

Archiving and Capacity Management Solutions For Microsoft® Exchange



**C2C Systems
September '01**

r.v. ar· chived, ar· chiv· ing, ar· chives

To place or store in an archive.

Computer Science. To copy or compress (a file) into an archive.

(source www.dictionary.com)

Table Of Contents

Introduction.....	3
Background	3
What is driving the need for Archiving and Capacity Management?	4
▪ System constraints and server roles	4
▪ Information management	4
▪ Regulatory requirements	4
What is meant by 'archiving' within Exchange?	5
Archiving Solutions.....	6
How does Microsoft's strategy affect the solution?.....	6
Front-end and Near-line servers.....	7
Exchange Indexing and Knowledge Management systems	8
So what is an Archiving and Capacity Management solution?	8
Corporate growth and Futures.....	9
The C2C Approach.....	9
Where can I find out more?	10
Archive One.....	11
Technical Overview	11
Architecture	12
Client.....	13
Access to Archive data.....	13
Ghost Messages	13
Server.....	13
Advanced Knowledge Management.....	14
Compressing Data.....	14
Security	14
Content Control	15
Virus Control.....	15
Language Variants	15

Introduction

For organizations that rely on Microsoft Exchange as a critical business resource, the concept of email archiving is emerging as one way to preserve Exchange's dual role as a time critical message delivery platform and the de facto business intelligence repository. This white paper -

- Explores and clarifies the business drivers for archiving in the MS-Exchange environment.
- Outlines the continuum of approaches to archiving – from simple deletion to complex and expensive off-line storage platforms.
- Comments upon Microsoft's mid and long term strategy for MS-Exchange as an information management system.
- Suggests a way to address the most critical aspects of today's archiving requirements, without limiting future flexibility to follow Microsoft's strategic directives.

Background

Archiving has become a generic term for a number of solutions attempting to meet a variety of needs. The solutions vary from stripping attachments from e-mails to off-line file-store, through complex backup and retrieval systems involving CD or DVD jukeboxes.

The primary problem administrators face remains consistent. The messaging system has unwittingly become the common repository of all operational data. While this is great news for end-users using the production email system as a personal file store, it is not so good to the well being of a system designed for rapid high throughput and fast backup/restore schedules.

The administrator is faced with the dilemma of how to provide high throughput, high availability servers and give end-users access to their stored data without the need for huge retraining exercises.

Although there is a common primary requirement to offload data from the central servers, the secondary needs of how then to manage, and most importantly make this information freely available, are varied. It is this variation that creates a number of different solutions, which vary greatly in terms of cost, complexity, ease of use and manageability.

This paper examines the strengths and weaknesses of these solutions, the Microsoft strategy for Exchange in relation to archiving and explores an idea for a flexible solution that meets a range of criteria.

What is driving the need for Archiving and Capacity Management?

The need for Archiving is the proliferation of data being stored in the Exchange Server. Industry figures tell us that message volumes are increasing at 35%-50% and more per annum, while individual message sizes are increasing around 25% per annum. This gives us a year on year growth of close to 70%!

Every organization has its own definition of Archiving but all these definitions are generally derived from similar business needs, which can be summarized as:

- **System constraints and server roles**

The need to ensure that we have a set of Exchange servers that are kept to a manageable size. To some, the cost of upgrading servers, ensuring backup/restore times are within manageable windows and administering rapid growth systems is high. With Exchange 2000 (E2K) the concept of server roles is much stronger. You can easily define different roles for servers, such as a front-end server dedicated to the fast, efficient delivery of email and near-line server connected fully to the Exchange system that is dedicated to the longer-term retention of email data or a public folder server.

- **Information management**

This is the need to be able to find relevant information in a structured fashion and can be referred to as Knowledge Management (KM). This can be further subdivided as

a) User KM, being some form of simple retrieval of data for individual use.

This can be simple or complex technology, but the expectation is that the data is for the individual user only, so constraints on storage and searching are not rigorous.

b) Organizational KM. Still creating a store for each user but we are also allowing organizational wide access to this information with the aim of creating a true Organizational Knowledge base. This is far more complex and expensive than option a) and is usually part of a company committed application requiring massive investment and user training.

