'är-kīv to file or collect (as records or documents) in an archive or other repository: a place where historical documents are preserved.

Corporations are mandated to store and maintain records like never before. Do you have an archive plan within your Information Lifecycle Management solution that can meet these mandates? Are you paying too much and what can be done to address your needs? Before you make an archive solution selection, make sure you select the most cost effective media to meet your needs.

Who is Pegasus and where do we fit?

As the value of information changes, it makes sense to move data to different online and off-line storage media that provides the right levels of protection, replication and recovery at the lowest possible cost.

A key piece to a Data Lifecycle Management product is the archive component, as new regulations now govern how long business-related e-mail and Internet communications need to be retained and how quickly they may have to be retrieved. Generally speaking, two kinds of data need to be archived; data that must be stored and is under some kind of external regulatory guidelines, and data that is not regulated, but is stored based on internal policies for a specific period. There are some 10,000 state and federal regulations, such as the Sarbanes-Oxley Act, Department of Defense Standard 5015.2 and the Health Insurance Portability and Accountability Act, just to name a few, governing the storage, availability and destruction of information. The data lifecycle model must take both data types into consideration and direct them to different online and off-line storage media that provides the right levels of protection, replication and recovery at the lowest possible cost depending upon where they reside in their lifecycle.

Most regulations stipulate either how long data is to be available or that the data must be retrievable in its original, unaltered format. In some cases the regulations stipulate both. With this in mind, an archive solution needs to address data protection and integrity for the life of the data, operational policies and procedures to manage the data during its useful lifecycle, and a secure file system on secure media with a useable life that meets or exceeds the life of the data being stored on it.

Our expertise at Pegasus is to provide a secure Archive File System (AFS) on removable media. With our AFS, in conjunction with unalterable media, corporations can protect their data and ensure its integrity for the full lifecycle of all compliant data.

An Information Lifecycle Manager (ILM) is designed to manage data from creation to disposal. ILM provides a means to apply business-driven policies across your entire network, matching the right application to the right data classification and putting it on the right storage platform. In doing so the IT administrator can catalog and organize data according to its value, type and requirements, including availability, recovery, security, cost and compliance. Pegasus provides the connectivity within the ILM infrastructure to unalterable media, thus providing a compliant archive solution that is auditable, secure and removable.

What's an AFS and why is it needed?

Our AFS provides a software lock at the file system layer. It does not allow for sector reclamation once a sector has been written to. There is no way to erase data from the file system regardless of security level or password protection. Even if hacked into, the data remains in a read-only state, and if changed the original data is still stored on the media.

The highest level of write-once security is using an AFS with write-once media. Write-once media provides a physically permanent record stored on the media whereby there is a lock on the media to prevent data sectors from being erased, changed or altered in any way. Most hardware that supports write-once media automatically senses if the media is write-once or rewritable format. If the media is write-once it is write protected such that once a sector has been used it cannot be changed. This hardware in conjunction with an AFS is the highest level of electronic record protection for security and integrity.

What constitutes WORM?

Write-Once media comes in two forms; physical and electronic write-protect. The most common form of write-once media is electronic write-protect. Electronic write-protect media is used with hardware that has firmware imbedded within that senses if the media is write-once or rewritable. If the media is write-once, the firmware prevents the media from being written more than once per bit sector. Electronic write-protect media can be found in disk, optical and tape formats. The other type of write-once media is physically write-once. Media that is physically write-once requires a physical change at the recording layer of the media that cannot be reversed once changed. This is the most secure form of write-once media and can only be found in optical technologies.

Optical storage has been used for long-term record management for 15 years. It was the original archive storage standard for records management, and is the basis for many regulations in existence today because it has never been successfully challenged in a court of law regarding the authenticity of data stored on it. In the last few years, other forms of storage technology like tape and ATA disk RAID have tried to mimic the optical WORM standard for compliancy reasons. However, none of the newer technologies have been challenged in any of the recent judicial cases involving willful destruction or manipulation of corporate records.

How long do archives need to last?

Challenges to data trustworthiness grow in proportion to the age of the data. Any media that can ensure long-term data retention without having to migrate data to new technology will strengthen any challenges to data trustworthiness as the data ages. If, however, data must be migrated several times over its lifecycle due to technology change or technology end-of-life, then that represents a challenge to data integrity.

The need for fast access to data reduces in proportion to the age of the data. The legal necessity for quick access reduces with time. Legal discovery normally requires response in days, weeks or months, not seconds. In most cases, off-line archives meet the data access requirement for archive grade data.

