
Archive Security in a Tiered Storage Environment

A practical assessment of the security implications of using different storage technologies for long-term data archiving



Author: Marketing
Version: 1.2
Date: 24Aug06
Status: Final
Distribution: General

Table of Contents

1	The Tiered Storage Model	3
2	Archive Data and Storage Technologies	3
3	Archive Security Defined	4
4	Controlling Access	5
4.1	Encryption	5
5	Data Integrity	5
6	Authentication	6
6.1	WORM Recording	6
6.2	Media Identification	6
6.3	Digital Signatures	6
7	Disaster Prevention and Recovery	7
7.1	Backup	7
7.2	Replication	7
8	Authorized Destruction	8
9	Validation	9
10	Data Availability	10
10.1	Media Life	10
10.2	Application Interface and File Formats	10
10.3	Technology Roadmap	11
11	Summary	11

Table of Figures

Figure 1 - Tiered Storage Architecture	3
Figure 2 - Archival Storage Technologies	4
Figure 3 - Principle Security Considerations	4
Figure 4 - Archive Replication Architecture	8
Figure 5 - Media Security Summary	9

1 The Tiered Storage Model



The use of tiered storage is becoming increasingly common within many organizations. The popularity of this architecture is driven by the need to meet the, often conflicting, needs of on-demand information access balanced against cost reduction and complicated by regulatory compliance and risk management. The proliferation of different disk, tape and optical technologies in combination with new software solutions enables significant cost savings by matching appropriate storage architectures to specific business requirements. The benefits of a tiered storage strategy can be very compelling, but it also introduces new security challenges particularly with respect to the long-term retention and accessibility of corporate information.

A well-constructed tiered storage strategy recognizes that data has different value to the organization throughout its lifecycle and optimizes the use of different technologies by distributing data to the appropriate storage devices at key points in the life of the information. Sometimes referred to as Information Lifecycle Management (ILM), the process seeks to capitalize on the strengths of different storage technologies in a way that meets business objectives while minimizing acquisition and operating expense.

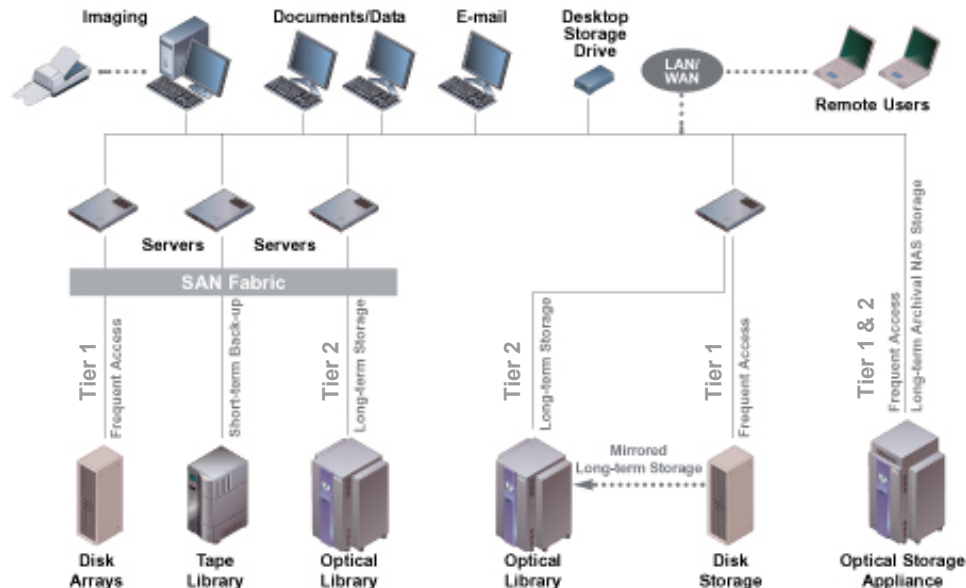


Figure 1 - Tiered Storage Architecture

This tiered storage strategy can be very effective. However, unique security considerations are introduced anytime data moves from one location to another or when the lifecycle of the digital records demands that they be retained for years or decades. This paper seeks to expose and explore some of the more important security issues when archiving data for extended periods of time as part of a tiered storage strategy.