- **Regulatory requirements**

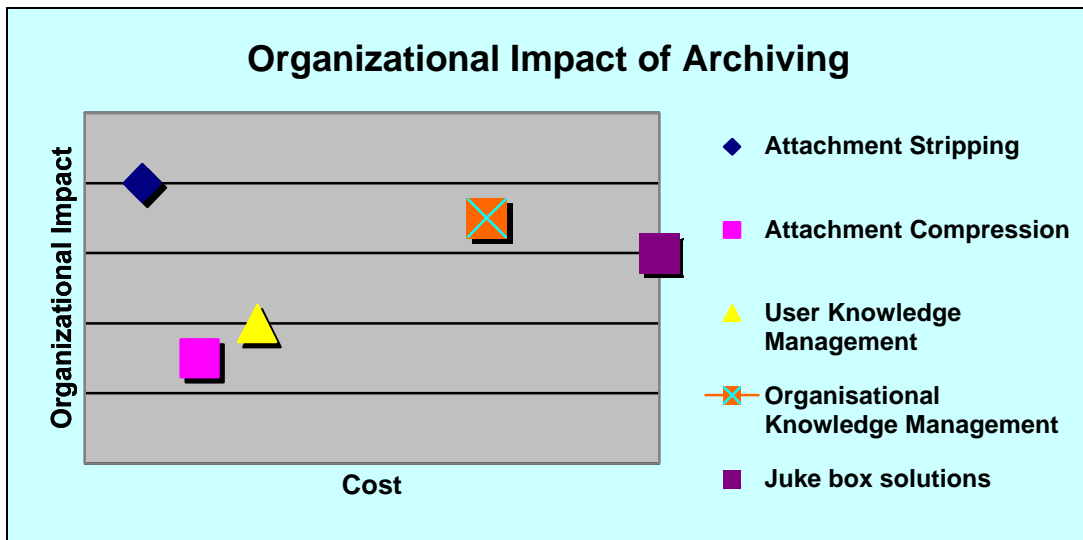
These are forcing many organizations to adhere to regulations requiring them to keep all or defined parts of their email transactions for a period of time.

Typical Regulatory requirements force organizations to:

- Keep copies of all emails
- Keep copies of all email transactions with third parties
- Maintain copies of the electronic calendars of key members of staff

Of course each of these business requirements and their associated solutions has a different cost and impact to an organization. A company should examine its needs and decide which of the methods mentioned (or combination of methods) would meet their needs and solve their problems.

The simple message is never underestimate the complexity of retraining if the solution does not fit directly within an existing framework.



What is meant by 'archiving' within Exchange?

So, if the overall need is to reduce the dependency on the Central Server resource loading, what are the individual business factors that help us define customers' requirements?

"Archiving Solutions" can assist with

- Reducing disk storage on the Exchange Server
- Reducing the time taken to backup/restore the Exchange Server
- Reducing the time taken to backup/restore individual mailbox or folder data
- Moving the data to another storage area to reduce bottlenecks and loadings on the primary messaging server
- Copying or moving data elsewhere to meet statutory or regulatory requirements
- Creating Knowledge Management systems where the corporate email system resource is searchable and views of this data can be personalized

One key point of examining archive solutions is that it will drive you toward defining roles for Exchange Servers. This can be for long-term retention of email or providing fast efficient routing of email, but the two can rarely co-exist in a single server.

Archiving Solutions

There are a variety of solutions to meet these requirements that have advantages and disadvantages. The solutions include:

- Attachment stripping
 - Simple to understand
 - Single instance storage is broken
 - Forwarding of mail becomes complex and often needs a central attachment server that is a less efficient file store than Exchange itself
- Attachment compression
 - Simple to understand
 - Can manage around single instance storage
 - If handled correctly compression and decompression can be a seamless way to reduce the size of the information store and also reduce transmission times making Exchange 'faster'
 - Does not enhance the knowledge environment
- Moving data to a near-line server
 - Will reduce the load on the front-end server
 - May enhance the knowledge environment
- Moving data to a near-line server and integrating with an index server
 - Will reduce the load on the front-end server
 - Should significantly enhance the knowledge environment
- Moving data to a file store and integrating an index server
 - Requires significant training and imposes rules on the way knowledge workers operate
 - Should significantly enhance the knowledge environment
- Moving data to a juke-box and providing restore capabilities
 - Cost and complexity of hardware and associated resource costs

How does Microsoft's strategy affect the solution?

