Backup Is Not Archiving: Reduce TCO and Improve Data Protection with a Customized Archive Solution
Executive Overview

Without a comprehensive archive strategy, the potential toll on an enterprise includes rising costs, increasingly complicated backup and recovery, inefficient search/access, potential regulatory issues, and the inability to unlock the value of data in the future.

Industry leaders are no longer just retaining data, they are leveraging it to find and create new revenue opportunities. While storing data for regulatory purposes is still driving data growth, proactive organizations are seeking ways to leverage additional data that will help them grow their business. Today’s high rate of data growth and increasing number of regulations mean that enterprises need to evolve their thinking beyond legacy backup and recovery data protection strategies. Companies that operate in compliance environments must be able to meet requirements imposed by internal and external organizations, which mandate retention and retrieval for decades or more. In addition, enterprises are recognizing that there are also other contributing factors that add to their need for an archive solution. These include content digitization, lean budgets, and the need for storage efficiencies.

It is critical to understand the difference between backup and archive and that backup is not archiving. Yet, according to an IDC study\(^1\), 57 percent of respondents use backups to meet long-term retention requirements. A backup solution is intended to satisfy short-term needs for operational recovery. On the other hand, an archive solution is designed to meet long-term retention and retrieval requirements, which include assurances that obsolescence and data corruption will not compromise the archive. Further, a successful archive solution can reduce overall enterprise storage and backup costs, simplify management, improve access to information and provide access to new business opportunities.

This paper will help IT managers understand why developing an archiving strategy is a business necessity and why Oracle’s archive solutions, with policy-based tiered storage, assure long-term business viability.

\(^1\) IDC, Rethinking Archiving, presentation at Storage and Networking World spring conference, 2012.
Introduction

Rapid data growth is spurred by factors such as increasing numbers of devices and users, server virtualization, rich media content and social media. IDC’s 2012 Digital Universe Study estimates that the amount of data is doubling every two years, and that 80 percent of it touches enterprise data centers at some point\(^2\). Gartner also predicts 30 percent to 60 percent compounded growth in data creation through 2015\(^3\). This is not just structured data growth. Both structured and unstructured data are growing and being retained for much longer periods of time.

In addition to data growth, other market drivers affect enterprise data protection strategies:

- **Storage efficiency**: Scrutiny of the value and access of data over time points to the importance of tiers of storage to reduce total cost of ownership (TCO).

- **Compliance and e-discovery**, which is driven by legal requirements in a variety of vertical markets: In healthcare, for example, the retention period in the United States is the life of the patient plus 10 years. In many financial institutions, certain data is required to be kept for at least seven years, and some financial archives house information *forever*.

- **Business opportunities**: The potential to extract value through big data analytics causes organizations to save data, once viewed as noncritical, for mining purposes either now or in the future.

- **Digitization**: Significant analog-to-digital activity is happening within the media/entertainment industry. Other vertical markets are using digitization to preserve photos, video, or historical documents for future generations to enjoy, for business opportunity, or for compliance purposes.

---


Backups and disaster recovery strategies alone cannot respond adequately to the majority of the conditions described above. Creating an archive strategy is the right choice for long-term data protection. A true archive strategy can help organizations lower overall storage costs, lower backup costs, improve file access, and ease management. While a long-term archive faces many challenges, including obsolescence and data corruption, Oracle’s archive solutions solve these top challenges, while delivering business value.
Archiving, a Business Necessity

Why is archiving considered a business necessity?

Archiving involves preservation of data that must be kept for business or legal reasons. This data often is fixed content that does not change. It’s essential to know where the data resides, that it's stored in an accessible format, and that health checks are performed on a regular basis.

When current data protection strategies are evaluated, it is clear only archiving stands out as a strategy designed for data longevity and organizational efficiency.

Comparison of Data Protection Strategies

The following backup strategies lack some of the basic requirements to be considered a comprehensive strategy for long-term data protection:

- Backups ensure short-term protection of data to enable operational recovery following an incident such as accidental deletion, corruption, or a local hardware failure. In this case, backing up data is relatively easy. Restoring data is more challenging because the backup solution must be able to address the need for a massive restore or retrieval of a single file from a backup set. If the data has been deduplicated, it must be rehydrated before being restored.

