tags, and global namespaces to unlock archives and enable high-performance search and retrieval.

BACKUP (Copies Data)	ARCHIVE (Moves Data)	ACTIVE ARCHIVE (Fast Access to Data)
Copies data for protection and recovery. Leaves source data in place.	Moves infrequently used data to more cost-effective storage. Frees up space on source devices.	A scalable solution based on intelligent active archive software, tiered media, and/or the cloud.
Restores lost files to desired point in time. Speed is critical factor.	Retrieves files for reference and analysis. Retrieval speed is not a critical factor.	Provides rapid file- or object-level access to archived data.
Short-term duration averages 1-120 days.	Protects permanent and long-term data from modification or deletion.	Protects data and provides efficient data access. Typically operates without manual intervention.

AAA'S TEN-YEAR CELEBRATION

10th anniversary perspective from Molly Presley, Founder of the Active Archive Alliance and Head of Global Marketing at Qumulo

"It's been 10 years since we founded the Active Archive Alliance. Its purpose was to help users retain and leverage massive unstructured data. This is still our purpose today."

Looking back 10 years, many ground-breaking technologies were just gaining momentum: IoT, sensor-data, high-res surveillance, AI, and more. But users struggled to find data management solutions that worked with large data sets and multiple storage locations, so they could incorporate the new technologies into their organizations.

Today, active archiving software allows users to extract data analytics, artificial intelligence, machine learning, and more for far greater value and usability."

WHAT IS AN ACTIVE ARCHIVE?

An active archive manages data for rapid search, retrieval, and analytics. It is a high-value, scalable storage architecture that grants fast access to archival data across a virtual file system and manages data between storage systems and media types typically based on user-defined policies. Users can implement active archives on-premises, in the cloud, or both.

Active archive systems use node-based architectures with highly parallel threads to move much more data than legacy archiving systems, scaling up to accommodate petabyte-scale data movement.

The active archive ecosystem can integrate SSDs, HDDs, NAS nodes, tape, optical, and the cloud. The system can support file, block, and object, including unstructured data, with intelligent data movement software.

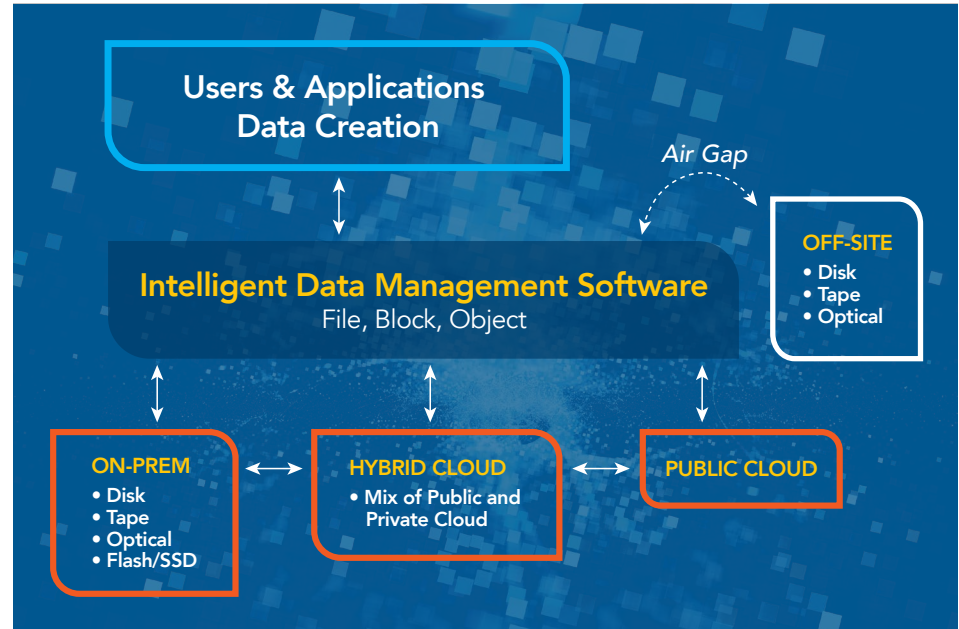
Open standard tape file systems such as LTFS provide highly functional and cost-effective active archives. Cloud service providers (CSP) and their partners are also making advances by architecting storage services to support active archives.

Rich metadata provides additional intelligence, and global namespaces eliminate data silos for faster search and retrieval in a single virtualized storage pool.

Meanwhile, ongoing AI development is adding more opportunities to harvest the untapped value of archival data.

THE ACTIVE ARCHIVE

Integrates Intelligent Software and Scalable Storage for the Optimal Archive Solution



"Data storage is critical to artificial intelligence and business development. Active archives help you get the right data and make sure you're storing it in the right way, which lets you recognize more revenue at a lower cost. In other words, use it or lose it!"