2 Archive Data and Storage Technologies

The necessity to understand the use and value of different data types is fundamental to a tiered storage model. If data cannot be accurately classified, it is very difficult to develop a strategy that will ensure appropriate storage devices are employed throughout the lifecycle of the data. Fortunately, the identification of archive grade data is reasonably straightforward in most environments making it a good place to begin for those organizations developing ILM storage strategies.

Archive data is distinguished from other types in several ways. Also known as “fixed content”, it is no longer being actively created, modified or accessed. In most cases it is essential that archive data be retained unaltered, since the ability to demonstrate authenticity is vital to the integrity of the archive. At the same time, less frequent access and discovery requirements create the opportunity to offload archive data from expensive, high performance hardware to less expensive and more secure alternatives.

The most common technologies used in professional archive environments are magnetic disk, magnetic tape and optical storage. Discounting strictly consumer products in each of these categories the leading mainstream contenders include the following product types.



Figure 2 - Archival Storage Technologies

When used in professional archive applications, all of these technologies can be combined with other components to create more redundant and automated solutions. These include specialized disk based archive systems that typically use high capacity SATA drives, as well as automated tape and optical libraries. While each has their own strengths, they also present specific security challenges.

3 Archive Security Defined

“Security” can mean very different things depending on perspective and application. For the sake of this document, the term is being used in a more general sense. This includes a number of major considerations required to secure the integrity and availability of archived information over time.



Figure 3 – Principle Security Considerations

This list includes more common security issues such as access control and data authentication, but also factors that are unique to long-term record storage such as technology evolution and data migration. Archive specific security issues are important to consider because even if the integrity of the information remains intact, the lifecycle of the data is compromised if it can no longer be read by an application or retrieved from the storage device.

4 Controlling Access



The primary purpose of controlling access is to prevent the theft or the tampering of already archived data. To achieve an acceptable level of access security, both physical and electronic tools should be employed. This often has less to do with the archive storage solution and more to do with the policy and procedures of the organization.

Controlling access begins by physically preventing unauthorized personnel from approaching the archive system, which can be achieved through secure facilities and biometric identification systems. For those users that can access the computer network, carefully maintained user accounts are essential to ensure that unauthorized users do not compromise the archive environment.

While this level of security is worth mentioning for the sake of completeness, there is nothing particularly unique regarding the access considerations for an archive environment. Generally speaking, it is exposed to the same risk profile as other storage tiers within the overall architecture and should be protected accordingly.

4.1 Encryption

Another way of controlling access to individual archive records is to prevent them from being viewed by unauthorized personnel through file level encryption. Cryptographic hash functions based on MD5 and SHA-1 mathematical algorithms prevent anyone from opening a file without the proper electronic key. File encryption techniques are independent of the storage media and can be used with disk, tape or optical. In direct response to the market demand for more secure data storage, encryption is being incorporated directly into next generation tape and optical drives. It is important to note that while the sophisticated mathematics used in encryption may be sound today; there is no guarantee that it will not be compromised in the future. It could be dangerous for companies to rely exclusively on encryption particularly in the case of archive data, which must be secured for many years.

5 Data Integrity



In order to control the authenticity and access to valuable archive data the number of copies of an individual record may be strictly limited, making the integrity of each record on the physical storage media absolutely critical. For this reason, it is important that the chosen storage technology employs verification procedures so that all records written to the media are correct and complete.

When data is written to a disk or to professional optical storage like UDO, it is fully verified so the application or the user can be certain the record has been accurately recorded. This is not the case for tape media or consumer optical products like DVD and Blu-ray. When records are written to tape or DVD, there is no way of being certain that the data has been properly recorded without reading the data back; a very time-consuming and impractical process for a professional archive.

Once written, the quality and robustness of the media plays a major role in maintaining the long-term integrity of data. Tape is a fragile material that degrades over time and must be stored in a carefully controlled environment to minimize data loss. If data is to be stored on tape for a long period of time, it should be carefully monitored and maintained. Corporate archives should also be very cautious about using low cost, high volume, consumer grade media such as DVD and Blu-ray. The manufacturing priority for consumer media is cost, not quality. Buying cheap media is a false economy that can result in considerable data loss.

