



# **Zimbra™ Collaboration Suite: Archiving, Discovery & Compliance Solution**

---

ZCS Archiving Methods & Technologies Whitepaper Version 1.02  
Mar 2007

## Table of Contents

1	About the Whitepaper .....	3
2	Archiving, Discovery and Compliance Enforcement Needs .....	3
2.1	Reduced Server Proliferation & Acquisition/Support Costs .....	3
2.2	Simplified Administration through Unified Cross-mailbox Search & Policy Management.....	4
2.3	Lower Storage Costs through Unified Archiving .....	4
2.4	Integrated Security Policies and Delegated Administration Access.....	4
3	ZCS Archiving & Compliance Enforcement Solution .....	4
3.1	SMTP-based Message Archiving .....	5
3.2	ZCS Archiving/Compliance Enforcement Implementation Options .....	6
3.2.1	Silo'ed Architecture: Collaboration and Archiving Solutions.....	6
3.2.2	Integrated ZCS Archiving & Compliance Enforcement Solution .....	7
3.3	Archiving Users' Existing Mailboxes.....	8
3.4	Hierarchical Storage Management (HSM) & Benefits of HSM for an Archiving Solution.....	8
3.5	Visual Search Builder for Cross-Mailbox Discovery .....	9
4	ZCS References .....	10
5	Appendix: Select US Government Content Retention Requirements for Employers	
	11	

# 1 About the Whitepaper

This document describes the methods and technologies that Zimbra Collaboration Suite version 4.5 supports to meet archiving, discovery and compliance enforcement requirements of businesses, educational institutions and government organizations. This whitepaper provides a framework for review and discussion by the key parties that would be involved in the detailed implementation effort.

We welcome any feedback or questions regarding the methods, technologies or features discussed in this document.

## 2 Archiving, Discovery and Compliance Enforcement Needs

The electronic communications archiving market is growing explosively due to the proliferation of retention and compliance policies that are often motivated by the increased regulatory overhead of Sarbanes Oxley (SOX), Health Insurance Portability and Accountability Act of 1996 (HIPAA), and many other corporate, state and federal regulations. Much of what is transmitted through an email system or an instant messaging (IM) system constitutes a ‘record’, or content that must be retained for a minimum period of time that is often established by statutes or regulations. A messaging archiving system that can index, store and purge these records according to corporate or other policies is an important component of an overall messaging management strategy, particularly in heavily regulated industries, such as financial services and healthcare.

However, messaging archiving systems can also provide other benefits, including the ability to automatically migrate messaging system content to other storage media, thereby making messaging servers more efficient and users more productive; supporting organizations during legal discovery efforts; and turning a messaging system into a corporate knowledge store, allowing users to mine data for a variety of purposes.

Unfortunately, third party archiving and compliance solutions available currently can be expensive, often doubling the necessary storage, server and administrative costs, as well as security overhead. Instead organizations around the world are looking for integrated archiving and compliance enforcement capabilities as part of their messaging and collaboration solution to leverage the following advantages.

### 2.1 *Reduced Server Proliferation & Acquisition/Support Costs*

As administrators work to address new requirements, they have generally had to resort to “bolt on” technologies for security (AS/AV), mobility, search, archiving, cross-mailbox search/discovery, and so on — each such addition requiring its own hardware footprint, as well as administration and security overhead. A collaboration solution with an integrated and flexible front-end and back-end architecture that supports capabilities including but not limited to the above capabilities can help

reduce the complexity of integration and acquisition, maintenance and administration costs.

## ***2.2 Simplified Administration through Unified Cross-mailbox Search & Policy Management***

While businesses hope to never have to access users' archives, complying with regulatory policies can sometimes require searching through user information between a fairly large duration of time and many user archives. Legal costs can become onerous if scouring through multitude of messages and attached documents to search for something specific becomes a tedious task. Sophisticated administration tools that allow setting up archival policies for membership, retention period and authorization based on Class-of-Service (COS) and search tools to perform syntactic search across the archives can help businesses reduce the amount of effort and time it would take for administrators to locate what they (or the regulatory bodies) might be looking for.

## ***2.3 Lower Storage Costs through Unified Archiving***

Users increasingly use their collaboration system as their repository to store their documents and messages thanks to the increased corporate mailbox sizes and sophisticated search capabilities that are built into the clients. IT executives are rightfully asking that if their organization must retain user email, why not utilize the same underlying software and storage for an archival solution for backup/restore and other users' mailbox failover mechanisms too.