Frequent migration of data represents a risk to records and demands an audit trail. It is also costly for the operation. Media that reduces the migration frequency dramatically simplifies archive management, reduces cost over time and improves the trustworthiness of the data.

What are the trade-offs of archive media?

Almost any media can be used for an archive, however, there are trade-offs associated with each media type. Random access media is more expensive than linear access media. Rotational media provides faster access to individual files, however, generally transfers large blocks of data slower than linear access media. Removable media can be stored off-line; and, therefore, the total cost of ownership goes down over time, rather than fixed disk, which goes up over time.

Table 1 reflects the archival attributes of the various media types and shows the trade-offs associated with storage technology choices. For example, tape technology is a very low cost storage medium, however, it does not provide true write-once data protection and the access time is slow. RAID on the other hand provides very fast access to data, while the overall cost of the system is high yet does not provide a means to off-line data.

Table 1

	RAID	*Tape*	*DVD*	*MO*	*UDO*
Permanent Media	No	No	Yes[1]	Yes	Yes
True Write Once Media	No[3]	No[3]	Yes[2]	No[3]	Yes
Removable Media	No	Yes	Yes	Yes	Yes
Professional Quality	Yes	Yes	No	Yes	Yes
Media Capacity	Mid/High	High	Low	Low	Medium
Read / Write Speed	High[4]	High	Low	Medium	Medium
Access / Seek Speed	High	Low	Low	Medium	Medium
Total Archive Cost	High	Medium	Low	Medium	Low

1 Bare DVD media is highly susceptible to physical damage and contamination.
2 The DVD-R and CD-R recording process does not include write verify
3 Available with software API WORM capabilities.
4 Current WORM RAID solutions demonstrate very slow transfer rates
5 DVD-R, DVD-RW, DVD+R, DVD+RW do not have addressable sectors only DVD-RAM

Magnetic disk can have WORM-like properties in a kind of softWORM API based application. The software overhead associated with hard drive softWORM is extensive. Today's read-write performance on specific API based ATA hard drive softWORM technologies is very slow, less than 500 KB/s. Optical media transfer rate in conjunction with a WORM archive file system exceed softWORM API's by several times. In addition, Optical WORM archive file systems also give users direct access to archive data. This is not so with a softWORM API applications such as EMC's Centera.

Not all removable media is created equal, or is all removable media best suited for archive applications. Removable media can be broken down in the several groups. One of those would be sector-based media vs. non-sector based media types. Tapes, DVD-R,

along with all the CD technologies, are sequentially accessed, non-sector based media that are written to by spiraling out from the center of the disk, or in the case of tape, stream from beginning to end. Starting from the beginning of the tape or center of the disk, each bit is written next to the last bit written in a sequential order, regardless if the information is related or not. This must be done, because there are no addressable sectors on the media's recording surface. The data can be read in a random sequence once written on a disk, however, the write process is sequential. Problems occur when bad areas on the recording surface are encountered during the writing process. Non-sector based media does not have any means to set aside those bad areas, and as a result, data blocks can be lost during or after the archive process.

Many removable media formats such as magneto optical, DVD-RAM, ultra density optical (UDO) and Sony's new ProData have addressing sectors on the recording surface. This allows for a file system to treat the media like a hard drive by taking corrupt sectors off-line during the formatting process and provide a means for the file system to do error correction and write verify. Most enterprise archive application of critical data rely on some form of sector based write-once media for maximum data protection.

The yellow highlighted boxes in Table 2 indicate the desired attributes in archive media.

Table 2

	RAID	***Tape***	***DVD***	***MO***	***UDO***
Permanent Media	No	No	Yes[1]	Yes	Yes
True Write Once Media	No[3]	No[3]	Yes[2]	No[3]	Yes
Removable Media	No	Yes	Yes	Yes	Yes
Professional Quality	Yes	Yes	No	Yes	Yes
Media Capacity	Mid/High	High	Low	Low	Medium
Read / Write Speed	High[4]	High	Low	Medium	Medium
Access / Seek Speed	High	Low	Low	Medium	Medium
Total Archive Cost	High	Medium	Low	Medium	Low
Sector Based	Yes	No	Yes[5]	Yes	Yes

1. Bare DVD media is highly susceptible to physical damage and contamination.
2. The DVD-R and CD-R recording process does not include write verify
3. Available with software API WORM capabilities.
4. Current WORM RAID solutions demonstrate very slow transfer rates
5. DVD-R, DVD-RW, DVD+R, DVD+RW do not have addressable sectors. Only DVD-RAM

Archival storage total cost of ownership (TCO)

Table 3 examines the TCO regarding archive solutions. Some basic questions need to be addressed before looking at the cost of ownership on an archive solution.

