

1 Introduction

Archiving of information is one fundamental Information Technology operating process that has been undergoing a renaissance. In the evolving history of storing information, archiving has been an integral part of operations in the mainframe world but has been not effectively exploited for its value in other types of operations. There have been limited tools both in software systems and in hardware storage systems to provide the real benefits that could be achieved with archiving. That situation is changing with new and maturing software and revolutionary storage systems.

The requirements of managing information now include the legal ramifications in dealing with regulated information. There are over 10,000 laws worldwide according to industry analysts that deal with some requirement for storing information. From the standpoint of IT operations this is an extra burden that goes beyond just the processing and storing of information. The requirements to meet regulations regarding the management and storing of information are highly visible within a company by executives and legal counsel and can have both career and potential legal implications for IT management.

The regulatory compliance issues regarding information are even more significant for information that is being archived. Many of the laws are specifically written with the anticipation that information will be stored for long periods of time. The length of time that the information must be stored, control of access to the information, the means of identifying the different copies of

information, and the ability to produce information on demand all have legal requirements.

This book will detail what constitutes archiving, explaining the elements and functions that distinguishes it from the more traditional process of data protection using backup. The economic implications of implementing an archiving system and its impact for the long-term are a very important part of an IT strategy and explored in detail in this book. The other major aspect of archiving which is part of the data preservation is meeting the regulatory compliance and business governance demands for archiving of information. The variety of regulations and their effect are a major part of the second half of this book.

The organization of the material is first to explain what is archiving of information in general terms with some of the considerations involved. The different options for archiving are explored with the new technologies highlighted. The economic impacts of archiving are covered next with examples of what is to be expected.

Next in the material is the explanation about the impacts of meeting regulatory compliance for archiving. The intent of the regulations and the real impact on storing the information is needed as a basis of understanding. Some of the more well-known regulations are discussed for their impact on an archiving system.

eDiscovery and what it means regarding potential litigation which in the United States is tied to the Federal Rules of Civil Procedure is discussed as a function closely aligned with regulatory

compliance. The demands on an archiving system to meet the eDiscovery and FRCP requirements are examined in detail.

What are the concerns for very long-term archiving and the efforts to address it are covered in the section about the future of archiving. Efforts in progress around standards and other advancements are explained and put into context of an evolution from solutions required today.

Finally, a summary of the economics and compliance issues regarding archiving of information puts perspective on the issues and helps in guiding the integration of archiving into an overall strategy for storing information.

What is Information Archiving?

Most Information Technology professionals already have an opinion of what information archiving (or data archiving) is. Many of the professionals may not be quite accurate or not have a clear understanding of what really constitutes information archiving. The reason for that may be rooted in the changes over time in the archiving technology (both software and hardware) and in the requirements for archiving. Certainly the mainframe operational environment matured with an archiving model using integrated system software and a model of hierarchical storage and generations of data. In the open systems world, the capabilities have not matured at the same level or uniformly between the different systems.

Historically, most IT operations have done what they considered archiving by moving data to magnetic tape with backup software while making two or three (or more) copies. One copy of the data would be put on a shelf or in a library in case there was a need to retrieve some of the archived data, another tape was sent to an offsite location as a protection from some potential disaster and a possible third tape was sent to another of the company's operations.

It was never expected that the data would have to be read from any of the tapes. But when there was a need, the complications started. Potentially there could be many issues with retrieving the information, beyond just locating the tapes themselves:

- The tapes may not be readable. Tapes deteriorate over time so there is a recommended lifespan of

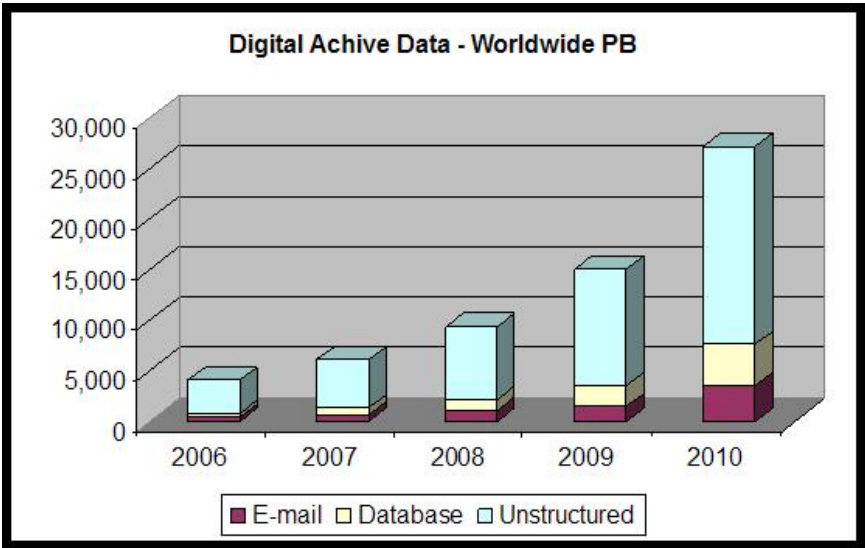
reading and rewriting the data. Additionally, tapes have a limit in the number of passes across the read/write heads before the tape stretching or abrasion renders them unreliable.

