

ZCP 7.0 (build 41322)

**Zarafa Collaboration
Platform**

Zarafa Archiver Deployment Guide



Zarafa

ZCP 7.0 (build 41322) Zarafa Collaboration Platform

Zarafa Archiver Deployment Guide

Edition 1.0

Copyright © 2011 Zarafa BV.

The text of and illustrations in this document are licensed by Zarafa BV under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at [the *creativecommons.org website*](http://creativecommons.org/website)¹. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Linux® is the registered trademark of Linus Torvalds in the United States and other countries.

MySQL® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Red Hat®, Red Hat Enterprise Linux®, Fedora® and RHCE® are trademarks of Red Hat, Inc., registered in the United States and other countries.

Ubuntu® and Canonical® are registered trademarks of Canonical Ltd.

Debian® is a registered trademark of Software in the Public Interest, Inc.

SUSE® and eDirectory® are registered trademarks of Novell, Inc.

Microsoft® Windows®, Microsoft Office Outlook®, Microsoft Exchange® and Microsoft Active Directory® are registered trademarks of Microsoft Corporation in the United States and/or other countries.

The Trademark BlackBerry® is owned by Research In Motion Limited and is registered in the United States and may be pending or registered in other countries. Zarafa BV is not endorsed, sponsored, affiliated with or otherwise authorized by Research In Motion Limited.

All trademarks are the property of their respective owners.

Disclaimer: Although all documentation is written and compiled with care, Zarafa is not responsible for direct actions or consequences derived from using this documentation, including unclear instructions or missing information not contained in these documents.

The Zarafa Archiver provides an integrated archiving solution for Zarafa installations.

¹ <http://creativecommons.org/licenses/by-sa/3.0/>

1. Introduction	1
2. Conventions	3
3. Archiving	5
4. Scenarios	7
4.1. Email archiving for storage purposes	7
4.2. Email archiving for keeping an email history	7
5. Choices	9
6. Hardware considerations	11
6.1. CPU	11
6.2. Memory	11
6.3. Storage	11
6.4. Virtual vs Physical	12

Introduction

This *Deployment Guide* provides more information about the Zarafa Archiver on different configurations and setups. This information can be used for architects, IT manager and engineers to support their decision making process on archiving strategies.

The Zarafa Archiver provides a Hierarchical Storage Management (HSM) solution for the Zarafa Collaboration Platform, with the following features and advantages:

- More rapid recovery from downtime incidents
- Improved messaging server performance
- Seamless End User Experience
- Flexible PST Migration
- Use less expensive storage tiers for long-term archiving

With the Zarafa Archiver older messages will be automatically moved to *slower* and thus *cheaper* storage. The slow storage consists of one or more additional Zarafa Archive servers which sole task it is to store archived messages.

Although the older messages are stored on different servers, the user will not notice this as the Zarafa Archiver provides transparent stubbing. Older messages can still be accessed from the main mailbox of the user, by using the (de)stubbing feature.

To activate the archiving function for users, they need to be coupled to an archive mailbox. The archive mailboxes will be located on one of the archive servers. The archive servers have exactly the same storage architecture as a normal Zarafa server, all mapi properties are stored in a MySQL database and all attachments are stored compressed on disk.

The Zarafa Archiver uses the Zarafa's multi-server technology to access archive stores in a seamless way using transparent user authentication. Nonetheless, the Zarafa Archiver can be used in a single-server setup with limited functionality.

Zarafa Archiver is an additional product and is not a default component of the Zarafa Collaboration Platform. Subscriptions of Zarafa Archiver can only be used with a valid subscription on the Zarafa Professional or Enterprise edition.

Conventions

Before starting to deploy the Zarafa Archiver it's strongly advised to read this chapter to understand the different terminology.

- **Primary Server**

The primary server is the server with the best performance and best IO subsystem, that contains the mailboxes of the users holding the most recent data.



Note

Although the term **Primary Server** suggests that there's only one primary server, multiple primary servers can exist in a multi-server environment. In this document no distinction will be made between a single-server or multi-server environment unless explicitly stated.

- **Archive Server**

The archive server is the server, with a substantial slower and cheaper IO subsystem, that contains the archives for the stores that reside on the primary server.



Note

An archive server is another zarafa-server with the sole purpose of providing storage for one or more archive stores. In a multi-server environment this server will be just another node in the cluster.