## ***2.4 Integrated Security Policies and Delegated Administration Access***

Controlled access and an audit trail are required in order to verify data integrity and prevent any loss, damage or misuse of the archived data. For example, undocumented access to and manipulation of the archives can be deemed as insufficient security for compliance by the regulatory bodies. Also, administrators familiar with authorization, authentication and access control lists (ACL) functionality of the collaboration solution are likely to find similar methods and user interfaces for their archival solution easier to manage. Hosted Service Providers often require that administration and ongoing management of archival policies be delegated to the administrators of their individual business customers.

# **3 ZCS Archiving & Compliance Enforcement Solution**

Zimbra Collaboration Suite version 4.5 offers integrated archiving and syntactic search capabilities. Administrators may utilize the already-deployed infrastructure and tools to enable archiving for a specific user or a group of users. ZCS also offers sophisticated search capabilities based on search criteria to locate messages from across the user archives.

### 3.1 SMTP-based Message Archiving

Zimbra Collaboration Suite uses SMTP messaging as a means to achieve archiving. Messages sent or received by select users are automatically copied (forked) to a pre-selected SMTP address. The Postfix MTA is used to fork these messages and deliver them to the respective user and archive mailboxes without the user knowing about it. Before messages are forked, the archive copy is expanded to include envelope headers, BCC addresses as well as all recipients of distribution lists.

For example, a message received by Richard Thomas ([rthomas@example.com](mailto:rthomas@example.com)) and Sales Team ([sales@example.com](mailto:sales@example.com)) that comprises of Tom, Greg and Susan may appear as follows in each of their mailbox:

```
Received: from localhost (localhost.localdomain [127.0.0.1])
  by mail.example.com (Postfix) with ESMTP id D2490F68C4;
  Mon, 4 Dec 2006 17:36:37 -0800 (PST)
Received: from thom ([173.254.143.70]) by 201.236.164.219 with
Microsoft SMTPSVC(6.0.3790.211);
  Mon, 04 Dec 2006 23:48:18 -0000
Message-ID: <015f01c717fe$70f042d1$b1d6f61d@thom>
From: "Bob Smith" <bobsmith@mail.com>
To: rthomas@example.com, sales@example.com
Subject: New License Key
Date: Mon, 04 Dec 2006 23:48:18 -0000
MIME-Version: 1.0
Content-Type: text/plain;
  charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
X-Priority: 3
X-Mail-Scan: ofni
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 6.00.2900.2527
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.2527

Hi Richard,
Can you send me the license key for the software again?
Thanks, Bob
```

The forked copy of the message above that is sent via SMTP to the specified archive address includes the X-Envelope headers information and may appear as follows:

```
Received: from localhost (localhost.localdomain [127.0.0.1])
  by mail.example.com (Postfix) with ESMTP id D2490F68C4;
  Mon, 4 Dec 2006 17:36:37 -0800 (PST)
Received: from thom ([173.254.143.70]) by 201.236.164.219 with
Microsoft SMTPSVC(6.0.3790.211);
  Mon, 04 Dec 2006 23:48:18 -0000
X-Envelope-From: bobsmith@mail.com
X-Envelope-To: rthomas@example.com, rthomas@example.com.archive,
tom@example.com, greg@example.com, susan@example.com
Message-ID: <015f01c717fe$70f042d1$b1d6f61d@thom>
From: "Bob Smith" <bobsmith@mail.com>
```

```
To: rthomas@example.com, sales@example.com
Subject: New License Key
Date: Mon, 04 Dec 2006 23:48:18 -0000
MIME-Version: 1.0
Content-Type: text/plain;
    charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
X-Priority: 3
X-Mail-Scan: ofni
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 6.00.2900.2527
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.2527
```