- The tape drive mechanism may not be available. The drive that a tape was written on may have become obsolete and replaced with newer generations that cannot read that generation of tape. The obsolete drive may not be available from any source. Drive obsolescence is a regular event that occurs on a 5 to 6 year interval which necessitates a forward migration of data as a planned activity. If the tapes on the shelves or at the disaster recovery facility were not migrated, there may be no way to retrieve the data.
- The software that was used to write the data to tape may no longer be installed or supported and there may be no migration path to new software. Many times a standard backup application was used to move the data. From the time the data was written until it was needed again, the application may have changed or been discontinued such that no backup application exists to retrieve the data any longer.
- The application software that created the data may have changed or have been discontinued such that the data, while it may be retrieved, may not be in a form that is usable.

The focus on information archiving has increased greatly over the last few years and is continuing to be a major topic within Information Technology. Why has this interest increased so greatly? There are many different reasons attributed but the reasons come down to two fundamental areas that are driving many other changes in IT.

The first major change that has brought renewed interest in archiving is the dramatic increase in capacity demand that has continued. The statistics are quoted often: an annual, compounding growth of the amount of data to be stored of 80 to 100%. The number really is staggering and the amount of storage needed represents serious impacts:

- Capital expenditures for new capacity continue to increase even considering the fact that primary storage is getting less expensive.
- More physical space is required for the storage systems – despite the expected 18 month doubling in capacity of disk storage technology.
- The additional storage will require more power and more cooling. Some mitigation is being attempted with spinning down disks that are not currently transferring data but even that has a limitation in the amount of benefit.
- The operational expense of administering more storage capacity continues to increase. Rarely does a company provide for the staffing necessary to accommodate the increase in capacity at the same ratio so either required functions don't get accomplished as effectively in administration or other changes have to be made. Sometimes new software tools can be used to enable staff to manage more effectively but these are one time improvements that do not track linearly with the capacity increases.
- Migrating data to new primary storage when the old systems reach the end of their useful life.



Source: Enterprise Strategy Group

Figure 1: Archive Capacity Growth

A driving factor for capacity increase besides the normal increase in business is the fact that 99% of all documents are in digital form now. Another well-known change is in the amount of emails that are received. Currently, there are over 50 billion emails sent per day. The type of data that is increasing so dramatically is not business data contained in databases but is the unstructured data representing documents and individual files.

The other major change that has caused new or renewed interest in information archiving is the advent of more regulations regarding information. These regulations present significant problems to Information Technology to comply with the laws and avoid any potential penalties. The regulations regarding

information are usually specific to business operations, human resources, and vertical industries such as health care. They apply to all information but have a special importance for archived data due to the long-term preservation implied for archived data. Using backup software does not lend itself to the discovery actions required against the information and does not meet the long-term compliance demands.

Meeting regulatory compliance demands regarding storing of information has increased in importance over time due to the penalties that have been assessed to companies for non-compliance. The penalties have been financial with ranges between the inconsequential to millions of dollars and have been threats for civil penalties.

Archiving is important again for those reasons. To do the archiving, some new approaches are being employed. These new approaches include both hardware archiving systems and archiving software. The new approaches for hardware include removable disk systems as the target for long-term archive and integrated archiving appliances to provide simplicity. In the archiving software area, movement of data to an archive is integrated into Enterprise Content Management software, file archiving and migration software, email archiving software, and with some industry specific solutions.

The economic benefits of archiving data – moving data off primary storage to where it no longer requires backing it up and frees up that capacity makes it a strategic technology for IT. Adding the ability to manage the information in the archive according to specific regulations is equally important – it eliminates the potential for financial penalties and keeps executives from

prosecution. For IT staff, protecting information and meeting regulations is also career protection.