Reasons to archive

Summary of Linkedin Forum Discussion

Author: David Barrett-Hague
Version: 1.0
Creation Date: 21 July 2011
Revision Date: 21 July 2011
Status: Final
Distribution: Unclassified

Revision History

Date	Who	Version	Comment
29th June '11	DBH	0.1	Created.
21st July '11	DBH	1.0	Finalised for publication

Contents

Executive Summary

Data archiving is something that all organisations know they need to address but often tackling the growing mountain of data is deferred until a compelling event forces the organisation to take it seriously. Whether it is a legal disclosure (eDiscovery) or a business continuity matter unfortunately the later it is done the more it costs.

Implementing a data archive brings a number of benefits to the organisation and this paper analyses the responses by a group of Archiving Professionals on LinkedIn who contributed to a forum debate (see http://www.linkedin.com/groups?gid=125782) which addressed the many reasons why companies archive digital data. The question poised was:

> "Reasons to Archive: Incontrollable backups, poor performance, ... share your reasons"

The main reasons offered were to lower costs, improve performance of existing IT infrastructure and reduce legal risks. It was generally agreed that the cost saving was a key driver in the implementation of an archive. Some argued that this should not be the main reason, as there were significant other benefits, however it was clear that a strong return on investment (ROI) was a key part of the decision to implement an archiving strategy.

A number of commentators clarified the difference between backup and archiving, one put it particularly clearly stating:

> "Archiving is about the long-term retention of information for legal, regulatory and compliance demands and back-up is for short-term disaster recovery."

Archiving relates to the preservation of corporate records which are individual pieces of information which provide a historical account of the actions of an organisation. Each has a value, often unknown, which may be an asset or liability for a company. Hence, based on the relevant regulations some records should be destroyed rather than retained preventing it causing future legal risks and also freeing up space.
An archiving strategy is part of an overall data management system and based on pre-defined criteria data is moved off expensive primary storage systems to an archive repository. This frees up the primary storage capacity and allows more appropriate management policies for the archive data. Another commentator concisely brought this together, stating:

> "Ultimately it's not an archiving discussion but one of data and information management and having a long term strategy"

The key benefit of an organisation implementing an archiving strategy is that they are starting to actively manage their data, taking seriously the responsibility of managing this corporate asset.

Like most business processes archiving has many stages, whatever the initial reason to implement an archiving strategy the repository often develops into a corporate asset

that can be mined as a research tool to assist with future business endeavours and promote organisational learning. If implemented in a sustainable way archiving can not only reduce cost but improve business efficiency and improve the competitiveness of the organisation.

Introduction

This paper summaries the response to a question poised on the Linkedin Achieving Professionals Group (see http://www.linkedin.com/groups?gid=125782) forum addressed the many reasons why companies archive digital data. This paper provides a summary of the responses from a wide range of industrial experts to the question:

> "Reasons to Archive: Incontrollable backups, poor performance, ... share your reasons"

Reasons to Archive

The main reasons offered were to lower costs, improve performance of existing IT infrastructure and reduce legal risks. The full list of reasons offered was as follows, they have been group into these three many headings:

- Cost saving
  - Reducing length of time taken to do data backups
  - Improving performance by unclogging primary systems
  - Reduced insurance premiums
  - Storage reduction, use of de-duplication technology
  - Archiving prior to system upgrade reduces downtime of system
    - giving organisations increased flexibility and agility when it comes to choosing their next mail platform.

- Improve performance of existing IT Infrastructure:
  - Performance of primary application eg Oracle, SAP, Exchange is improved by removing old data
  - Part of DR/Business Continuity planning
  - Gain control over the unstructured data allowing future [data] mining
  - Gain in productivity (reduce time to find archived information)
  - Reducing *.pst sprawl

- Meet Legal and Regulatory Compliance requirements
  - Legal discovery or eDiscovery ". Ever since Enron, however, the legal advantages of being able to quickly search an online repository of archived data has steadily overtaken all other reasons for archiving. "
  - Retention management, including the deletion of information

It was generally agreed that the cost saving was a key aspect to the implementation of an archive. Some argued that this should not be the main reason, as there were significant other benefits, however it was clear that a strong return on investment (ROI) was a key part of the decision to implement an archiving strategy.

