



Disk Based Repositories

An alternative to SQL based archiving



Overview

The purpose of this whitepaper is to define and explain the architecture and benefits of Archive One data stores, known as repositories. This paper details how establishing criteria-sensitive archive repositories on appropriate storage media, can assist organizations to achieve reliable and fully functional email archiving while reducing the high cost of ownership associated with alternative archive technologies typically requiring interaction with SQL.



Disk Based Repositories: Definition

Archive One can store the data extracted from Exchange Servers in multiple virtual units defined as disk based Repositories. The structure of a Repository was designed to optimize requirements of fast search of the index, rapid retrieval of data and minimal back-up while supporting full security and disaster recovery initiatives.

Each repository consists of email content held in multiple data blocks of 10 Mb each which are compressed and encrypted, and stored with their associated index. This Repository architecture can support more rapid data retrieval and simplified backup / restore processes than alternative technologies, where thousands or even millions of data files and large indexes requiring high levels of resourcing have been seen to negatively affect email performance, prove impractical in terms of backup strategies and high-risk for Disaster Recovery.

Archive One does not limit the number of repositories per archiving server, therefore it is simple and indeed desirable at a business or administrative level to define a separate Repository for data based on an archiving criterion or a combination of criteria. Examples include establishing a Repository per department, year or both.

Repositories offer the administrator the flexibility to set a retention period or time limit on data availability. Default retention and deletion periods can be set at Repository level, within which individual or groups of messages and attachments can have unique retention and deletion options. Retention periods can be changed at any stage by the administrator.

Mailbox users are associated with a Repository via the Archive One Administrative console. Users can be moved between repositories at any stage, and can be retrieving from multiple repositories simultaneously. The storage location of repositories can be on local drives, network shares and large storage devices such as a NAS or SAN.

Archive Data Categorization: ILM for email archiving

Using multiple repositories allows the placement of archived data to the most appropriate storage based on performance requirements and appropriate data ownership costs – an approach mirrored in the storage industry by its adoption of “Information Lifecycle Management” strategies.

An example of how a company (Company A, see diagram on next page) could use multiple repositories follows-

- Define repositories for each major business department within a time period (e.g. Finance 2009, Finance 2008, Finance 2007, Human Resources 2009, Human Resources 2009, etc)
- Allocate user mailboxes to the current relevant Repository
- Set an archiving policy to archive data between date ranges

The location of each of the repositories can be based on a combination of performance and cost of ownership considerations, for example:-

Old data – over 3 years: All of the departmental data archived for years 2005 and 2004 could be located on low cost / low performance network storage such as IDE Disk. The reason for this is this archived data is unlikely to require regular access by the users, so the slow performance of this storage is not a major factor. However the placement of this type of archive data onto low running cost storage is a huge reduction in cost of ownership compared to alternative models.



Medium age data – 2-3 years old: All of the data archived for the year 2007-6 could be located on a medium performance file and print production server. Greater storage performance is available than using the low cost storage used by the 2000 - 2001 archived data, but cost is justified as users are more likely to want to access the data stored.

Recent data: All of the year 2008 and 2009 archived data could be hosted on a High Performance NAS or SAN. Users will actively wish to retrieve data from this section of the archive and the provision of high performance storage would be appropriate.

In subsequent years the departmental and annual categorized archived data can be very simply moved onto progressively lower cost/performance storage. Archive One provides a simple wizard to assist with this activity.

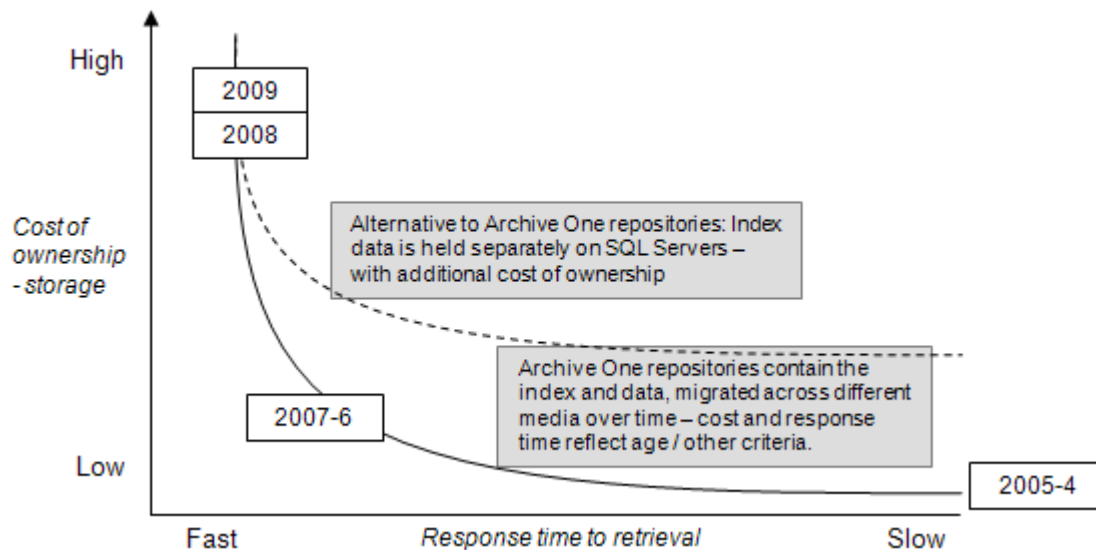
Example for company A: media migration repository model for Finance Department

