

Data Storage Solutions:

How to Reduce Costs with a Data Archiving Tier

by Timothy C. Summers, V.P. Product Engineering

Overview

Almost every information technology organization is faced with managing explosive data growth. The data is valuable and must remain available. However, even with the cost of storage technology continually declining, the business data that must be kept and managed is ever increasing. As a result, IT personnel are under extreme pressure to reduce costs and budgets.

The easy path is to consider placing all data on Tier 1 storage resources. This facilitates access to data for end users and application infrastructures, however, simultaneously perpetuates an increased long-term cost to the business and great strain on IT professionals due to continual storage management challenges. Business continuity planning and associated disaster recovery initiatives, the high costs of Tier 1 storage, technology refreshes, data migration, restricted floor space, and power and cooling costs are just some of those challenges.

A data archiving (Tier 3) solution can greatly reduce IT budgets and simplify operations. This is not a revolutionary new concept, but for many the fundamentals are misunderstood or mischaracterized. Due to these perceptions, ongoing operations and implementation roadblocks, the concept is passed over as a viable solution.

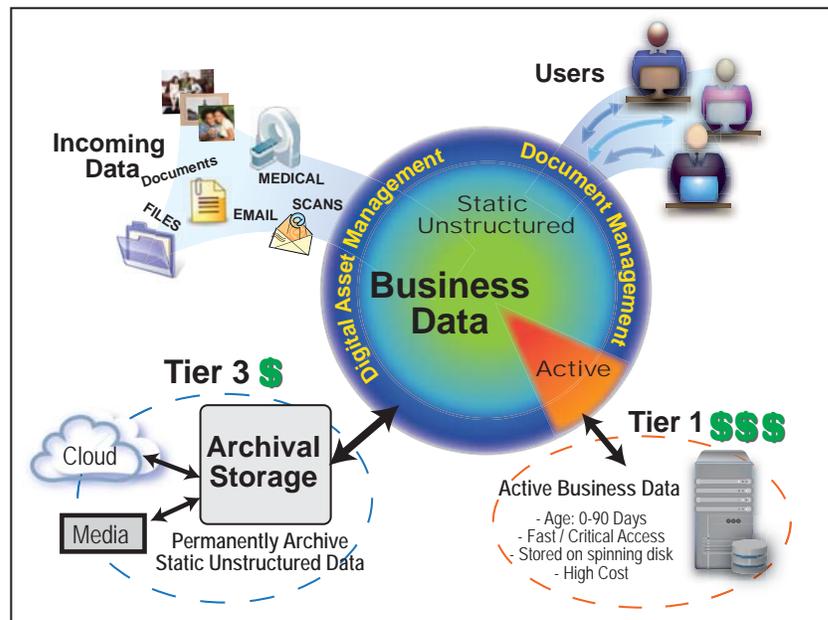
Data archiving is the answer. Taking the approach outlined in this document can greatly simplify implementation, ensure a seamless transformation for users, and reap significant cost savings.



Transitioning to Tiered Storage with an Archive Tier

Key steps for transitioning to a tiered storage environment with a Tier 3 data archive storage layer.

- Understanding your data and its applicability to being archived
- Find the correct balance between Tier 1 cost savings and possible impact on accessing data
- Implement a transparent archiving environment with seamless archive data access
- Understand the savings provided by freeing Tier 1 storage and replacing it with Tier 3 archival storage
- Selecting the correct Tier 3 archival storage to maximize savings.



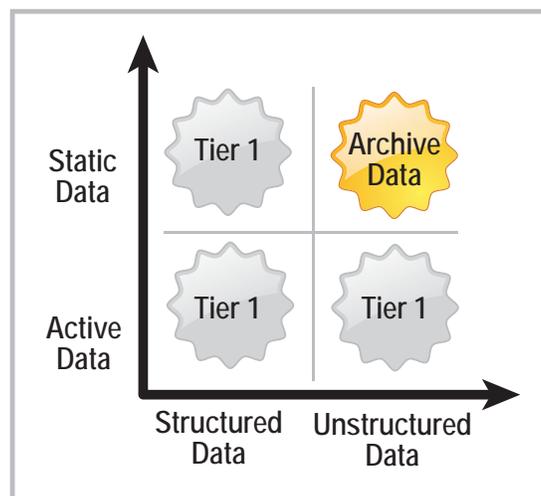
Understanding Your Data

Industry analysis confirms that data storage requirements double on average every two years. The type of data created by the organization is a significant factor in determining what data to archive.