- The objective of a Disaster recovery plan is to minimize recovery time objectives (RTO) and recovery point objectives (RPO). The lower the RTO and RPO, the more expensive the disaster recovery solution.

- Tiered storage is a solution that relocates less frequently used data to lower performance and lower cost storage devices. This is not necessarily data that qualifies for archiving. It is data that is moved transparently to another tier of storage without changing the application’s visibility of the data. The primary goal of tiered storage is to reduce storage TCO by eliminating some percent of expensive, high-performance storage. Tiered storage can be implemented in primary, backup, and archive storage environments.

Implementing a customized enterprise archive with tiered storage is the ideal strategy for satisfying long-term data protection requirements. Policy-based archiving software does its work automatically, including ingesting, archiving, recalling, migrating, and deleting data. Advanced features focus on data integrity, health checks, and proactive healing within the system. This choice reduces risk and cost, while supporting long-term business viability.

The Benefits of Archive Solutions

The major benefits of deploying an archive solution includes lower storage costs, lower backup costs, improved access to information, and simplified management. Over time, the operational and financial outcomes are substantial.
Lower Storage Costs

The combination of tiered storage and dynamic management lowers storage costs. When policy-based decisions are applied automatically to determine data locations and recall, the administrative burden is removed and the savings on storage can be substantial.

In an archive tiered storage solution, a typical content distribution scheme reflects a small disk cache tier, disk archive tier for frequently accessed files, and digital tape for the majority of the data. It is always advisable to have two geographically separated copies of any files for data protection. Digital tape copies are typically used for remote storage requirements. Digital tape storage is an integral part of a tiered solution because it provides the least expensive, most scalable storage medium for the 80 percent of data that is rarely accessed after 90 days. A recent Clipper Group TCO study, which includes equipment, media, maintenance, energy, and floor space considerations, concludes that disk costs 26 times more than a digital tape based solution for a long-term archive. Additionally, because tape’s areal density improvements are expected to outpace disk almost two to one over the next few years, tape’s TCO advantage over disk is expected to increase over time. This is why the optimal archive environment deploys both disk and tape.

Lower Backup Costs

When implementing any storage strategy, enterprises need to understand how the integration of an archive with a backup environment can provide savings up to 80 percent over legacy backup-only solutions. In general, the less data backed up, the lower the costs of data protection. With the integration of an archive and backup solution, organizations can reduce the amount of data they are continually backing up. By separating less frequently accessed backup data from the legacy backup environment and sending that data to the archive, organizations can reduce the amount of data they are backing up, ultimately reducing costs.

The following example illustrates how an organization can save up to 80 percent over legacy backup-only solutions. Consider an environment responsible for backing-up 100 terabytes daily with the following requirements:

- Daily incremental backups are kept for 30 days, weekly full backups for 90 days, and monthly full backups for one year.
- Data that is less than 90 days old is kept on disk.
- All full copies of data go to tape.
- The data has a 10 percent rate of change

After one-year the raw storage requirement for backup is more than 4 petabytes. However, if 50 percent of the data can be moved to an archive environment because it is no longer changing but

---

4 For more information, visit [http://www.clipper.com/research/TCG2013009.pdf](http://www.clipper.com/research/TCG2013009.pdf)
needs to be retained for long-term retention requirements, the overall backup storage requirement drops to only 2.1 petabytes, and the archive environment with two copies of the data requires 100 terabytes. Combined, the raw storage requirement is 2.2 petabytes—that's nearly 50 percent raw storage capacity savings, even with two copies of the archive data.

The capacity savings continue to grow as more data is sent to the archive and less data is sent to the backup environment. When raw storage capacity savings are translated to TCO savings, as described in Figure 1, a combined backup and archive strategy can provide savings up to 80 percent over backup-only solutions. As demonstrated below, using tiered storage always provides savings over a disk-only solution, but the greater the percentage of data that can be removed from the backup environment and sent to a true archive environment, the greater the TCO savings.

**Save up to 80% over Backup Only Solutions**

*Backup to Archive and Disk to Tape Ratios Affect TCO*

![Chart showing TCO savings]

Figure 1. Back up to archive and disk-to-tape ratios affect TCO.