Note

Unlike primary servers, there's no need for a multi-server environment to have multiple archive servers.

- **Primary Store**

The primary store is the store that resides on the primary server and on which a user normally works.

- **Archive Store**

The archive store is the store that resides on the archive server and which is used for storing the archived messages from the primary store.

- **zarafa-archiver**

The archiver is the application to manage the archiving. Basically it can be used to attach primary stores to archive stores and execute archive runs. It can be installed on any Zarafa server to connect to the primary or archive server using SSL authentication. It can also be used on a single server, using **zarafa-server**'s unix socket.

- **Stubbed Message**

A stubbed message is a message in the primary store that acts as a placeholder for the archived message. These messages occupy virtually no space, but allow a user to see that a message was once there. On top of that it acts as an entry point to the archived version of that message.

- **Archive Configuration**

An archive store can be configured in two ways, one-for-one and one-for-many. This is not a system wide configuration and can be setup for each archive independently.

This allows for hybrid systems where N users with small to medium stores can be placed on M archive stores (where M is significantly smaller than N) and users with big to huge stores can be placed on dedicated archive stores.

- **One-for-One Configuration**

In a one-for-one configuration one archive store is attached to one primary store.

The advantage of this configuration is that it's faster as the archive store itself is kept smaller.

The disadvantage is that for each user an additional non-active user needs to be created (since there's a one-to-one mapping between stores and users in Zarafa).

- **One-for-Many Configuration**

In a one-for-many configuration one archive store is attached to multiple primary stores. For each attached primary store a folder is created in the archive store that will act as the root of the archive for that particular primary store. The advantage of this configuration is that less additional non-active users are required. The disadvantage is that the archive will become slower if the total amount of archived data in it grows.

Archiving

Archiving solutions have different stages to process the email data. Below the four stages of the archive process are described.

1. **Copying** - In this stage all messages that are eligible for archiving are copied from the primary store to the archive store. When a copy of a message is made, an internal reference to this copy is placed in the original message.
2. **Stubbing** - In this stage all messages that are eligible for stubbing are stubbed. A stub is defined as the original message with the body and attachments removed. A message is eligible for stubbing when it reaches the specified minimum age AND archived copies are present. So a message is never stubbed if it's not yet copied to the archive store.
3. **Deleting** - In this stage all messages and stubs that are eligible for deletion are removed from the primary store. A message is eligible for deletion when it reaches the specified minimum age AND archived copies are present. So a message is never deleted from the primary store if it's not yet copied to the archive store.
4. **Purging** - In this stage messages that reached a specific age will be deleted completely from the archive store. In this case messages will not be available on both the primary and archive server.

Every of these stages can be configured in the configuration file of the Archive Controller. Before deploying the Zimbra Archiver make sure the configuration options are carefully set according to the archive strategy. The default installed configuration will only copy message items from linked mailboxes to the archive mailboxes.