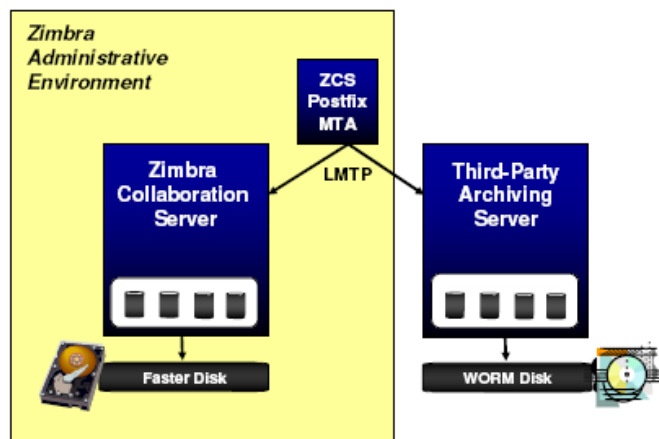
```
Hi Richard,
Can you send me the license key for the software again?
Thanks, Bob
```

## 3.2 ZCS Archiving/Compliance Enforcement Implementation Options

ZCS offers a couple different options to businesses depending on the requirements, architectural preference and existing investments.

### 3.2.1 Silo'ed Architecture: Collaboration and Archiving Solutions

A typical and non-integrated architecture for email archiving today is to deploy two distinct systems: an email server such as the Zimbra Collaboration Server and an archive server such as EMC Legato or Veritas KVS. ZCS uses SMTP forwarding to fork messages to the archive server. The archive server stores user messages and provides admin tools for user selection and searching. The archive server may or may not employ WORM (write once, read many) storage depending on the business requirements and cost constraints. Zimbra supports such a configuration. The Postfix MTA included within the Zimbra Collaboration Suite (ZCS) can easily be configured to act on a per-user basis as the forwarding agent, as illustrated below.



In this configuration, the ZCS super administrator enables archiving for a set of user accounts using the built-in `zmarchive` command line tool.

e.g. `zmarchive enable rthomas@example.com`

The tool also provides a facility to specify the forwarding address through the use of configuration parameter `amavisArchiveQuarantineTo` (e.g. `rthomas@archive.example.com`) to which all messages sent or received by the user would be forwarded. Alternatively, the archive address can also be generated by defining a regular expression that will be expanded by the system based on user profile and system information.

e.g. `zimbraArchiveAccountNameTemplate (\${USER}@\${archive}.\${DOMAIN})`

Once the configuration is completed, the MTA sends a forked copy of each message to the address specified including all envelope information and recipients of distribution lists as described above.

Note: Zimbra does not recommend that multiple users' messages be archived in a single mailbox. The setup is a lot less maintainable in the long run.

### 3.2.2 Integrated ZCS Archiving & Compliance Enforcement Solution

Zimbra Collaboration Suite version 4.5 introduces integrated ZCS archiving and compliance enforcement capabilities. ZCS facilitates mechanism to store a copy of a message for a select set of enabled users as well as provides administrator authorization and search capabilities to locate and dump certain messages that meet the search criteria for regulatory and corporate compliance enforcement.

As before, the super-administrator of the ZCS solution is authorized to select and enable archiving for users through the `zmarchive` tool. The tool also allows setting up individual archive mailbox names.

e.g. `zimbraArchiveAccount (rthomas@example.com.archive)`

The archive address can also be generated by defining a regular expression that will be expanded by the system based on user profile and system information.

e.g. `zimbraArchiveAccountNameTemplate (\${USER}@\${DOMAIN}.archive)`

Note: Zimbra recommends adding the `.archive` suffix in archive mailbox names to prevent direct mail delivery to the user archive or spoofing of the archives. As discussed before, the envelope information including the recipients of the distribution lists will be included before the message is archived in the archive mailbox.

In order to be able to easily sort archive mailboxes, ZCS provides a facility that administrators may choose to use to insert the creation date in the archive address. This can be achieved by setting the `zimbraArchiveAccountDateTemplate` as shown below.

e.g. `zimbraArchiveAccountDateTemplate(yyyyMMdd)`

Note: ZCS will not try to parse archive addresses. The date part of the archive account name is supported to make it easier for administrators to temporally sort their archive mailboxes if necessary. Administrators may choose to remove or not include the date in which case the archive mailbox sorting capability would not be available.

In addition, the integrated ZCS solution offers administrators sophisticated cross-archive search capabilities also using the `zmarchive` tool.