Before trying to match the requirements, one should understand Microsoft's positioning of Exchange as an Information Store. Current Microsoft promotions show Exchange as the repository of Unstructured Data and SQL Server as repository of Structured Data.

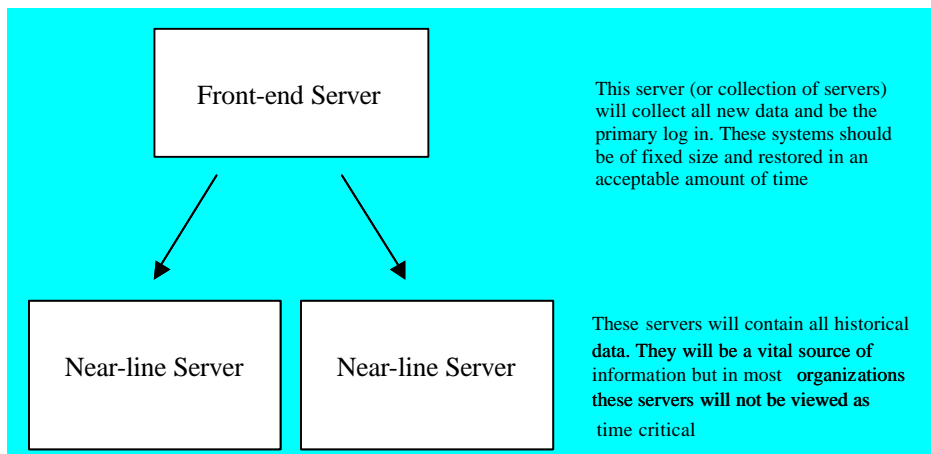
So, if Microsoft is developing Exchange to be the repository of unstructured data, then future developments of Exchange are likely to enhance the scalability of the message store and abilities to search within it. One has to ask if the solution being chosen fits within the Microsoft strategy. Failure to do so could result in premature obsolescence. Worse, you could find yourself with a useless system– unable to migrate or upgrade.

The most obvious statement to make after looking at the future strategy of Microsoft is that information should be left in the Exchange system for as long as is possible. ‘Archiving’ systems that involve moving all or part of the Exchange data to other file stores (such as NTFS or FTP Servers) will generally prove far less efficient. There is also a large question over the long-term viability of such methods.

Users of archiving solutions not only need to qualify their requirements but also must be aware of Microsoft’s strategy before making a decision.

Front-end and Near-line servers

Prior to Exchange 2000 (E2K), Microsoft Exchange was limited in its ability to scale. For example, you could only have a single database on each server to hold the public folders. With E2K a more flexible storage system exists where Exchange can be split to provide multiple message stores, which gives it more scalability and configurability.



E2K offers a more resilient and scalable system but requires other tools to identify messages by criteria and then move or archive these messages. It is also a small logical jump to realize that these back-end servers (unlike the front-end servers) are

prime candidates for indexing and turning into knowledge management repositories. It is this point that leads us to believe that Archiving is the location and moving of the data and it is Knowledge Management systems such as Index server that will allow the searching and recall of the data in the most efficient manner.

There are also other important points to make here. The first is that the back-end servers need not be an Exchange server. They could just as easily be a SQL Server or a simple directory structure. However, it would seem that as Microsoft is aiming for the Exchange Server to be '*the*' store for unstructured data in the enterprise, then it would be most logical to conform to their architecture and use an Exchange Server as the Near-line Server. Further, there is likely to be another group of servers behind the back-end servers and these are the off-line servers. These are likely to be a physical backup system (CD Juke Box) and retrieval catalogue so the data can be brought back to the user.