**Improved Access to Information**

For many archives, access to information is a keystone of the strategy due to compliance requirements. Information must be recalled in a very specific amount of time and must be verifiable. In a backup environment, recall costs are high. In order to restore a file with an old backup copy, it’s necessary first to find the backup set and then to restore the individual file(s). If the data has been deduplicated, it must be rehydrated as well, which can be time consuming. Items in archives are stored at the file level, and the application always has access to the file(s). It’s simple and easy to retrieve a file or file segment whenever the application requests it. In an archive, even if the initial file recall fails, the archive application search function offers a secondary route to the file. And, if two copies of the data exist, the archive application automatically pulls the next copy of the file and returns it to the host application.
Simplified Management

Throughout the archive data lifecycle, policies that automate data movement lower the administrative burden substantially.

Archive software allows users to set policies using a myriad of factors that control when a file is archived, how many copies are made, when the data can be released from the primary system, how many versions of a file to keep, which copy to recall first, when to migrate the data from one media type to another, and even when to expire or delete the files. Advanced functions can control quotas that limit the size of certain users and offer queue management for prioritizing application access to the archive. Once policies are set, the archive software automatically implements them.

Oracle’s Approach to Archiving

Oracle’s archive solutions allow enterprises to meet their specific long-term data retention and protection requirements as well as provide business value. Flash, disk, and tape storage products work hand in hand with software to contain costs, simplify backup and recovery, provide efficient search/access, and support timely retrieval. Also, Oracle’s approach solves the most common challenges when deploying an archive—TCO, data retention, data corruption, and obsolescence—so the investment is both smart and safe.

Achieve Compelling TCO

The TCO of an archive solution is dependent upon several factors, such as data growth rates, migration policies, retention policies, and mix of storage tiers. The previous example of data protection versus archiving demonstrates that the more data that can be migrated from primary storage into the archive environment, the lower the TCO. Even within the archive, faster recall requirements often dictate the amount of data stored in the disk cache or within the disk archive. It’s common for performance reasons to put the metadata in flash, so a typical archive environment includes tiers of flash, disk, and tape. By adopting a tiered storage archive, customers can improve performance while decreasing their total cost of ownership. Here are a few enterprise examples.

One Oracle customer, the USC Shoah Foundation Institute, which maintains one of the largest digital libraries in the world, saved millions of dollars in video replicating and replacement processes for an instant return on investment and saved 35 percent of the cost of staying on previous technology by moving to the Oracle solution.

Another Oracle customer in the meteorological industry, Deutscher Wetterdienst, saved $1.3 million due to the backward compatibility of the Oracle archive and increased storage capacity by 400 percent compared to the previous tape library.

And, finally, an Oracle customer in high-tech tripled the amount of critical customer information it was able to store, cut data retrieval times by 30 percent, cut cartridge storage costs by 20 percent, and reduced administrative costs because the system can be maintained by one person instead of three people.
Retain Data for Years or Forever

The ability to scale an archive’s performance and capacity is necessary to sustain enterprise growth. Each element in Oracle’s archive solutions contributes to a virtually unlimited archive size while maintaining lifetime access to data. Oracle’s archive data is always accessible even while adding or removing storage devices. Figure 2 displays the maximum scalability within each system in Oracle’s storage hardware product lines.

Figure 2. Maximum scalability within each of Oracle’s storage product lines

Oracle ZFS Storage Appliance ZS3 Series is the next generation of application engineered storage systems providing unique Oracle software integration, record-setting performance, and superior efficiency. The ZS3 Series includes innovative technologies that deliver significant performance advantages while lowering capital costs: Hybrid Storage Pool, an intelligent and adaptive multi-tier caching architecture; a multi-threaded SMP OS that fully takes advantage of multi-core CPUs; and advanced deduplication and compression. Unique management tools provided by DTrace Analytics software enable fine-grained IO and VM visibility for quick issue determination and resolution, reducing complexity and operating costs. This capability is also available on Apple iPhone and iPad through Oracle’s Sun ZFS Appliance Monitor which provides mobile monitoring and the ability to check storage services, logs, I/O statistics, real-time analytics, component status, faults, and recommended repair procedures.