A simple approach to use of archive repositories is illustrated here. Repositories are established by year and stored on media appropriate to the probable access requirements of users and the relative storage costs of those media.

For example, Year 2005-2004 data is on old, slow, and low-cost media. 2007 - 2006 is on medium cost and response media, while 2007 and the new 2009 repositories are on fast, high performance media which is more expensive but supports the users' access requirements.

At the end of 2009, a 2010 repository will be established and previous years' repositories will be migrated back by one profile of media – so 2008 will go to medium response storage and 2006 will be moved to slower, cheaper media.

Note that Archive One Repositories contain the associated index data; alternative SQL based models have directly attributed additional cost of ownership with separate databases for indexes.



Backup Strategies

The construction of an Archive One Repository, comprising 10 MB packages of data with their relevant index, allows for the archive backup process to be very simple and low cost. The archiving service does not need to be suspended for a



Repository back-up to occur. No special agents or technologies are required and even NT backup provides full support to Archive One Repositories.

Furthermore, if a company has deployed Archive One Repositories with date range criteria, only the present year's repositories will need to be actively backed up on a daily basis, in significant contrast to alternative approaches. In the example for Company A, the data stored in departmental repositories for years

2005 – 2004 will never change and as such does not require the daily backups, but only needs 2-3 disaster recovery copies in case of server failure.

One of the greatest costs of data ownership lies in the provision of daily backups

Therefore, the above strategy can have a massive impact on the reduction of the cost of ownership of the legacy archived data with no trade-off in maintaining operational user access to that data.

Other archiving technology models do not provide this flexible approach for data placement, but require that SQL servers are provided to hold the index, irrespective of the access demands and performance profile associated with the archived data. In addition to SQL hardware and license requirements, its use for archive indexes requires the archiving service to be stopped for a full back-up of data and indexes, as well as specific database administration support and a full resilience / DR strategy.

Archive Data Replication

Windows provides direct support for Distributed File System Replication (DFS) and File Replication Service (FRS). This provides the ability to select network directories and perform data replication to alternative servers and sites.

Disk based repositories do not require implementation of sophisticated and expensive replication technology but provide organizations with an immediate facility (via DFS and FRS) to include email archiving in their Disaster Recovery strategy.

Under most alternative archiving strategies, SQL database is chosen as the storage component for the archive index and on occasion the email data too, though a complex design is required to enable this replication facility to work.

Using SQL in turn can require further significant costs associated with secondary SQL servers, backup agents and maintenance. SQL replication is very complex and expensive when considered against DFS Replication of disk based repositories.

Repository Retirement

Removal of a Repository from the organization's operational environment is a common requirement and quite simple to achieve. The repository definition can be deleted from within the Archive One Administrative application and the relevant directory structure deleted from disk. An immediate regain of disk space is then made.

It should be noted that even though the Repository is retired, simply restoring the repository back to disk (which does not even need be the original location) and then re-defining the Repository in the Archive One Administrative application will re-establish the archived data for access.



About C2C

We hope that this paper has helped you to plan your way forward. C2C offers automated data archiving and management for email, files and SharePoint content. With over 15 years experience delivering solutions for capacity, e-policy enforcement, compliance and eDiscovery, C2C optimizes performance, reduces storage management costs and minimizes risks associated with email - helping you to control your data before it controls you.

The Company, a Microsoft Gold Certified Partner, supports organizations in the government, manufacturing, finance, education and healthcare industries, including Fortune 1000 companies. Established in 1992, C2C is a privately held company with offices in Springfield and Westborough, Mass. and Reading in the UK.

Copyright

All trademarks, trade names, service marks and logos referenced herein belong to their respective companies.

Disclaimer of liability: While every precaution has been taken in the preparation of this document, C2C assumes no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

Where can I find out more?

For more information and free evaluation software, visit www.c2c.com or email info@c2c.com.

C2C Systems, Inc.
1 Federal Street
Bldg. 101-R
Springfield
MA 01105-1199
USA

T: +1 413-739-8575
F: +1 413-739-4980

C2C Systems Ltd
6 Richfield Place, Richfield Ave
Reading
Berkshire
RG1 8EQ
UK

T: +44 (0) 118 951 1211
F: +44 (0) 118 951 1111