There are risks to storing data on any media type, but given the value of archive data it is a significant security lapse if the highest media quality is not used and the integrity of data is not fully verified at the time it is initially written. This first level of data integrity is often overlooked,

but it is an important consideration given the operational sensitivities of an archive environment.

6 Authentication



One common archive requirement is the need to demonstrate the authenticity of individual records. For many organizations, this is being driven by senior executives in order to comply with industry regulations on record retention and to satisfy their own internal policies on risk management. The inability to convince a court of the authenticity of electronic records in the face of a legal challenge can have crippling political and financial consequences. With absolute authenticity as the ultimate goal, there are a number of very competent strategies that can be used to provide the maximum possible protection beginning with the physical recording process.

6.1 WORM Recording

Disk, tape and optical all offer WORM (Write Once Read Many) versions of their products, which are designed specifically to prevent the alteration of data once written to the media. However, all WORM products do not provide the same level of protection. Only optical storage including UDO provides physical WORM functionality. Phase Change recording used on UDO true Write Once media cannot be modified in any way. By contrast, both tape and disk use rewritable magnetic media and employ WORM emulation through software and firmware controls. While it can be argued that WORM emulation is adequate for some applications, true Write Once technology provided by UDO provides a higher standard for secure storage.

6.2 Media Identification

The ability to identify and track individual pieces of storage media can further strengthen the case for record authenticity. The best example of this is a unique, software readable, serial number provided on each piece of UDO media. This serial number allows applications to track when and on which specific piece of media records have been archived. This makes it very difficult for data to be copied off UDO, changed and rewritten to another piece of media. This ability to identify and protect against the introduction of fraudulent media is not possible with disk or tape storage.

6.3 Digital Signatures

Digital signatures can be a very effective way of identifying who created an archive record, when it was created, and if it has been modified. They normally use public and private key technology to clearly identify the originator of the signature and to prevent misuse by unauthorized individuals. Like encryption, digital signatures operate on a file level so can be used together with disk, tape or optical storage technologies. However, Digital signatures cannot preclude the creation of a fraudulent record generated by an authorized person making it critical to have other physical and electronic checks within the overall system.

Authenticity is clearly a multifaceted requirement since no single product feature or capability can guarantee absolute record authenticity and immutability. The best an organization can do is provide sufficient evidential weight to show, beyond any reasonable doubt, that their archive records are authentic and have not been modified. Any archive strategy seeking to address this issue should use a combination of WORM recording, media identification, encryption, and digital signatures.

7 Disaster Prevention and Recovery



Protecting data from damage and recovering a system in the event of disaster is another key consideration to the overall security and availability of an archive. In first tier primary storage environments, active data is often protected by short-term backup and on-going duplication or system replication. Some of these same strategies can be used to protect an archive storage tier, but how it is deployed is greatly dependent on the criticality of the archive to the operation of the business often defined by how quickly the data needs to be accessed.

Here again, the behavior and priority of archive data can differ dramatically from the service level requirements of transactional or active data. If active data is unavailable it can have serious business and financial consequences whereas a temporary interruption to the archive may have very little impact on the business. This difference means that while protecting the existence of archive data is absolutely essential, putting in place an expensive high availability infrastructure may not be required.

7.1 Backup

The need to provide short-term backup for an archive is very dependent on the selection of the archive media. Even with the redundancy provided in disk based archive products, they are the most volatile because of the mechanical nature of magnetic disks. With disk archives it is possible to have multiple disk failures that can result in significant loss of large pools of archive data. SATA disks are the most common drives used in archives. While these provide excellent capacity, they have much lower duty cycles than SCSI and Fibre Channel drives, increasing the risk of multiple disk failures. The only way to protect against this is to implement some form of short-term backup or replication to recover the lost data.