An application-aware storage solution, Oracle’s Pillar Axiom 600 storage system enables enterprises to optimize performance based on application I/O needs. The Pillar Axiom 600 facilitates policy-based sharing of storage resources, CPU, capacity, and cache. Patented quality of service (QoS) architecture ensures predictable performance and efficient consolidation. The Pillar Axiom 600 requires no special training to deploy, provision, manage, and maintain tiered storage, and the modular design enables nondisruptive near-linear scaling of storage capacity and performance.
Oracle’s tape solutions include a full line of StorageTek tape libraries, StorageTek T10000 and StorageTek LTO tape drives, and tape media that together offer the most scalable storage available today. Oracle’s tape libraries scale from the entry-level StorageTek SL150 modular tape library with 300 slots, to the midrange StorageTek SL3000 modular library system with almost 6,000 slots, to the industry-leading enterprise library, the StorageTek SL8500 modular library system that scales to over 100,000 slots. The StorageTek SL8500 modular library system is the world’s first exabyte storage system. With advanced library management software the StorageTek solution can scale to 3.2 million slots or 27 exabytes of data, all managed behind a single pane-of-glass. In addition, Any Cartridge Any Slot technology in the StorageTek SL8500 modular library system and the StorageTek SL3000 modular library system protects media investments by enabling mixed media use.

The recommended tape drive for archives is Oracle’s StorageTek T10000D tape drive because it maximizes archive footprints with its 8.5 terabyte maximum capacity feature and leading 252 megabyte per second throughput. The StorageTek T10000D tape drive provides data integrity validation, linear tape file system (LTFS) support, built-in encryption, and WORM technology.

Eliminate Data Corruption

An essential component of an archive is integrity checking, which provides assurance that the data in the archive has not been altered or compromised. As data is moved between tiers of storage in the archive—for example, between the application server, flash, disk, and tape—it is critical to check that the data has not been corrupted as it is sent from the server and stored on disk and digital tape.

In Oracle’s archive solutions, software, and hardware work together to provide continuous archive validation. Validations are performed on data throughout its lifecycle, upon creation, routine access and automatically through periodic validity checks. Oracle’s disk systems provide end-to-end data integrity for data residing in disk cache or in the disk archive. Oracle ZFS Storage ZS3 appliances provide automatic data integrity checking and correction with block-level checksums. When a corrupt block is identified, ZFS software automatically repairs it with another copy, providing for self-healing capabilities at the block level. For data on digital tape, Oracle’s StorageTek T10000D tape drive, which is designed for 24/7 enterprise data centers, checks a file's cyclic redundancy checks (CRCs) internal to the drive as each record is received from the archive manager. In conjunction with StorageTek software, the StorageTek T10000D tape drive also has the ability to run automated periodic CRC checks on archived data, without transferring data back to the host. The integrity of the data is also checked on reads to ensure the proper CRCs are in place.

Periodic media validations are available for data stored on tape through Oracle’s StorageTek Storage Archive Manager or Oracle’s StorageTek Tape Analytics. Customers can automate periodic media validations based on health indication, days since last accessed, random sampling, volser ranges and more. This ensures data is still accurately recorded on tape and will be available to the application when recalled.

Prevent Obsolescence

An archive solution must provide a route to the data in 10, 30, or 100 years, but both software and hardware can be subject to obsolescence over that timeframe. Oracle addresses obsolescence concerns
with open storage formats and data migration. Open archive formats, like LTFS, enable future access
to data without the application that wrote the data, an essential advantage for the longevity of the
archive. When data is written to tape in an open format, the data is accessed the same way that data on
flash or disk is accessed—with the use of an open-source driver. StorageTek Storage Archive Manager
is one of the few widely adopted archive applications that supports open formats. It stores data using
the tape archive (tar) format, which has been in use for more than three decades.

If data is written to an archive in a proprietary format (without an open format), the application that
wrote the data is required to find and recall the data. Open formats eliminate the requirement for
historical applications to be present when recalling data, ultimately reducing the barriers to accessing
archival data in the future.