As the diagram depicts, data can be classified into four quadrants. Active data is constantly being modified while static data is possibly modified for a short period after creation, and then rarely or never again modified.

Structured data includes files such as databases, whereas, office document files are typically considered unstructured. Unlike active data, 70-80% of unstructured data (estimated) is rarely if ever accessed. Furthermore, lifecycle management of unstructured data is very simplistic compared to structured data.

For almost all organizations, the archiving sweet spot is unstructured static data.

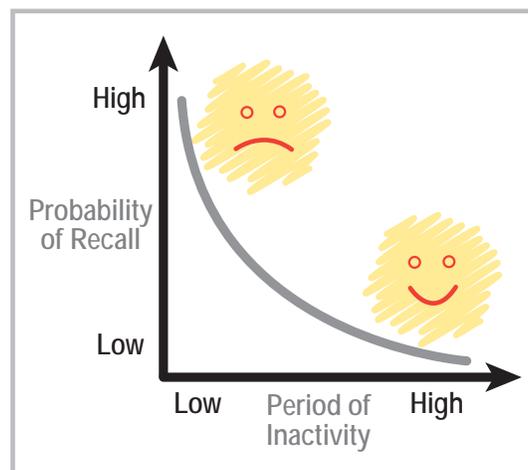


Finding the Correct Balance: Cost Savings vs Impact to Users

A proper balance of Tier 1 storage combined with Tier 3 storage will provide both cost savings and fast accessibility to the data. Having identified the class of data for archiving, the next task is to determine how much of that data can be migrated to lower cost archival storage. Once the data is archived, the amount of time it takes to access the data can be extended by milliseconds to hours depending on the archival storage that is implemented. At the same time, the larger the amount of data archived, the greater the savings results.

As exemplified in the illustration, the simple key decision is what period of inactivity should occur before data is archived. If you archive too quickly, then the probability of users accessing the data increases along with corresponding longer data access times. If you archive too slowly, this results in lost savings leaving money on the table. Key considerations include:

- **Tier 3 Archive Storage access times:** There are many types of archival storage, and some can add seconds to access time while others can add hours. Selecting the appropriate archival storage is extremely important.
- **The Business requirement for immediate data access:** If extending data access times will have a high impact on productivity or cause business concerns, that issue must be balanced against the access characteristics of the Tier 3 archival storage.



Ultimately, the simplest approach is to start conservatively. Identify the 10% oldest static unstructured data and see what type of access patterns occur with the data once stored on Tier 3. In general, it should be negligible. From that point forward, increase the amount of data archived over time, freeing additional Tier 1 storage resources and further reducing costs related to continued capacity expansion and management of Tier 1 storage. Data lifecycle management modeling tools can be used to model data offloading based upon policy selections.

Transparent Archiving Environment with Seamless Data Access

Now that we know how much and what type of data to archive, the next step is to determine how to avoid impact to the organization's business operations. The key ingredient is transparency for IT and user processes involving data access. More specifically, once data is relocated to Tier 3 archive storage, how will business users continue to access data seamlessly without affecting their existing data access processes?

To achieve this, a data migration process to relocate data to the Tier 3 archive storage must be considered. There are three general categories of solutions to be considered:

Managed Placement

Relocation of data can be achieved manually by simply utilizing applications such as Windows Explorer, selecting the data to be relocated and then moving it to the archive storage tier. In general, the archive storage tier is then setup to mirror the structure of the Tier 1 storage.

Pros

- No software requirement lowering implementation cost.