— Shawn O. Brume Sc.D., Global Hypergrowth Storage OM, IBM

ACTIVE ARCHIVES AT WORK

Not every organization places the same value on active archiving. Active Archive Alliance's survey reported that 67% of respondents did plan to deploy an active archive where data is available online and always accessible. However, 43% of respondents were not ready to implement or modify an active archive strategy within the next year.

Understandable. IT projects are balancing acts between business challenges, strategies, and resources. Data archiving rarely makes the top of the project list, especially if the organization treats archive as backup.

But organizations may not have that luxury too much longer. The ability to mine large datasets for value is a significant competitive advantage, and active archiving is the means to achieve it. Businesses without active archiving spend money to store and manage growing data but miss out on extracting business value.

Key usage cases for active archiving tell the real story, starting with healthcare and high-performance computing (HPC) in life sciences.

Healthcare

Healthcare has been in the forefront of active archiving for years. Active archiving enables healthcare organizations to track patient diagnoses and treatments across multiple providers, ensures regulatory compliance by proving data security and availability, offers streamlined workflows to records management and clinical users, and supports telemedicine (remote clinical services) and telehealth (remote non-clinical services).

Machine learning is a major component of healthcare IT innovation. The technology identifies meaningful patterns in big data sets, which improves data ingestion, indexing, and information identification. Active archiving makes machine learning possible on a large scale. It also supports the quadruple aim in healthcare to lower costs, increase patient satisfaction, reduce physician burnout, and achieve better outcomes.

"Applying chatbots and algorithms to active archives of legacy medical records could yield new ways to discover, communicate, research, diagnose and treat illness."

— Shannon Larkin, Vice President Marketing & Business Development, Harmony Healthcare IT

DATA AND COVID-19

When COVID-19 hit, innovative active archives were already in place. New medical research workflows are analyzing and correlating human, demographic, geographic, contact tracing, disease treatment, and vaccine research.

"Increasing demands for discrete data will drive healthcare to adopt sophisticated archiving solutions. Reporting tools capture emerging trends, including population health initiatives."

— Julie Fogel, Director of Marketing, MediQuant

COVID-19 data and analysis does not stop with healthcare. Multiple industries are consuming pandemic information to make data-driven decisions. Large-scale data analysis, made possible by active archiving, drives legal, governmental, and business decisions in response to the pandemic.

Active archiving also protects organizations involved with COVID-19 from phishing and malware attacks. Hacking attempts have increased since the announcement of the pandemic. Active archives enable organizations to decommission multiple storage silos in favor of highly secure archives, many with offsite air gap capabilities.

HPC in Life Sciences

Scientific HPC generates extreme scale, data-intensive workflows allowing researchers to run complex calculations in the life sciences, space exploration, volcanology, energy, physics, meteorology, and many more.

Life sciences researchers are among the heaviest HPC users. For example, gene therapy is an extremely competitive area. To remain competitive, life science organizations must quickly perform sophisticated analyses to reveal the insights hidden within billions of objects.

However, massive data sets and large imaging files tax legacy storage. For example, the bandwidth ratio between compute and I/O performance continues to diverge, with compute performance outstripping IO performance. A compute-intensive scientific workflow needs exploit all available IO capabilities, including faster access to large amounts of archival data. And managing life science data also requires high security, intensive analytics, continuous data generation, and scientific collaboration.

Active archives remove legacy storage roadblocks, which enables the life sciences to offer life-saving research and products faster than ever before.

“First, innovative active archiving helps business users to manage, maintain, and control data growth in the easiest and most cost-effective way. Second, it enables users to extract high value out of their stored data for competitive advantage.”

— Betsy Doughty, Vice President of Corporate Marketing, Spectra Logic

More Major Use Cases

Other top use cases include security, business efficiency and continuity, media and entertainment, and IoT, including autonomous vehicles.

- **Security.** Tape-based active archives enable the “air gap defense.” Tape becomes an electronically disconnected copy of data that prevents rolling cybercrime disasters from attacking backup and archive copies. Files are largely inaccessible to online hackers, as the files can be offline more than 95% of the time.
- **Business Efficiency.** Active archiving benefits business by improving workflows, business continuity, and consolidating multiple storage systems into a single interface. Instead of operating independently in isolated siloes, active archiving aggregates archival data from different storage locations into an active archive. This allows applications to quickly access archival data while improving business workflow and is transparent to users. Active archiving improves business continuity by enabling business to restore high-value data faster and more efficiently.
- **Media and Entertainment (M&E) Workflows.** M&E relies heavily on digital archives for ongoing monetization. For example, a studio can extend movie profits by tapping the archives for a sequel, video game tie-in, or action figures. Traditional workflow involves the creative team repeatedly requesting archived assets from IT, and IT continually fielding the requests, and locating and restoring tapes. Active archives streamline the workflow by granting transparent access to approved users, who quickly locate and restore the digital assets they need.
- **Autonomous Vehicles.** AV manufacturers receive real-time data each day from each autonomous vehicle. The massive data sets include details of driving decisions, events, object identification, and driver interaction. Active archives enable researchers to quickly analyze data for R&D, insurance, laws and regulations, and business intelligence.