The newest open format technology for tape is LTFS, introduced in 2010. The format is being adopted
quickly by media and entertainment customers because of its interchange and portability benefits.
Oracle’s newest LTFS software offering is StorageTek Linear Tape File System, Library Edition, which
provides a file system interface to an entire tape library, allowing users to see every file in a large tape
library just as though the data were stored on flash or disk.

Oracle is committed to developing and supporting open formats and is taking a leadership role in this
area. Currently, Oracle co-chairs the Storage Networking Industry Association (SNIA) committee that
is driving standardization of LTFS.

Data migration is driven by data expiration and by technology refreshes. Data migration can be
automated by the archive manager software, which migrates data from one storage medium to another,
thereby ensuring constant access.

For technology refreshes, it is important to remember that with a media lifespan of 30 to 40 years,
StorageTek T10000 media also addresses concerns about migration, long-term retention, and TCO
because, compared to disk, data won’t have to be moved as frequently during its lifetime. Since the
average tape library system sits on a customer’s data center floor for around 10 years, migration cycles
are one quarter of a typical disk systems. Fewer technology refreshes and automated migrations
further improve customer TCO models.

Customer Use Cases

Oracle’s archive solutions are deployed by entry-level customers all the way to very large enterprises,
and in all cases, the solutions provide the tiered storage, performance, scalability, data integrity, open
storage formats, and data migration capabilities discussed above. Oracle’s solutions can meet the needs
of specific industry archive requirements, as well. For example:

For more information on LTFS, download a free copy of *LTFS for Dummies*. 
• Content management—management of the digital content lifecycle, including creation, capture, storing, versioning, indexing, searching, distribution, publishing, and retention.

• Media and broadcast—video, image, and audio repositories that support content creation, capture, digitization, editing, distribution, and preservation.

• High-performance computing/big data—support for data modeling and simulations based on structured and unstructured data to solve complex problems.

• Healthcare—digital repositories that focus on electronic medical records, health exchange data, and medical imaging.

• Preservation/higher education—maintaining or restoring access to artifacts, documents, and records through the study, diagnosis, treatment, and prevention of decay and damage.

• Video surveillance—real-time monitoring and control of sensitive areas using cameras with centralized recording and long-term video retention.

Two use cases, one in media and entertainment and one in healthcare, are highlighted below.

T3Media Licenses Sports, News, and Creative Footage

Ongoing technology improvements in the media and entertainment industry are creating both opportunities and challenges due to increases in the quality and quantity of digital assets. Companies are using high-resolution formats for audio, video, and graphics and producing more digital assets to satisfy broadcast and consumer demand. Higher bandwidth capabilities allow the assets to be distributed and repurposed more easily than ever via the Internet or private channels. Many assets have a longer useful life, for continuous online viewing or for specific licensed uses, and this requires longer-term storage and protection. To make the best use of their assets in a competitive marketplace, companies need to be “lean and green”—to be able to store, access, and share assets efficiently and cost effectively. For this reason, media and entertainment businesses like T3Media are turning to archives instead of traditional storage solutions.

T3Media serves the media and entertainment industry through its Platform, offering sophisticated access and monetization services for enterprise-scale video libraries. The company licenses sports, news and creative footage to producers in advertising, entertainment, publishing and emerging media. Each year, T3Media uploads tens of thousands of hours of programming content and more than 100,000 media assets. As Mark Lemmons, chief technology officer, sees it “What is the difference between the content and the archive? There is no difference. There is no distinction between a live asset and the archive.”

T3Media needed to deploy a cost-effective, cloud-based, digital storage environment consisting of primary and secondary storage environments that could scale to an exabyte level to meet clients’ future needs. Secondary storage is not viewed as backup. According to Mark Lemmons, “Having multiple copies is as much about ‘load balancing’ and ‘managing bottlenecks’ in content delivery as it is about traditional back-up.” For digital assets that have a lifespan of 30 years or more, the company needed to ensure short- and long-term integrity for the content it is trusted to store. Other requirements included a highly available environment for its clients and swift digitization.
The solution consists of StorageTek Storage Archive Manager, Oracle’s StorageTek SL8500 modular library system, Oracle’s StorageTek T10000C tape drive, and Oracle Solaris operating system. On this foundation, T3Media built its three-tier storage environment—metadata, disk, and tape—this includes more than ten petabytes of client data on tape alone. T3Media did not always view tape as an important tier of storage, but now sees it as integral. “I was dragged kicking and screaming to tape. I thought customers would need content right now, and I was right. They do need it now, and we are providing it with tape.” states Lemmons. “Don’t get tricked into thinking disk is fast and tape is slow. We are meeting customer expectations on tape.”