Exchange Indexing and Knowledge Management systems

Many organizations are very excited about the indexing abilities of E2K, and as a core component of Windows2000, indexing seems very exciting. However, the performance implications on systems such as Exchange are high. If you are keen on keeping the Exchange Server as a performance engine for message throughput and robustness, then this is an excellent reason for using near-line servers, and indexing on them.

While Exchange is one of the most important sources of business information, it is not the source. Many tools and products exist to merge together all information that is relevant to users. These systems or applications are Knowledge Management systems. Applications such as Microsoft SharePoint Portal server are designed to 'crawl' through information in Exchange, merge this with information from other sources (web sites and file stores for example) and then present this to users in a searchable fashion. Clearly integration with systems such as this would make the Exchange system an even more flexible business information repository.

So what is an Archiving and Capacity Management solution?

We would promote that the best and most efficient archiving solution is a methodology for removing data in a structured (rule-based) and reliable fashion.

Retrieval of the data is not necessarily a part of the removal method, but can be a separate product integrated into the solution. Retrieval is probably best thought of as Knowledge Management, but not essentially so.

So why is retrieval separate? Because to some administrators retrieval is the all-important part of archiving. To others the storage of the data is key, and retrieval capabilities less important. One scenario exemplifies this is the FDA regulation requiring executives' calendars to be kept for a period of seven years. The archive solution can copy off this data daily, weekly or as needed, but the need to retrieve it will happen rarely, so why pay for a retrieval system when a manual system will work adequately.

Corporate growth and Futures

Two key elements of archiving that must not be ignored are corporate growth and future proofing of any solution.

Archiving is a complex task and implementing a solution that meets the company demands for today and even two years hence is unlikely to occur. You need to take care that a chosen solution does not become a lock-in should (or more likely when) corporate needs change.

One area you must consider is that of data availability in say seven years from now. Courts are demanding that companies search e-mails to prove innocence (or guilt) and as such you have the responsibility to your organization to ensure data can be recovered. You need to ensure that whatever archiving solution is used, the data will remain accessible in many years. This either means ensuring your supplier will remain in business or that the architecture chosen fits in with the Microsoft ongoing strategy for Exchange.

The C2C Approach

One solution that offers organizations these features is C2C's Archive One (for Exchange 5.5 and 2000). Built on the concept of an Application Service dedicated to Folder and Mailbox management, the product it provides is a range of folder and mailbox routines that coupled together form different archiving solutions.

Archive One is unique as it addresses the needs of managing folders. Remember that Mailboxes are folders with sub-folders of Inbox, Calendar, Sent Items, etc. Archive One was developed for messaging administrators who needed to be able to analyze folders, set criteria and action according to those criteria.

The problems shown to C2C ranged from simple loss of control of the folder system through poor establishment of permissions, through to needs to remove or identify specific data within messages and to inform other parties that the items existed.

Archive One provides a variety of routines that allow a user to

- Identify messages by a wide variety of criteria
- Automatically create new folders as repositories for the data on near-line Exchange servers
- Move, Copy or delete the data
- Maintain the security structure in the new archive store

With all solutions, it is the volume of data, the need for on-line or near-line storage that will dictate the options available, and obviously the associated costs. Archive One gives a step into the Archiving world and a chance to explore the variety of solutions available to the administrator without huge expense and long implementation times.

Where can I find out more?

For more information and free evaluation copy
www.c2c.com



Archive One

Technical Overview

For organizations that rely on Microsoft Exchange as a business critical resource, archiving technologies are fast becoming one of the preferred cost effective methods to improve flexibility, resilience and performance of the Exchange Server.

C2C Systems Archive One is an archiving system that addresses the classical archiving requirement of identifying and moving items to a less critical store but also addresses the need for cleaning the information stores, (e.g. only allowing messages up to 60 days old in a sent items folder). This unique approach is referred to as 'capacity and content control'.

What is driving the need to archive?