“Identifying when primary data should be archived is critical, and AI will play an increasingly important part in this over the next decade.”

— Dave Thomson, Senior VP Sales & Marketing, QStar Technologies

“The past decade has been a period of massive change for media workflows, and archives have taken a new role. Active archives are part of the daily workflow and play an essential role in creating new productions.”

— Ferhat Kaddour, VP Sales & Alliances, Atempo

ACTIVE ARCHIVE KEY BENEFITS

FUNCTION	KEY BENEFITS SUMMARY
Ease of Use	Provides end users with file-level access to all their data, all the time without IT intervention.
Streamlined Workflows	Improves business workflow performance for tasks that use archival data.
Business Continuity	Helps secure operations and core business functions from a disaster by providing fast and secure access to high value archived data.
Lower Cost	Supports data movement from costly SSDs and HDDs to lower cost storage, which significantly reduces overall storage TCO.
Higher Capacity	Enables low activity data movement from SSDs or disk to tape, which releases production space for future growth. The economic impact is considerable, since at least 60% of data presently stored on SSD or HDD are suited for tape archives.
Reduced Backup Window	Streamlines the backup process by removing terabytes of data from the backup set on primary storage, and greatly reduces the duration of the backup window.
Improved Primary Storage Performance	Reduces the amount of infrequently accessed data on primary storage by moving data an active archive. Organizations realize substantial improvement in production system performance.
Improved Access to Archival Storage	Enables a cache front-end on archival systems, which significantly improves performance
Cybercrime Protection	The "tape air gap" prevents cyber-attacks on archival data. Data is stored offline without electronic access and cannot be hacked.
Cloud Usage	Presents a new cloud service that easily integrates with private, public, and hybrid cloud environments.

HIGHLIGHTING THE VALUE OF ACTIVE ARCHIVES

Blogs:



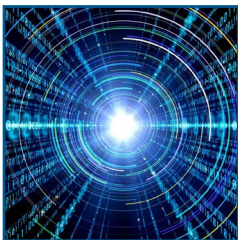
To Active Archive or Not to Active Archive? That is the Question

[READ](#)



The Right Combination of Tape and Cloud Storage

[READ](#)



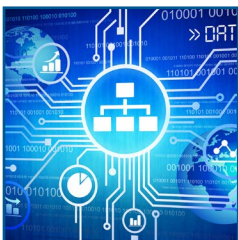
Active Archive: Object Storage Matures into Hierarchical Storage Management

[READ](#)



Active Archive Supports the Quadruple Aim in Healthcare

[READ](#)



An Active Archive Strategy Can Help Solve More Than Just Archiving

[READ](#)



IoT and AI Generate Demand for Active Archive

[READ](#)

"Archiving inactive data on a suitable storage technology and providing transparent access with a storage & data management software is the most efficient way to reduce costs without productive restrictions and at the same time to fulfill requirements for long-term compliant storage."

— Sebastian Klee, CMO at PoINT Software & Systems

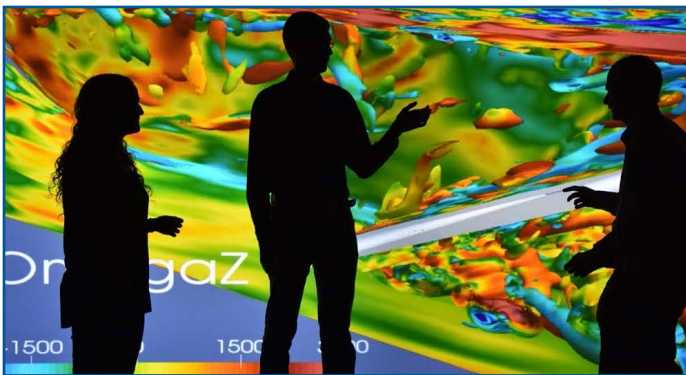
HIGHLIGHTING THE VALUE OF ACTIVE ARCHIVES

Case Studies:



John Muir Health Cuts Costs, Secures Data, Sheds Legacy Systems

John Muir Health and MediQuant collaborated to move multiple terabytes of clinical, financial and ERP data into a single user-accessible, secure archive with one common framework across the entire enterprise, using DataArk® software and MediQuant's data stewardship program.