Cons

- Users must manually search both Tier 1 and Archive locations for data.
- Users must ensure proper placement of data (on migration, when recalling, when modifications are implemented).
- Not automated, not policy based.
- Data must be moved manually and repetitively over time as data ages.

Data Lifecycle Management Software Products

Data Lifecycle Management software products, such as QSTAR's Network Migrator or Rocket Software's™ Arkivio Autostor, seamlessly automate the process of identifying and relocating data to be archived while simultaneously providing users transparency to data access.

Pros

- Policy based rules that aid lifecycle management of data, facilitating migration of data to new storage locations. By simply defining policies, data meeting the criteria will then be migrated to archival storage.
- File stubbing capabilities ensure data access from the original Tier 1 storage location. Though file data is relocated to the archive storage tier, access is achieved via the original storage location. Data is seamlessly recalled from Tier 3 archives.
- Support for open and closed storage file systems including Windows, Unix, Linux, MAC, NetApp, ONTAP, EMC (Centera, Celerra, VNX, Data Domain, Isilon), Hitachi Data Systems (HNAS, HCP), Quantum, etc.
- Some products optionally provide data indexing and contextual search capabilities, facilitating data searches without requiring immediate recall of archived data. If users perform a lot of searching against the data, this can result in high archive data recall rates. Content management or data indexing systems will index the data and allow searches without actual file retrieval avoiding high archive data recall rates.
- No modification of user behavior required for accessing data.

Cons

- Some products require agents or file system drivers to be installed on the Tier 1 storage servers, while others do not. This can cause concern for some IT administrators, as well as require considerations in the area of maintenance activities.

Content Management Systems

Content Management Systems are another option but will require a change in user's work processes, as these systems provide a framework that supports the creation and modification of digital content via a dedicated user interface. If your organization has more complex requirements, CMS's may be an attractive alternative.

Pros

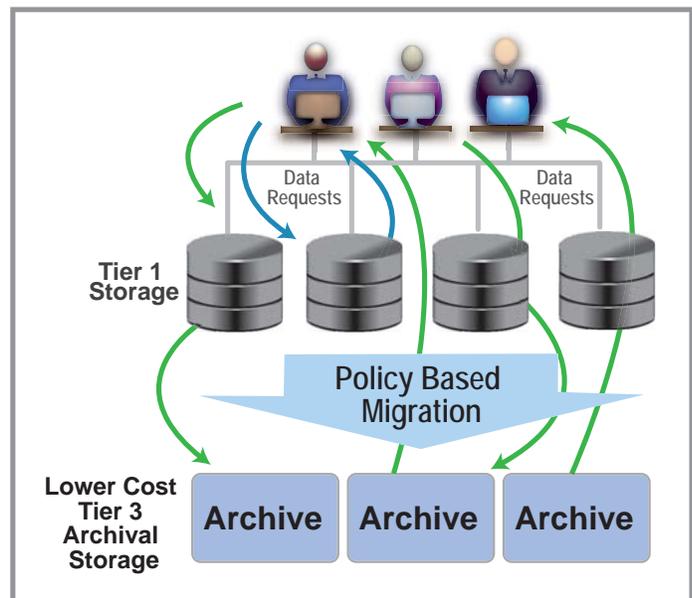
- Provides a framework for managing the creation, modification, organization and collaboration of and around digital content.
- Typically has dedicated web-based user interfaces that support simultaneous multi-user collaborative environments.
- Typically has extensive features for managing content, such as multiple presentation and editing interfaces, web publishing, version control and others, adding a wealth of features.
- Optimizes data retrieval by providing optional features for indexing and searching content, ensuring data is not recalled from archive storage until you are sure you wish to view the data.
- Includes a Hierarchical Storage Management system that implements a tiered storage strategy, placing data on storage tiers based on administrator defined policies.