The other two main reasons could also be quantified in financial terms and argued to be a cost saving however they represent clear benefits to the organisation beyond the crude cost analysis.

By moving aging data out of primary systems it not only frees up space but also improves the performance of the system. This little used data clogs up business applications like databases and email servers and results in frustration from the users as the systems slow down over time. Proactively archiving this data restores this performance and has other benefits. For example by archiving emails in Outlook users can be given large inboxs and hence don't need to create local .pst files (separate personal archives) thus eliminating a major headache for IT Administrators.

Another important performance improvement comes from the separation of an organisations active and inactive data. By archiving the inactive data you allow the different types of data to be treated differently, more appropriately. You active data continues with your existing disaster recovery backup strategy, of course this is now much quicker. Whilst your archive data can be handle by a separate disaster recovery strategy. For example if data is only moved to the archive once per week, then any disaster recovery strategy only needs to run once per week as no data will change in the intervening period.

Meeting legal and regulatory compliance requirements was also seen as a key reason to archive information. It was stressed by a number of commentators that this includes the need to destroy information at the appropriate time. The term Retention Management is often used to describe this process of identifying records for destruction usually after a period of time has elapsed. However, many organisation are reluctant to delete data, as one respondent colourfully draw the analogy of his father's garage, one man's junk is another man's archive.

One specific comment was made that related to the broadcast industry due to the sheer capacity required to archive the massive back catalogues of film and videos. Digital Asset Management (DAMs) systems are often used to manage this data and provide access to the movies. Each digital asset (film footage) may not be accessed for many years and then due to a sudden event it may become very valuable.

General advice

In addition to the reason why to archive there were numerous pieces of advice as to what should and should not be part of an archive solutions. The issue of security was a key area of consideration as an archive repository is a pooling of corporate information and knowledge which will include confidential information. In this regard encryption was mentioned as a solution however, like de-duplication, when considering the longevity of an archive you also need to consider the separate long term storage of the encryption keys, or de-dup indices, as without this valuable corporate data becomes worthless random digits.

One often over looked aspect when developing an archive strategy is that due to the long term nature the information needs to survive beyond the economic life of the physical hardware it is stored on. Therefore future technology and data migrations should be considered in the design. This also relates to the storage management software including any encryption and de-duplication features. One only has to look at Microsoft's Windows operating system and the issues caused going from one version

to the next to imagine the potential problems after 20 years. Hence, it is important not to get locked into a proprietary system.
Experience has shown that archiving is addictive. That is once an organisation starts it realises the huge benefits and widens the deployment to other areas of the business what were not part of the initial scope. As such an archive must be easy to scale with the ability to partition different areas to provide for different departmental needs. For example the legal department may need a retention management policy whilst the R&D just wants long term preservation of their research information.

Archive Technologies

Little was discussed about possible technologies used as part of the archive, this is probably due to the fact the question poised was why and not how. The suitability of the various technologies depends on the particular requirements of the company. Most archiving management solutions automatically move data into a secondary storage, based on pre-defined policies, and provide a transparent link to the end user. This removes the need for intervention from the IT Administrator to archive or retrieve information.

In terms of physical storage medium there are three options, namely magnetic disk, magnetic tape and optical disc. Each has there pros and cons which again depend on the requirements of the company implementing the archive specifically around the longevity of the media, data access profile and long term cost of ownership. A more recent consideration is the energy consumption which is particular important considering the length of time this data remains inactive and un-accessed.

Conclusion

A strong theme from the forum discussion was that the key benefit of an organisation implementing an archiving strategy is that they are starting to actively manage their data, taking seriously the responsibility of managing this corporate asset. Like most business processes archiving has many stages, whatever the initial reason to implement an archiving strategy the repository often develops into a corporate asset that can be mined as a research tool to assist with future business endeavours and promote organisational learning. If implemented in a sustainable way archiving can not only reduce cost but improve business efficiency and improve the competitiveness of the organisation.

Authors note:

The information above was taken from the comments made by members of the Archiving Professional LinkedIn group as part of a forum discussion and has been edited and expanded to make the paper readable by a wider audience. Whilst the author was one of the respondents for privacy reason the other respondents have not been identified. However, the reader may wish to read the forum comments in full and these can be found at:

http://www.linkedin.com/groups?gid=125782