[READ](#)

STFC Rutherford Appleton Laboratory to Process and Preserve More Than 60 PB of Scientific Research Data with Spectra

"Our new Spectra tape library will allow the massive volumes of environmental data we collect to be made available to environmental scientists and support their research for many years to come." — Professor Bryan Lawrence, University of Reading

[READ](#)

Videos:



This video highlights how active archives enable reliable, online and efficient access to archived data. Active archives are compatible with flash, disk, tape, or cloud (public or private) as well as file, block or object storage systems, helping move data to the appropriate storage tiers to maximize cost savings.



In this video, Molly Rector, co-founder of the Active Archive Alliance, explains how an active archive can provide visibility into your applications and machine-generated data with actively assigned metadata no matter on what tier it is stored.

TREND: EXABYTE GROWTH DRIVES INNOVATION

“There will continue to be strong exabyte growth in read-centric applications in the data center, from AI, ML, and big data analytics to a variety of business intelligence and accessible archive workloads. These at-scale use cases are driving a diverse set of performance, capacity and cost-efficiency demands on storage tiers.”

— Phil Bullinger, SVP and GM,
Western Digital

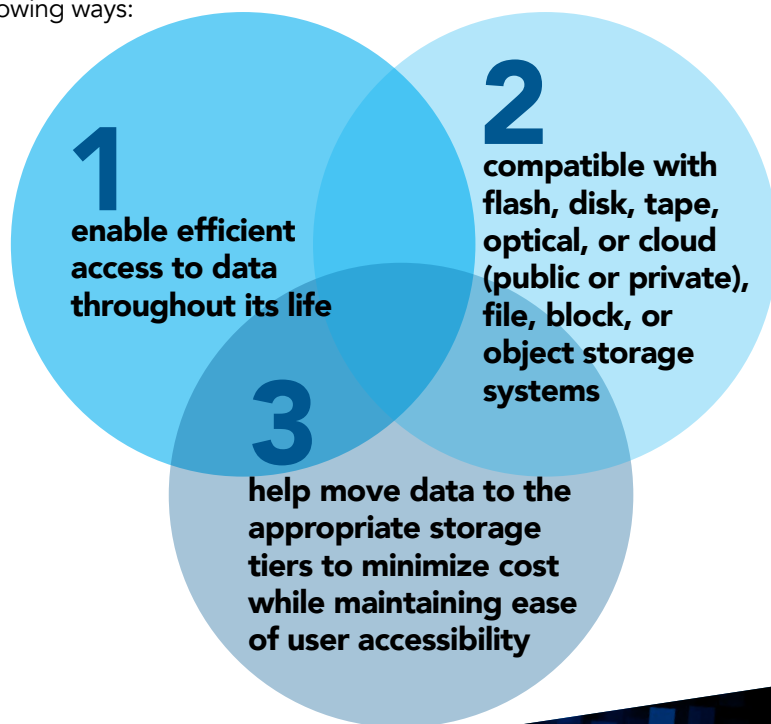
CONCLUSION AND OUTLOOK

The growing reliance on archival data makes it ground zero for unlocking game-changing data strategies. New applications such as the Internet of Things (IoT) with billions of nodes and boosted by the arrival of 5G networks, will help fuel insatiable demand for more intelligent active archives.

Breaking down the great wall of untapped archival data has become a strategic goal for many leading-edge organizations. They want active archiving solutions that effectively handle all data types and scale quickly, without the cost and burden of traditional siloed data warehouses. The capabilities are here in place to make that a reality.

ABOUT THE ACTIVE ARCHIVE ALLIANCE

The Active Archive Alliance is a vendor-neutral, trusted source providing end users with technical expertise and guidance to design and implement active archive solutions for intelligent data management. The goal of the Alliance is to encourage a multi-vendor effort to promote and align the awareness and technologies needed to meet the rapidly increasing requirements for archival data in the following ways:



“Inexorable data growth has a serious impact both on storage technology choices and the strategies needed to manage these ever-increasing data sets. The problem drives the trend: the increasing need to know what the data is, and how to extract value from it.”

— Floyd Christofferson, CEO, StrongBox Data Solutions

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

PR@activearchive.com



activearchive.com

Active Archive Alliance Members & Sponsors

FUJIFILM



Quantum



Western Digital

ACTIVE ARCHIVE is the trademark of the Active Archive Alliance. Third party trademarks used herein are the trademarks or registered trademarks of their respective owners used with permission.

© 2020 Active Archive Alliance. All rights reserved.