Cons

- The CMS Administrator must "setup" the CMS business architecture.
- All historical data must be imported into the CMS system. This may result in a significant effort across the business.
- All users across the business must now utilize the dedicated CMS system interfaces on a daily basis. This requires user training and adherence to new business policies and procedures for management of data.
- All data housed in the CMS must be accessed via its dedicated UI. Any modifications must be committed to the system as well.
- CMS systems require a much more complex implementation process that will take longer to implement

Though all three techniques can be employed to achieve archiving of legacy data, the Data Lifecycle Management products offer the most transparency and ease of implementation. This transparency is depicted in the illustration. Based on administrator defined policies, unstructured static data is first migrated from Tier 1 storage to lower cost Tier 3 archival storage.

Hence, the oldest data is located on Tier 3 storage resulting in savings. When users access the data, it is dynamically retrieved from either Tier 1 storage or Tier 3 archival storage. The user does not have to know where the data resides, but simply accesses it as they have in the past. This total transparency to the requestor eliminates any impact to users.



Understanding the Tier 1 Storage Cost Savings

Moving static unstructured data from Tier 1 storage to Tier 3 archival storage has two main areas of impact:

1. Offloading data reduces Tier 1 storage utilization - This enables you to:

- Reclaim Tier 1 storage, resulting in eliminated or reduced future storage purchases. This provides significant savings by eliminating not only future storage purchases but also reduced backup requirements, associated long-term costs such as power, cooling, data refresh (including technology refresh every 3 to 5 years) and related data migration costs, etc.
- Reorganize storage resources to free storage devices for complete removal from the data center knowing your data is secure for the long-term on Tier 3 storage. This reduces the IT storage footprint and the associated cost of these devices as identified above.

2. Migrating data to Tier 3 archival storage increases utilization. Costs associated with this include:

- Selection and installation of data archiving devices. Archival storage is generally overall lower cost than typical Tier 1 storage, providing significant savings. Media costs will be competitive with some lower cost Tier 1 storage devices, but after the initial purchase cost of Tier 3 (lower than Tier 1), ongoing costs are greatly reduced and the need to refresh technology is eliminated. Overall power and cooling lessen since true archival storage does not require power to store the data.
- Additionally, long-term data retention is a key requirement. Storage such as optical should be considered due to its greater than 50 year life span, ensuring the secure storage of the data. This also eliminates the need for migrating data for technology refresh reasons, providing the most significant long-term cost savings.

Other technologies can cost 10 to 30 times as much when considering long-term cost. HDD software-based WORM requires refresh every 3 to 5 years, requiring major capital investments repeatedly over time. Technologies such as tape require refresh every 7 years (can be slightly longer depending on tape technology), are prone to data loss due to tape stretching and damage, are highly inefficient on data recall functions (lack random access capabilities), and carry a risk of potential data loss.

Initial Tier 3 acquisition costs can pay for themselves when balanced against deferred or eliminated Tier 1 storage purchases. Hence, immediate capital outlay will be the same or slightly reduced. However, significant capital and operational cost savings are attained through elimination of future technology refreshes and power and cooling expenses.

Tier 3 storage provides piece of mind knowing that a trustworthy copy of last resort is available and that your critical data is secure for the period of time that you want or need to keep it. Archiving on the right technology further enables you to take control of technology refreshes which require data migrations and significant capital investments, as opposed to technology obsolescence forcing the occurrence and timing of these costly events.

Selecting the Correct Tier 3 Archival Storage

Key business considerations for selecting archive storage technology include:

- Corporate or Regulatory Mandate Requirements - HIPAA and the HITECH Act, IRS, Financial, and other regulations require unalterable permanent storage of artifacts with absolute protection of data involving user/patient personal information.
- Data Recall Requirements – If static data continues to be referenced frequently, the archival storage should have fast recall capabilities in order to meet the demands of business operations. An example of this might be patient medical images.
- Business Continuity Requirements - Issues can (and do) occur at the most inopportune times. Archive storage must be capable of supporting both short and long Recovery Time and Recovery Point Objectives (RTO/RPO). In essence, the capabilities can range from recovery from backup to high availability replication.
- Data Protection from Unauthorized Access – archive systems must protect data while online, at rest or in transit.
- Ensures a confident strategy for securing valuable data assets permanently over the long-term.