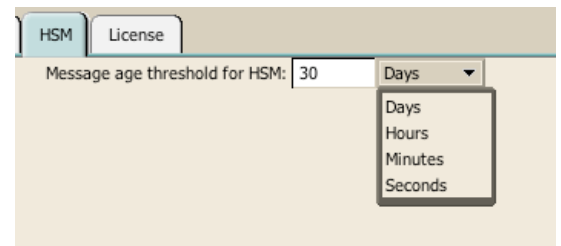
### 3.3 Archiving Users' Existing Mailboxes

ZCS supports the ability for administrators to import .pst files in bulk for all users through the use of an XML config file that contains the file location and names, user account information and optionally start date of imported messages. This capability allows organizations (and users) that wish to upload existing messages in users' mailbox (stored in .pst files) into the users' archive mailboxes for future discovery and compliance enforcement purposes.

### 3.4 Hierarchical Storage Management (HSM) & Benefits of HSM for an Archiving Solution

Email administrators are under increasing pressure to cost-effectively accommodate growing mailbox quotas and appropriately large archived mailboxes. The answer lies in leveraging commodity storage (e.g. SATA drives) as part of an efficient Hierarchical Storage Management (HSM) architecture to store rarely accessed (generally older) data on cheaper drives. It would be imperative, however, that such an architecture not affect user experience or reliability of user access to his or her mailbox messages.

The Zimbra solution natively supports volume management and auto-aging of messages from fast spindles to slower ones, such as via HSM hardware. The fact that all Zimbra data-storage is file based (as opposed to relying on raw partitions) makes HSM integration relatively straightforward (just as it simplifies supporting arbitrary network-attached storage).

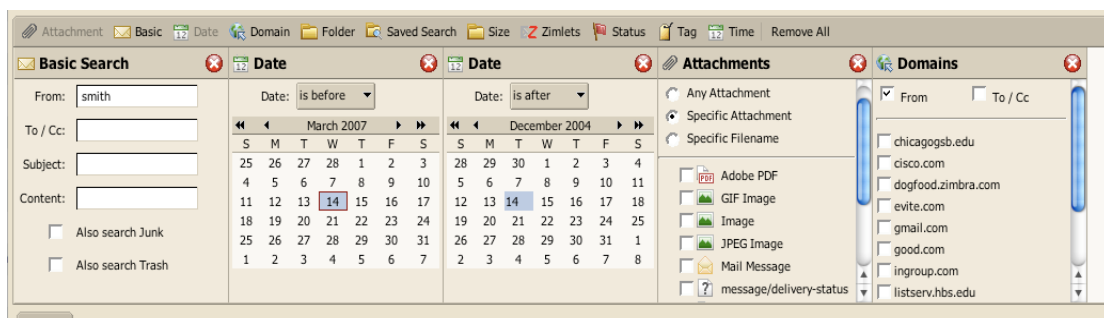


Integrated HSM capability in archiving systems offers a number of benefits to organizations. As mentioned above, HSM allows companies to reduce their overall storage costs through the use of powerful, high-value primary disks to store recent messages supplemented with cost-effective means (e.g. SATA drives) as secondary storage. By tuning the system to automatically move messages to secondary disks based on age, administrators can achieve both higher overall throughput and lower costs. Another advantage of HSM is its ability to reduce complexity and manageability. By partitioning messages based on age, administrators are better equipped to handle routine tasks such as managing backups and system maintenance for a tiered storage system. Finally, organizations (especially service providers) can also use HSM to provide varying levels of service to their customers.

### 3.5 Visual Search Builder for Cross-Mailbox Discovery

As discussed in section 2, complying with regulatory policies can sometimes require searching through user information between a fairly large duration of time and many user archives. Legal costs for “data discovery” can become onerous if scouring through multitude of messages and attached documents to search for something specific becomes a tedious task.

Sophisticated administration tools that allow setting up archival policies for membership, retention period and authorization based on Class-of-Service (COS) and search tools to perform syntactic search across many archive mailboxes can help businesses reduce the amount of effort and time it would take for administrators to locate what they (or the regulatory bodies) might be looking for.



The Zimbra Solution offers a sophisticated and integrated tool called Visual Search Builder to meet the complex data discovery needs. As shown above, the Visual Search Builder allows authorized users to perform structured searches for data across a range of (or all) archived mailboxes between certain dates and from/to certain people. Queries can be built to locate messages with certain attachment types, words in the subject or body of the messages, in select domains etc.