- The numbers of messages being sent and received are increasing at 35% - 50% per annum
- The average message size is increasing by 25% per annum
- Organizations will require a faster, scalable and more efficient mail storage approach as email becomes the focal point of corporate communications
- Industry and national regulatory bodies will require you to retain more corporate mail/calendar information

The need to archive is usually seen as the need to offload old unused data from the servers, but the continuing demands on server loadings are not going to go away. This is why Archive One includes a mechanism to quickly retrieve any archived message without any delay or impact on the user. Archived messages need not be thought of as 'gone', just 'further away'.

Automated Archiving

Archive One can be set to automatically process according to its rule set. The Archive Policy can be set to work

- Specific time of the day
- Specific intervals per month
- Multiple Archive Policies may be defined, each with its own message selection and start time criteria
- The list of mailboxes to be archived can be defined using distribution lists

Capacity Management

Archive One is designed to assist the administrator in the task of removing data from user mailboxes or folders. Archive One can be set to run at periodic intervals (see Automated Archiving) and can move or delete items or attachments according to the Policy rule set. For example administrators may refuse users to keep attachments in sent-items if they are more than 50KB and more than 40 days old.

Policy rule set

The Archive One client offers the administrator a simple and intuitive interface in which they construct Policy rules. Policy rules contain four key elements

What is to be found?

- Mail/item criteria, attachment criteria or combination (such as last modified date, size, number of attachments, attachment type, importance, etc.)

Where the search is to be performed?

- Individual mailbox, sub-set of mailbox(s) (e.g. sent items), distribution list, etc.

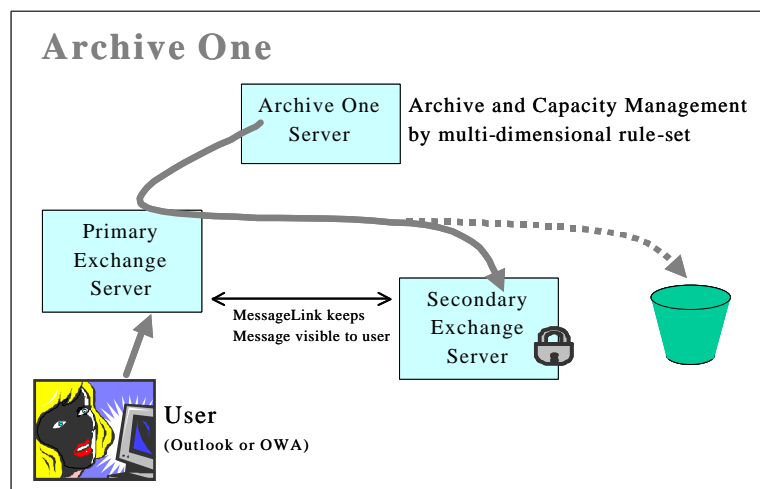
When it is to be run?

- Time of day, etc (see Automated Archiving)

What is to happen when a match is found?

- Move, Copy, Copy Latest, Delete, List

It is the ease of change of any Policy rule that make Archive One both a Capacity Management application and an Archiving solution.



Architecture

Archive One architecture is a single technology solution, this being Exchange. There are some major benefits to this:

- It harnesses the robustness of Microsoft Exchange.
- The Exchange Administration team is fully conversant with this technology
- It does not require scarce technical skills from other technology teams to implement or support
- Associated subjects such as backup are already under the control of the Exchange team

Archive One utilizes the new features of Exchange 2000 server roles, front end and back end, while still extending the functionality of Exchange 5.5. The concept of front end and back end Servers is as follows:

Front End Server. This is the main 'live' mailbox engine, whose purpose is to provide fast efficient mail services to the user. Performance, robustness and fast back-up/restore are key elements of this server.

Back End Server. This is a secondary server, whose purpose is to store archived messages. Here, high performance is not critical and being off-line occasionally will not be disastrous to the organization. This can also be referred to as near-line storage.

Client

Archive One does not need any client software installed. This makes deployment easy.

Archive One will work with Outlook 97, 98, 2000, 2002 (XP) and Outlook Web Access (OWA).

Access to Archive data

Archive One moves the messages being archived to another Exchange server in a secure public folder structure easily understood by the administrator. Users can gain access to this store simply by having a shortcut to this store placed in outlook.