One of the key financial benefits of archiving data off primary storage is to eliminate the need to repeatedly backup static data, shortening backup windows and reducing administration overhead. Being forced to implement a backup strategy for a disk archive negates any financial benefit, imposing additional cost and complexity, and can compromise the authenticity of the individual archive records.

With tape there is also a high risk of massive data loss. Individual tapes can be damaged, degraded or broken. The high capacity of tape technology actually compounds this risk since damage to one tape can cause the loss of hundreds of gigabytes of archive data. Tape damage is a very real concern since it is a fragile media that wears with use and is not highly resistant to a wide range of temperature and humidity conditions. While tape based archives may not require a short-term backup strategy like disk, multiple copies of duplicate media are essential and the tape media should be periodically retensioned, monitored and rewritten to new tapes to ensure archive data is not lost.

Optical technology provides the most stable of the three storage technologies, as it is a non-magnetic, non-contact media. Once data is written on UDO there is little risk of data loss in normal operating conditions, eliminating the need for short-term backup. Unlike tape, UDO media has a very high tolerance to environmental conditions and requires no on-going maintenance.

7.2 Replication

The possibility of site disaster is a distinct risk regardless of the storage technology. Whether using disk, tape or optical, a duplication or replication strategy providing offsite data copies is necessary to preserve the archive in the event of a major site failure. The way this is achieved is again heavily influenced by the storage technology.

Tape and optical archives have the option of installing either replicated libraries at a second site to deliver quick failover, or one can choose to use less expensive offline media copies stored in a secure location such as a fireproof safe. The choice of offline media will not

provide rapid failover, but does offer a very inexpensive disaster recovery capability. The strategy that is chosen by a given organization will depend on the service level required for the archive data in the event of primary site failure and their available budget.

Because disk technology is not removable, replication can only be achieved through the installation of a second fully redundant disk archive. While this strategy will provide the best possible performance, it is by far the most expensive disaster recovery strategy and may not be justified by the service level requirements of the archive data.

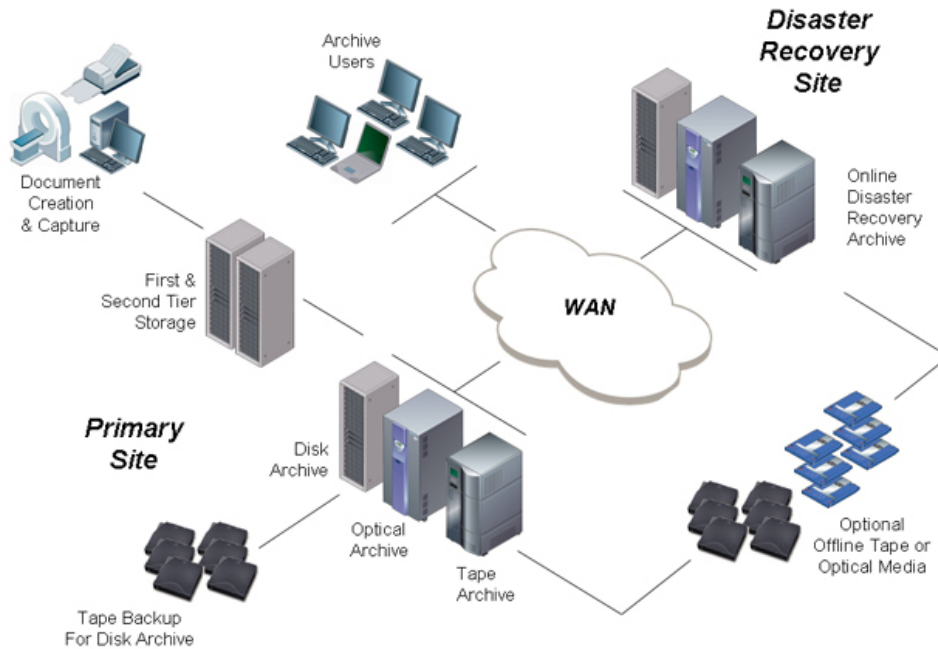
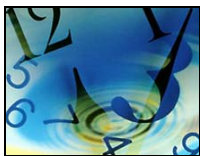


Figure 4 - Archive Replication Architecture

It is interesting to note that UDO also has a special feature not available with tape or disk designed specifically to protect offline media from unauthorized access. Called UDO Guard, it is a software key mechanism assigned to each piece of media that prevents unauthorized users or applications from reading any of the archived data. This feature is one additional tool that protects the contents of the UDO media from theft or mishandling especially during transport and storage. This is a key security consideration when working with removable media.