Now T3Media operates with these new capabilities:

- Access any tier from standard calls to the file system
- Add storage incrementally on a per-client basis, a competitive advantage in the cloud-storage sector
- Validate fields as content is transferred from one storage tier to another
- Achieve granular, incremental control for the purpose of maintaining and pricing storage units for clients

Compared to its legacy storage environment, T3Media realized a three- to five-times performance improvement, and the company saved at least 25 percent in fiber duplication, card infrastructure, and infrastructure management costs.


Novant Health Stores Nearly 1 Billion Files

Healthcare organizations are dealing with massive data growth, regulations, and an industry trending toward a patient-centered service model. Advancements in medical imaging and other medical specialties are causing more data to be generated than ever before, and regulations require data to be stored and protected for a patient’s lifetime plus ten years. Currently, many medical departments or facilities develop and maintain their own storage solutions, leading to varied vendors, standards, formats, and strategies. This siloed approach makes it very difficult for healthcare providers to collaborate and share data, a requirement in the patient-centered model. And, silos increase cost because data management is not centralized. These factors are causing healthcare organizations like Novant Health to re-evaluate their storage solution with an eye toward standardization, retention, accessibility, data integrity, and cost.

Novant Health is a not-for-profit integrated system of 13 hospitals and a medical group that consists of 1,124 physicians in 355 clinic locations, as well as numerous outpatient surgery centers, medical plazas, rehabilitation programs, diagnostic imaging centers, and community health outreach programs.
To operate efficiently in a regulated environment, Novant Health needs to be able to store and duplicate medical data, including radiology images, at an affordable, long-term cost. Its objectives for an updated solution included the following:

- Eliminate the need for clustered servers to simplify management and accelerate file system backups
- Provide disaster recovery replication of the archive across two data centers
- Satisfy multiple applications with standardized technology and cost efficiencies, along with the flexibility to allow users to access data quickly.

The solution, which consists of StorageTek Storage Archive Manager, the StorageTek SL8500 modular library system, the StorageTek SL3000 modular library system, and StorageTek T10000C tape drives, allowed Novant Health to migrate 65 applications, including radiology and cardiology imaging, desktop shares, streaming video/audio, and insurance documentation. As a result of two-to-one lossless compression capabilities offered by the new system, the company realized significant cost efficiencies. And, it is able to tailor any performance requirement to satisfy an application’s needs. For example, disk access performance for data between six months and three years old, or automatically making multiple tape copies to reduce risk.

According to Robert Dick, Datacenter and Archiving Resources Lead, IT Technical Services at Novant, “Oracle’s Sun Storage Archive Manager—in conjunction with its StorageTek tape drives and modular library system—has enabled us to develop a cutting-edge, high performance long-term archiving solution that easily and affordably handles our high data volumes while providing the flexibility our users need when accessing critical healthcare files.” Currently, the solution stores more than 900 million files with combined storage of more than 775 terabytes.


**Conclusion**

Data growth, regulations, content digitization, lean budgets, and the need for storage efficiencies are all complicating the operational and financial pictures of enterprises, which must store and protect information for years. While data retention challenges are growing organizations are also looking to leverage their data to gain insight into and grow their businesses. In the past, many enterprises have relied on their backup solutions for long-term data storage and protection, but this approach is not recommended. The best practice—and business necessity—is to develop and implement a scalable archive solution.

Oracle’s customized archive solutions incorporate software, disk, and tape into tiers of storage that are managed automatically based on policies. In addition to driving down costs and delivering compelling TCO, an Oracle archive solution eases the concerns about data growth, data corruption, software and hardware obsolescence, and retention requirements.