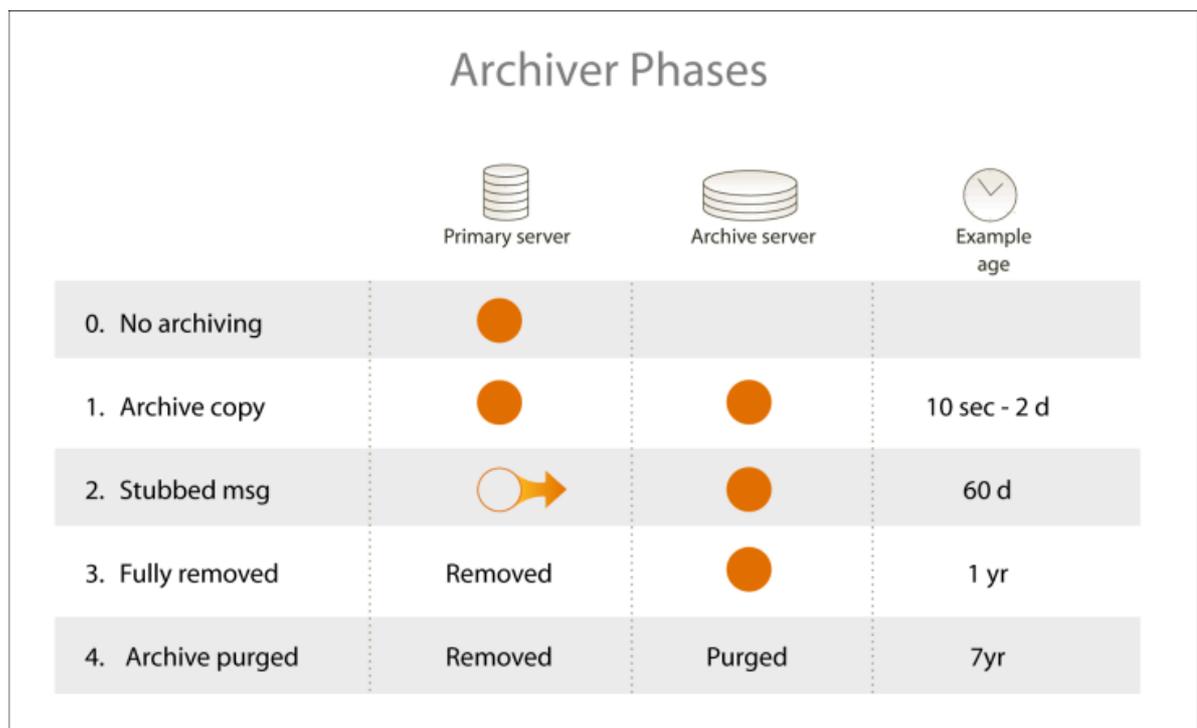


Figure 3.1. Archive phases

Scenarios

Email archiving are typically designed for two different scenarios.

- Email archiving for storage purposes
- Email archiving for keeping an email history and compliancy

Zarafa Archiver can be configured for both purposes.

4.1. Email archiving for storage purposes

The volume of incoming emails is still growing at most organisations, where most mailbox quotas will not be increased. Users will organise and archive emails to pst files to keep their mailbox clean.

By implementing the Zarafa Archiver all older emails can be stored on central located archive server. Where the archive database is typically stored on slower and cheaper storage. Storing the older emails on a different server will result in a smaller mailboxes of the primary server and therefore a better performant primary server. By having the archived emails stored centrally, the disadvantages of pst files can be eliminated:

- Not possible to access archive pst by multiple users
- PST are unmanagable by organisation
- PST files are often stored locally and therefore are not part of the backup
- PST files have slow network performance

With the stubbing feature the user can transparently access emails from the primary mailbox, where the actual body and attachment of the emails are accessed from the archive server. With the direct connection to the archive mailboxes, users can access there archived emails which are not available as stub in the primary mailbox.

4.2. Email archiving for keeping an email history

As email is used for more and more official communication, companies want to ensure no emails are lost. Email archiving solutions for keeping email histories will automatically archive all incoming and outgoing emails. All archived emails will be stored on a central archive server and can be kept for long-term accountability.

As normally all users access to their archive mailbox, so users can still access older emails.

With the Zarafa Archiver all incoming and outgoing emails can be directly archived after the send/delivery process.

The Zarafa Archiver controller can still archive, stub, and cleanup primary mailboxes to reduce the size of mailboxes, however in this case the emails that were already in the archive database will not be modified.

Choices

In the previous chapter the two typical archiver setups were described:

- Storage based
- History based

This chapter can be used as a guideline for configuration choices in the archiver rollout, based on these two scenarios. All configuration options will be explained together with typical value for the setting.

Archive permissions

When the archiver is deployed for storage optimisation users will require the same permissions on the archive mailbox as on their primary mailbox. In a history based archive the permissions on the archive mailbox will be typically read-only, so the user can delete no items from the archive.

With the Zarafa Archiver the default permissions can be set when the a user is attached to an archive mailbox.

Copying time

In a storage based archive the message will be typically archived after 30 or 60 days, so only the more important will be copied to the archive mailbox and non-important email will not take any space on the archive server as it will be removed prior to the archiving actions. In a history based archive the messages will be automatically archived at delivery time. In this case it's not necessary to set the archive time, as the messages are already in the archive store available.

To use the history based archive feature both the delivery agent and spooler need to have the option **archive_enabled** set to yes in order to store directly all emails at receive/sent time.

Stubbing time

In both store and history based archives, stubbing will typically be set to 30-60 days. When choosing the right stubbing time, please keep in mind the body and attachments of stubbed emails can only be accessed when there is an online connection to the archive server. When using Outlook with caching modes the items can't be accessed when no server connection is available. For this reason the stubbing time should not be set to low.