Once the query is composed and run, the server leverages the built-in indexing mechanism of the Zimbra solution to scour through conversations from the selected

archives that match the query pattern. The matching conversations are automatically stored in a newly created mailbox for further refinement if need be. Access to this mailbox can also be made available to authorized legal or regulatory authorities for further manipulation without the risk of negative impact to the integrity of the archives or the overall system.

Thus, through the use of a web-based visual search builder, administrators and other authorized users can easily locate relevant information through volumes of data in a timely manner, thereby mitigating discovery costs and meeting the needs of regulatory compliance laws.

## 4 ZCS References

ZCS Documentation: <https://www.zimbra.com/products/documentation.html>

ZCS Wiki: <http://wiki.zimbra.com/>

ZCS Architecture Overview:

<http://www.zimbra.com/pdf/Zimbra%20Architectural%20Overview.pdf>

ZCS Blog: <http://www.zimbra.com/blog/>

ZCS Bug & Enhancement Tracking (*Bugzilla*): <http://bugzilla.zimbra.com>

ZCS Software Download: <https://www.zimbra.com/community/downloads.html>

ZCS Roadmap: <https://www.zimbra.com/products/roadmap.html>

## 5 Appendix: Select US Government Content Retention Requirements for Employers

Type of Information	Retention Period	Applicable Statute(s)
Written training agreements, summaries of applicants' qualifications, job criteria, interview records and identification of minority and female applicants.	Duration of training program plus three years	Fair Labor Standards Act, National Labor Relations Act
Hiring documents, including job applications, resumes, job inquiries, and records of refusals to hire.	One year from date of personnel action	Age Discrimination in Employment Act
Promotion, demotion, transfer, selection for training, layoff, recall or discharge.	One year from date of personnel action	Age Discrimination in Employment Act
Payroll records, collective bargaining agreements, including any changes, individual contracts, written agreements under the FLSA, sales and purchase records, and certificates and notices of the Wage and Hour Administrator	Three years	Fair Labor Standards Act and National Labor Relations Act
Personnel records of an individual whose employment has been involuntarily terminated	One year from the date of the termination	Americans with Disabilities Act
Consumer allegations of personal injury or harm to health, reports of occupational disease or injury and reports or complaints of injury to the environment submitted to the manufacturer, processor, or distributor from any source	Thirty years for employee claims of occupational disease or occupational health problems	Toxic Substances Control Act
Records of medical examinations required by law	Duration of employment plus thirty years, unless OSHA requirements provide otherwise	Occupational Safety and Health Act
Records concerning complaints of handicap discrimination, (in programs and activities receiving or benefiting from federal financial assistance) and relevant employment records of charging party and employees in similar positions.	Three years	Rehabilitation Act of 1973

Type of Information	Retention Period	Applicable Statute(s)
Personnel records concerning any discrimination charge brought by any agency or individual (e.g., records about charging party and all other employees holding similar positions, application forms, or test papers completed by all applicants for same position)	Until final disposition	Title VII of the Civil Rights Act of 1964
Records of monitoring exposure to hazardous materials	Thirty years	Occupational Safety and Health Act
Manufacturers, processors, or distributors of any chemical substance must retain records of employees' "significant adverse reactions" to health or the environment.	Thirty years from date such adverse reaction first reported to or known by person maintaining record	Toxic Substances Control Act
Employee benefit plans (such as pension or insurance plans); seniority and merit systems which are in writing	Duration of plan and for at least one year after plans' termination	Age Discrimination in Employment Act
Any correspondence, inquiries or notes relating to individual eligibility determinations	Indefinitely	Employee Retirement Income Security Act of 1974
Wage and hour records for the purpose of determining retirement benefits.	Indefinitely; alternatively six years following date of lump sum distribution.	Employee Retirement Income Security Act of 1974 and the Fair Labor Standards Act
All personnel or employment records including application forms, resumes, other hiring records; records regarding promotion, demotion, transfer, layoff, discharge, pay rates, or other compensation terms	One year from date records made or personnel action taken, whichever is later	Title VII of the Civil Rights Act of 1964
Personnel records concerning a charge of discrimination filed or an action brought against an employer under Title VII or the ADA	Until final disposition of the charge or the action (the date of expiration of the statutory period within which the aggrieved person may bring an action in U.S. District Court or the date such litigation is terminated)	Americans with Disabilities Act

---

## Copyright

Copyright Zimbra, Inc. 2007. All rights reserved. Zimbra and the Zimbra logo are trademarks of Zimbra, Inc.