Alliance Storage Technologies, Inc. offers advanced data archiving solutions that meet the needs mentioned in this document and more. Systems offer many other features that enable seamless access to archive data and simplify management.

Features include:

- Easily integrates with applications, users, and lifecycle management products via standard Network Attached Storage interfaces such as CIFS and NFS. These File Shares provide an excellent and easy integration path for Data Lifecycle Management products, Content Management Systems, and manual management via products such as Windows Explorer.
- Multiple archive storage tiers:
 - True WORM optical media meets corporate and regulatory mandates by providing 50+ years of storage life.
 - Cloud-Integrated Storage – deploy for archives without mandates or as a secondary offsite copy.
 - RAID Cache - stores recently archived data and limits the requirement to recall from optical media or cloud.
- Multi-layered Data Protection
 - Supports multiple user authentication services such as: Windows Active Directory and LDAP to ensure authorized access.
 - FIPS 140-2 compliant data encryption with integrated key management for the protection of sensitive government, corporate or personal digital information.
- Vulnerability protection to ensure data is secure
 - Built-in policy-based data management capabilities enable easy definition and management of purpose-defined archives.
 - Business continuity options meet standard to high-availability recovery time objectives, including high availability replication with failover/failback.
- Easily scalable via extensible hardware and software defined infrastructure.
- Simple network integration via Gigabit Ethernet connections with bonding and redundancy.

Summary

With the proliferation of unstructured data consuming countless petabytes of Tier 1 storage across almost every data center, reducing the associated costs of managing and preserving that data is of utmost concern. IT departments everywhere can benefit by implementing an archiving solution that offloads expensive Tier 1 Storage to an affordable Tier 3 archive.

It is important to deploy a lifecycle management product that offers user defined data movement policies and data stubbing capabilities while ensuring dynamic relocation of data without encumbering normal business and user operations. Initially, a conservative data movement strategy can be implemented in order to ease into Tier 3 archive storage limiting data movement and providing insight into specific data inactivity measurements.

Selecting the right Tier 3 archival storage solution is paramount to success. Solutions must offer cost savings in raw storage and reduced operational costs and should also meet governance and mandate requirements. Scalability and elasticity decrease initial investment and deployment costs while easily accommodating future expansion.

ASTI's data archiving solutions meet the archiving needs of tens of thousands of customers across many industries including Financial, Medical and Healthcare, Law Enforcement, Legal, Security and Surveillance, and Government including Municipal, County, State, and Federal Agencies.

To discover more about ASTI's archive solutions and products and their benefits, please visit www.alliancestoragetechologies.com.

About Alliance Storage Technologies, Inc.

Alliance Storage Technologies, Inc. (ASTI) develops and manufactures products and solutions specifically designed for professional data archiving. Innovative network attached storage solutions feature modular, flexible architecture and integrate seamlessly with leading content and storage management platforms. Trusted worldwide for the protection of archive data, ASTI's solutions provide permanent long-term, authentic storage of critical records to ensure data security and compliance with regulatory requirements.

Archiving products can be configured for each customer's specific environment and are suitable for businesses of any size from small organizations to the largest enterprise or government. ASTI partner affiliations include world class value added resellers, distributors, and integration partners. Best-in-class Service Direct support programs are designed to meet the demanding needs of global 24x7 operations.

Thousands of businesses worldwide trust ASTI quality to secure their valuable investments in corporate data.

Windows, Unix, Linux, MAC, NetApp ONTAP, EMC (Centera, Celerra, VNX, Data Domain, Isolon), Hitachi Data Systems (HNAS, HCP), Quantum, QStar, and Arkivio Autostor are trademarks or registered trademarks of their respective companies.



Copyright ASTI 2016

Alliance Storage Technologies, Inc.
10045 Federal Drive
Colorado Springs, CO 80908
alliancestoragetechologies.com
719-593-7900 Fax: 719-593-4164 WP003