1. How often will archive data need to be accessed?
2. How long will the data need to be retained?
3. Does the data need to be stored in a non-alterable state?
4. What other problems are you trying to solve with the implementation of this system?

When selecting an archive solution, keep in mind the data is fixed and is required to be accessible for a number of years. The data set will grow at an increasingly faster rate over time and must be managed, indexed, searched and if needed retrieved and presented from its original state. A plan will need to be in place to expand the archive solution as your data expands, however, you will need to have access to the data during the expansion period.

Hardware costs

Hardware costs include the storage system itself, the storage media (if separate) and any storage server needs you may have. It's important to look at usable storage capacity when doing a price/performance analysis. This price performance analysis should cover all costs over the archive period based on your requirements, including power consumption, cooling costs, physical space, and annual operating and administration costs. Ideally it is always best if existing hardware infrastructure can be used.

Software costs

On the software side you will also want to consider application and data lifecycle management software costs and the impact they will have on the current administration and operating costs. Include annual maintenance and associated software upgrade costs.

Will data need to be migrated or converted?

Often overlooked is the cost of data migration. Consider the cost of system hardware upgrades over the life of the data to be managed, migration administration, and associated cost for audit trail management. Data migration triggers include:

- Upgrades to system hardware
- Migration to newer/higher capacity media
- Useful life of hard drive technology 3-5 years
- Refreshing existing media (tape)
- Rebuilding existing media due to system failure

Additional costs to consider depending on media selection and regulation requirements:

- Remote site mirroring
- Data disaster recovery and backup
- Removable media vaulting

Table 3

8.5 TB Archive Managed Costs over 5 Years of Operation.

Product Type	Est. HW List Price	Managed Capacity in GB	Est. Media Price	Storage Mgt SW Price	HW and SW Maint.	Power and Cooling	Total Space Cost	Est. Total Server	Grand Total	Est. Price per GB
UDO - Write Once	$84,677	13,140	$17,082	$16,995	$30,487	$1,325	$10,200	$5,000	$178,905	$14
UDO - Write Once	$61,687	7,140	$16,660	$11,995	$39,914	$1,325	$10,200	$5,000	$153,921	$22
MO Rewritable	$186,560	7,972	$93,741	$35,995	$67,159	$2,845	$12,825	$5,000	$412,098	$52
Disk system	$124,000	9,600	$0	$202,000	$135,560	$11,773	$9,000	n/a	$491,933	$51
Disk system	$186,000	8,600	$0	$303,000	$203,340	$11,773	$9,000	n/a	$721,713	$84
DVD-R	$32,690	8,225	$13,913	$9,995	$12,261	$711	$10,200	$5,000	$92,995	$11
DVD-RAM	$32,690	8,225	$19,600	$14,995	$12,261	$711	$10,200	$5,000	$103,682	$13
LTO Tape	$44,595	13,000	$12,766	$8,500	$45,889	$9,905	$6,418	$5,000	$146,074	$11

Does not assume file conversion costs over time.

Considerations to reduce archive management costs:

- Select a strong data lifecycle management solution to automate the archive and data retention process
- Archive data stored in a file format that can outlive the operating system as it gets updated over the life of the archive
- Select a durable medium-designed to be relatively impervious to environmental contaminants and protected by a robust cartridge
- Non-rewritable, non-erasable media-offering protection of electronic records at the media and storage management component level
- Removeability-offering the ability to off-line inactive archives or near-line active access to electronic records. Also facilitates in creating and retaining disaster copies of electronic records.
- Media longevity-the ability to archive data for longest possible shelf life of any digital media (Ideally the media shelf life spec should be twice as long as the data need be available)
- Backward compatibility-a history of successfully providing the ability to read older media generations with newer write/read generations. This reduces the number of data migrations, thus reducing the risk of data loss.

It is clear there is a place for disk, tape and optical within the data lifecycle management model. The key is to give users the ability to direct data, based on its current value to the company, and to the most relevant storage medium so that corporations can extract the maximum value from the data at the lowest possible cost of management. Archives must be protected and the user able to demonstrate that the corporation meets or exceeds any

regulatory requirements and can gain access to data in a timely manner. Pegasus storage management solutions help to retain and manage critical archive data and other digital assets while providing a means to meet strict regulatory compliance requirements in a number of fields such as, Financial (SEC, Sarbanes-Oxley), Medical (HIPAA, DICOM, PACS), Government (DOD), Legal Admissibility (ISO-15801 and 18509) and many others.