

AWS Partner Network Storage Solutions: Data Archive

The data archive use case allows the preservation of important, but infrequently used data with appropriate cost and resiliency. AWS Partner Network (APN) Partners offer solutions that leverage Amazon Glacier for durable and cost-effective long-term data retention.

Data Archive Challenges

New regulatory statutes have lengthened data retention requirements and consequences for not being in compliance have become more severe. Data needs to be archived in real time and be easily searched, all while being stored in an affordable fashion. Archive solutions reduce primary storage and backup infrastructure requirements and emphasize both cost and performance advantages.

- **Data growth:** Unlike backup workloads, archives are a single authoritative copy of data. This introduces recovery and discovery requirements, which make offline storage impractical. As a result, on-premises archive infrastructure and operational expenses scale up in a sawtooth pattern. When a tape is full, you move on to another tape. When a tape library is full, you install another tape library. Magnetic tape degrades over time, with the best results approaching 30 years and the first signs of bit rot typically appearing after eight to 10 years. Archives often have long retention periods, sometimes indefinite, which increases operational overhead to refresh media.
- **Archive is not backup:** Backup solutions make point-in-time consistent copies of data. Archive solutions are tasked with preserving a single copy of data, which does not exist anywhere else. In order to satisfy data governance demands, archival solutions maintain extreme granularity and discovery capabilities. A backup data set might include an entire disk, a set of files, or application-specific data. Archives are often unstructured, consisting of not only backup datasets but also documents and email messages. Extracting and exposing metadata for search purposes is the core competency of most archive solutions, whereas backup solutions protect data from physical or logical errors, such as system failure, application error, or accidental deletion.
- **Slow access times:** Tape often is the storage media choice for archive data because its density and streaming performance are excellent when compared to an individual tape's acquisition cost. Random access capabilities of tape are poor, which makes unlocking value from archive data a challenge on tape. Search-based retrieval and archive validation are the only ways to leverage the data. Concepts like machine intelligence, complex queries, and leveraging analytics frameworks are nonstarters.



Amazon Glacier is designed for 11 nines of durability and provides a cost point on par or below tape while also enabling utility pricing and simplifying data management.

AWS for Archive

When we speak about how to migrate tape data to the cloud, the service we talk about most is Amazon Glacier. Amazon Glacier is designed for 11 nines of durability and provides a cost point on par or below tape while also enabling utility pricing and simplifying data management. For long term protected archives, many customers capitalize on an Amazon Glacier feature called Amazon Glacier Vault Lock. Amazon Glacier Vault Lock allows you to easily deploy and enforce compliance controls for individual Amazon Glacier vaults with a vault lock policy. You can specify controls such as “write once read many” (WORM) in a vault lock policy and lock the policy from future edits. Once locked, the policy no longer can be changed. This is ideal for protecting archival data against accidental deletion due to user error, bad actor scenarios, and conventional problems like bit rot.

Archive solutions typically build indexes using two sources of data. The first source is system metadata, such as file ownership, creation times, size, and other access control list properties. These attributes are important elements in data lifecycle decisions. The second source is content metadata such as document keywords, authors and audiences of both documents and email messages, revision histories, and image metadata like geotags and sensor values. This information enables the archiving system to make decisions based on the business value of the data itself.

AWS Competency Partner solutions integrate with Amazon Simple Storage Service (Amazon S3) and Amazon Glacier for long term preservation of data. Deployment options range from purpose-built appliances to software-defined solutions. Integration with complementary backup software suites is common, but not required. In some use cases, archival engines directly interface with primary systems. Email archival and document library support often require directly interfacing with primary systems to facilitate archival.

AWS Storage Competency Consulting Partners

AWS Storage Competency Consulting Partners possess deep domain and consulting expertise to help enterprises adopt and deploy complex storage solutions in one or more core storage categories, including data archive. These APN Partners have demonstrated success fully implementing storage solutions and projects on AWS, and are familiar with existing AWS Storage Competency Technology Partner solutions on AWS. For more information on AWS Storage Competency Consulting Partners, please visit https://aws.amazon.com/backup-recovery/partner-solutions/#Consulting_Partners

Benefits of AWS Storage Competency Partners for Archive

Archive with AWS Storage Competency Partners enables companies to harness the benefits of cloud economics while still maintaining their existing archive investments.



Affordable archival with no infrastructure:

Amazon Web Services (AWS) storage services are not only affordable, but they can reduce the total cost associated with maintaining legacy data storage infrastructure, such as tape management, onerous refresh cycles, security, and manual data transfer.



Ease of solution deployment: AWS and APN Partner solutions have been designed for push-button deployment and architected to take advantage of AWS' scalability and durability. AWS storage services have achieved numerous compliance standards, security accreditations, and help achieve an archival strategy that meets the requirements for your business.



Granular control: APN Partner solutions support a range of AWS storage services, providing a cost efficient storage target for the data. Lifecycle policies enable fine-grained movement of data to the appropriate storage tier, even as long-term requirements change. Additionally, you can extract value from dark archive data stored immutably on AWS through big data processing, machine learning, and artificial intelligence.