Ghost Messages

To enhance the user experience, ghost messages can be used. Once you have identified the message you wish to archive, (move to the near line server), there is an option to create a ghost message. These are small 'ghost' messages left exactly where the user left the original mail. They will look like the original email; the preview pane will work as normal. When the user clicks on the message it will be automatically transferred from the near line server. This process requires no client side software or configuration.

Server

The Archive One server is an NT/Win2000 process, and should run either on the back end Exchange server or an alternative NT/Win2000 server. A single Archive

One server can process multiple Exchange servers in the same Exchange Site. Archive One will work with either Exchange 5.5 or 2000 Servers.

Advanced Knowledge Management

Archive One can be integrated with advanced knowledge management solutions such as Microsoft Share Point Portal Server or Fulcrum. This will allow you to index all documents stored and combine them with your existing knowledge management system.

Compressing Data

Archive One does not directly affect the size of an attachment; the task of automated zipping/unzipping of documents is left to other C2C products. C2C as the supplier of the most popular auto-zip application for Outlook/Exchange are proven experts in the compression of email data. Archive One will integrate with either or both of:

- MaX Compression – Outlook based auto zip/unzip
- Active Folder Compression Server – Server based zip/unzip of folders and mailboxes

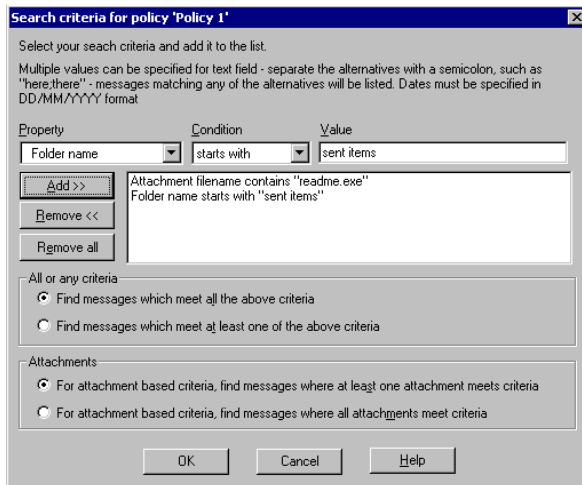
MaX Compression has more than a million client installations and is used at some of the largest companies around the world. The benefits of MaX Compression are in reducing message size, consequently improving disk and bandwidth utilization.

Compression Server can selectively ZIP attachments in mailboxes and/or public folders offering large reductions in disk utilization.

Both MaX Compression and Active Folders use standard ZIP techniques, making any zipped file readable by any standard ZIP decoder.

Security

Archive One conforms to your Exchange security model. Archive folders on the Back-End server will automatically create the same access permissions as the mailbox that is being archived. Where the administrator desires, these permissions can be changed. For example, where data is being retained for statutory purposes, you may not want the owner to have any access to the archive.



Search criteria for policy 'Policy 1'

Select your search criteria and add it to the list.
Multiple values can be specified for text field - separate the alternatives with a semicolon, such as "here;there" - messages matching any of the alternatives will be listed. Dates must be specified in DD/MM/YYYY format

Property	Condition	Value
Folder name	starts with	sent items

Add >> Remove << Remove all

Attachment filename contains "readme.exe"
Folder name starts with "sent items"

All or any criteria
 Find messages which meet all the above criteria
 Find messages which meet at least one of the above criteria

Attachments
 For attachment based criteria, find messages where at least one attachment meets criteria
 For attachment based criteria, find messages where all attachments meet criteria

OK Cancel Help

Archive One – building a rule

Content Control

Archive One can be used to monitor content of mailboxes and/or folders, and selectively delete, copy or move matched data.

Virus Control

Archive One can be used to monitor mail items of mailboxes and/or folders, and selectively delete or move matched data when the characteristics of a virus outbreak are known (e.g. messages with a subject of "I love you").

Language Variants

Archive One administration is available in English. With no client software, Archive One can be used in most language variants.

For Further information please visit

www.re-soft.com

info@re-soft.com