8 Authorized Destruction



Depending on the nature of the business and the data, destruction of archive records can be as critical to system security as their preservation. As the lifecycle of business records are being defined by industry regulations, organizations are required to retain more data for specific periods of time. A good example of this is emails. Generally speaking, email represents a greater liability than asset to an organization, and they would prefer to destroy them as soon as the statutory retention period has expired. Retaining them longer can actually increase their risk profile. As a result, many companies are now very concerned about the secure destruction of archive records when they reach their legal end-of-life. There are two primary methods for data destruction, logical and physical.

Records archived on disk, tape or optical are normally managed through some form of online index or file system. This makes it possible to delete the file reference, rendering the data

inaccessible without actually touching the media. If a record has been written onto the archive using a key based encryption technique, the key can also be destroyed for further protection. Either of these techniques may be adequate for some customers, but others require the physical destruction of the data.

With UDO, physical record destruction is possible using Compliant Write Once media, which provides WORM authenticity but permits the physical destruction of individual records. The nature of Phase Change records means that the UDO destruction process is very quick and it leaves no residual or shadow image that could be recovered.



This type of dynamic physical destruction is not possible with tape media, making discrete record deletion impossible. The only way to physically destroy individual records on tape is to copy all the contents of the tape back to disk, remove the unwanted files, rewrite the data to a new tape, and physically destroy the old tape. This is an extremely impractical process that exposes the authenticity of the records to a number of serious risks.

Record destruction on disk is handled in a different way. When records are “deleted” on disk the block address for the file location is initially unassigned, allowing the space to be used for future write operations. The file is still intact but the data is not accessible. Since disk is a random access, rewritable media it is possible to physically destroy individual records, but this can only be done by repeatedly overwriting the area in question to ensure that the record cannot be recovered. There are commercial and military standards for overwriting data on magnetic disk that involve using specific data patterns overwritten between 3 and 30 times. Secure data destruction on magnetic disk is possible, but it can be a time consuming and awkward process.

A common problem and an additional security risk with disk archives is that the deleted records often continue to exist on tape backups. Proper destruction procedures require that records be destroyed in the archive and on all tape backups. Numerous examples exist where this problem has been exposed during litigation. Records that were “destroyed” are later recovered from backup media, bringing to light undesirable information and severely compromising the credibility of the organizations record management procedures.

Companies using offline media for disaster recovery or making archive backups need to make additional provisions for data destruction. If physical data destruction is essential, media will have to be brought back online periodically so that targeted records can be removed from the offline media copies. It is important that all copies of the destroyed records are no longer available.

<i>Media Type</i>	<i>WORM Recording</i>	<i>Record Level Destruction</i>	<i>Media Identification</i>
Magnetic Disk	Emulation	Multiple overwrites	Not available
Magnetic Tape	Emulation	Not available	Not available
UDO Optical	Physical WORM	Physical destruction	Unique media ID

Figure 5 - Media Security Summary

9 Validation



Security processes and procedures are only as good as the organization’s ability to validate them. If a company cannot prove that they have managed their records as part of a well-controlled process, they are seriously undermining their ability to comply with regulations, as well as the legal authenticity of their archive data.

All secure environments must have the ability to track and audit the management of their archive data. This should include how data is committed to the archive, who has accessed the data, changes to the status of the data throughout its lifecycle, and authorization for data destruction at end-of-life. The last issue is a particularly good example since it could prove important to justify in court the corporate policy for data destruction and to document who authorized destruction and when. The existence of a well-defined archive management

process with audit capability builds evidential weight for legal defense and generally strengthens the overall security of the archive environment.

10 Data Availability



In addition to hard security issues there are other considerations that act to ensure long-term access to archive data. An archive cannot be considered “secure” if the data is, for whatever reason, no longer accessible. These considerations are a function of the extended length of time that archive records need to be retained and are of particular concern to an archive environment.