Deleting time

As the deletion time will make items unavailable in the primary mailbox, the value will be typically have a value of 6 months or 1 year. After this time users normally don't open those emails that frequently anymore and in case the emails should be access, the archive mailbox can be added as a delegate mailbox. Having a high time for this setting will result in larger amount of items in a folder. When users should always access all items from the primary mailbox, the delete_enable setting of archiver should be set to *No*.

Purge time

The purge setting will delete items completely from the archive mailbox, so emails will not be available at all. In case of a history based archive, so users will not be able to cleanup their own archive mailbox, this setting is recommended to use. When this setting is not configured, items will be never removed from the archive server. The typical time for this setting is 5-7 years, depending on your considerations and reasons for using the history based archive. When using the storage based archive the preferred time can be configured based on user experience and the average lifetime of items in mailboxes.



Note

This configuration option should be carefully considered including any legal aspects, as it will remove items from the email system.

Quota

Like a normal installation of Zarafa Collaboration Platform archive servers can also have quotas configured. When using a history based archive setup a reached quota will stop the history archiving process, so in this case quota should not be set. In case of a storage based archive the archive is typically configured to reduce the data size on the primary server. When the quotas are set on the archive server the archive process will not be able to archive emails when the quota is reached and emails will be kept on the primary server, so therefore it's not recommended to configure quota on the archive.

Quota on primary server

Also when the Zarafa Archiver is integrated with a running ZCP environment mailbox quotas can be configured. When having the stubbing setting enabled, the stubbed items will only use around 1kB of size in the database.

Hardware considerations

Sizing and implementing Zarafa Archiver requires careful planning to ensure that the product can perform to expectations and scaling as the customer requirements grow and fits in the existing infrastructure.

6.1. CPU

In most deployment scenarios the Zarafa Archiver will be installed on a dedicated server which doesn't provide any other services. The most heavy actions performed on the archive server is the actual archive run, which is typically done once or twice a day. This archive run will in general do a lot of database transactions and is not running CPU intensive calculations.

The archiver run will archive the mailboxes one by one. For the right choice of the amount of CPU's it's recommended to use a multi-core system where one core is reserved for the Zarafa-server, one for the MySQL database server and one for the actual Archive controller.

The advised amount of CPU cores is 4 where the preferred architecture is 64bits, so more than 4Gb of memory can be allocated by processes.

6.2. Memory

In normal setups of ZCP, memory is one of the most important hardware components as both MySQL and Zarafa-server cache requested data, so the second time the calendar or inbox is opened all items will be retrieved from the cache.

In an archiving setup with the stubbing feature enabled, the archived data is only accessed when an archived email is really opened. In this case the advantage of caching is limited, as the archived emails are normally not accessed over and over again.

However when the users open the archive store as a secondary mailbox, caching is more important as the user will directly access the archive store.

The advised amount of memory is:

Amount of archive users	Stubbing enabled	No stubbing
<500	2Gb	4Gb
>500 < 2000	4Gb	8Gb
>2000	8Gb	16Gb

6.3. Storage

In most cases, you require RAID-based storage to achieve your storage requirements. To maintain performance and reliability, consider hardware-based RAID rather than software-based RAID. To achieve redundancy on striped arrays while maintaining performance, consider the RAID scheme carefully.

RAID level 5 is popular to arrange cost-effective methods of achieving redundancy while keeping a good read performance. However, write actions on RAID level 5 will cost performance for storing parity bits. Therefore, in most cases discussed, a RAID1 or RAID10 should be considered. The RAID controller should also provide a battery-backed read and write cache to aid performance and prevent data corruption at power failures.

Before you use partitions on a storage area network (SAN), consider the I/O load together with any other applications that are already using the SAN to ensure that the performance can be maintained.

Ideally, discuss the implementation with the storage vendor or responsible to ensure that you achieve the best performance. Typically, you should create LUNs across as many suitable disks as possible, using entire disks rather than partial disks to prevent multiple I/O-intensive applications from using the same disks.

6.4. Virtual vs Physical

When above storage and memory requirements are matched, the Zarafa Archiver can run on both virtual as physical servers. Most organisation use the Zarafa Archiver on a virtual platform, as this is the default for most of these servers.