10.1 Media Life

Because archive data is retained for many years, the longevity of the storage media plays an important role in lifecycle of the data. If records are kept for only a few years, it may be possible for data to remain on the same piece of media until end-of-life. However, if records must be preserved for decades, it will be necessary to repeatedly migrate the data to new technologies overtime. In this case, the objective should be to minimize the frequency of data migration in order to control cost and mitigate the disruptive impact on record authenticity.

Generally speaking, the longer the life of the media, the less frequent data will need to be migrated but this needs to be balanced against a potentially shorter hardware and software obsolescence cycle. For example it doesn't matter if the media life is 100 years when the hardware and application life is less than 10 years. What's is important is that within a normal obsolescence cycle the integrity of the data on the media remain secure. To this end, the more robust the media, the more certain one can be that the data will safely survive any practical migration period. Companies should not be forced to migrate data because of the failure or deterioration of the storage media.

Disk has the shortest media life with operating cycles of only several years. This is particularly true of lower duty cycle SATA drives that are commonly used in archive environments. These products address this issue by combining the disk into redundant (RAID) configurations so when one disk fails, they can swap it out and continue running. Over the course of many years, this means that dozens of disks may need to be replaced in order to preserve the operation of the system and if too many disks fail simultaneously, data will be lost.

Tape offers a longer life, but as noted earlier it is a fragile media, which requires monitoring and maintenance if data integrity is to be ensured. Leaving tapes untouched for several years is dangerous to the archived data. Tape is often used as an archive media, but the overhead of monitoring and refreshing tapes must be factored into the security considerations of a tape archive.

UDO provides the longest and most stable media life when compared to disk and tape. It can securely preserve archive data well beyond a normal obsolescence cycle with a high degree of confidence and with no reoccurring media maintenance. This affords significant security benefits because it provides data integrity while reducing both migration frequency and any impact on record authenticity.

10.2 Application Interface and File Formats

The hardware and file interface represent another consideration for the access security to any long-term archive. There are two primary ways of interacting with an archive, via a standard file system or using a proprietary Application Programming Interface (API). Both these approaches have technical pros and cons, but from a long-term security point of view it is always best to make use of “industry standard” interfaces to maximize support. Using vendor proprietary interfaces introduces an additional risk to long-term access and vendor changes may force data migration or maintenance sooner than desirable.

This same logic also holds true for the file format used to store the archive records. Many organizations are looking to archive their data in a file format that is most likely to be readable in many years time such as ASCII text or PDF. They are specifically looking to avoiding file formats that are very application or version dependant.

10.3 Technology Roadmap

Long-term product availability and support is also critical to archive security. If storage vendors have short product lifecycles and support programs, it can be very difficult to keep an archive running and may force the premature migration of the data.

Disk vendors have the shortest product and support lifecycles. Their primary market is storage of short-term active data and not archival storage, so it is typical for them to launch and discontinue a particular model in less than 12 months. If the disk archive is not capable of upgrading to a new generation of disk technology, the customer may be forced to replace their entire system when the original drives are no longer available. This can increase the frequency of migration, jeopardize the integrity and authenticity of data and can add enormous cost and overhead to the management of the system.

By contrast, the vendors of tape and UDO have a much longer-term support philosophy with less frequent product generations, backward compatibility and extended support commitments. Customers using tape and UDO will benefit from longer product lifecycles, with media available for years or decades in the future. In most cases, these vendors understand that their products are being used for archival storage and have a business philosophy to support their customers long-term.

11 Summary

Providing data security in a tiered storage environment is a complex task and the unique requirements of an archive make this all the more difficult. In examining archive security, it is clear that the fundamental choice of the storage technology plays a significant role in identifying the risk profile. Disk, tape and optical technology can all be used, but each introduces weaknesses that must be accounted for in an overall risk assessment. Which technology to use will depend on the business needs and security requirements of each company. Choosing an inappropriate storage technology can have a major impact on security, as well as the cost and operation of the archive.

