Retain 4.9.2
How Retain Works

May 2021

Legal Notice

For information about legal notices, trademarks, disclaimers, warranties, export and other use restrictions, U.S. Government rights, patent policy, and FIPS compliance, see https://www.microfocus.com/about/legal/.

Contents

Preface

About Retain

Retain Unified Archiving

- Archives email and text messages.
- Catalogs them for search and retrieval.

Retain is not a backup/restore system.

About This Guide

This guide presents a conceptual understanding of Retain Unified Archiving version 4.9.2.

For Documentation Feedback

Scroll to the bottom of the online page and enter a User Comment.

To Join the Retain Idea Exchange

Got an idea for a new Retain feature or enhancement? Log in to Microfocus.com and join the conversations happening inside the Retain Idea Exchange.

For Additional Documentation

See the Micro Focus Knowledge Base website.

To Contact Technical Support

Browse to the Micro Focus Support Page (https://www.microfocus.com/support-and-services/) and begin typing *Retain Unified Archiving* in the **Search for a Product** field.

For Sales

Go to the Micro Focus Contact page (https://www.microfocus.com/en-us/contact).

For Professional Services

Contact Micro Focus Professional Services by sending an email to sales@microfocus.com or calling (877) 772-4450.

1 Key Concepts

What Retain Does

Retain

- Provides organizations with legal compliance and litigation protection.
- Frees up disk space on messaging systems.
- Enhances message-data management.
- Lets administrators
  - Perform advanced searches on archived messages
  - Review archive status and activity
  - Retrieve archived messages and data
  - Generate reports on archived messages and data
- Optionally, users can access and search their archived personal data.

IMPORTANT: Retain is NOT a backup or emergency-restoration system.

You must ensure that your messaging systems are backed up by other software and systems designed for that purpose, as required by organizational and governmental regulations.
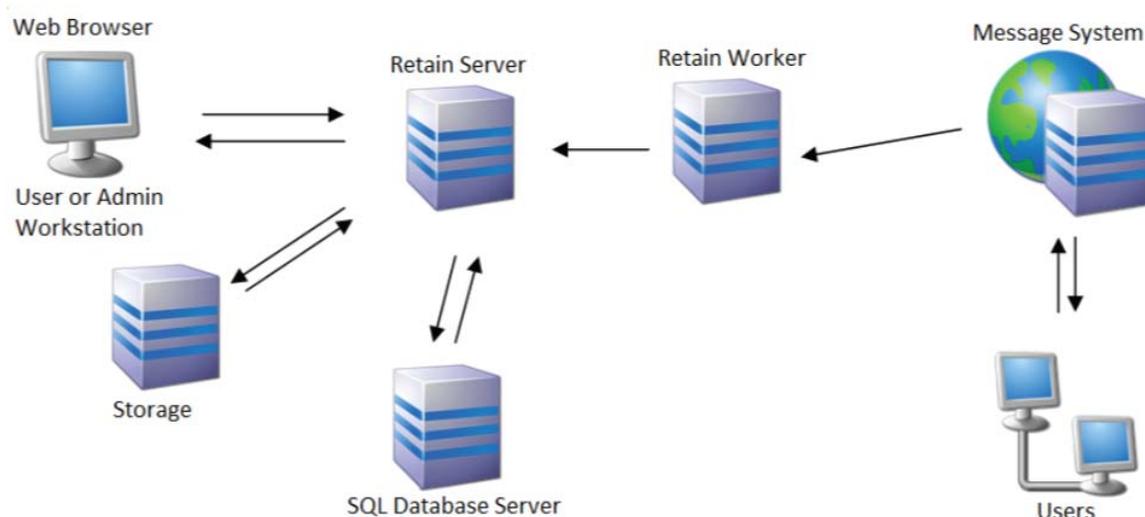

How Retain works

Briefly, Retain

1. Connects to targeted message systems.
2. Collects data by using each message system's defined APIs (for example, SOAP for GroupWise and Exchange).
3. Archives the collected data.
4. Indexes the archives in an SQL database.
5. Provides search-access to users with sufficient access rights, as managed by Retain administrators.

Retain Components and Their Roles

Retain consists of several main parts which can be installed on the same server or on different servers, depending on the size and complexity of your messaging infrastructure.

- **Retain Server:** One per system.
  - Controls all Retain functions.
  - Houses the archive.
  - Manages Retain Workers
  - Stores index-targeted data in the database.
- **Retain Workers:** One or more per system; often one per messaging server.
  - Can be installed with Retain, on the targeted mail server, or on a standalone server.
  - Collects data and transfers it to the Retain server.
- **SQL Database:** One database per system.
  - Can be installed with Retain, as a single-server, or clustered.
  - Stores message header data, user data, and links to archived messages.

  IMPORTANT: The Retain software doesn't include a database. You must install and maintain one of the supported SQL databases.

- **Reporting and Monitoring Server:** One per system.
  - Can be installed with Retain or on a standalone server.
  - Keeps job and server statistics.
  - Monitors mailbox errors.
- **Indexing Engine:** Installed on the Retain server (standard indexer), or separately in a High-Availability Indexer cluster.
  - Indexes all the data.
- **Stubbing Server:** Installed on the Retain server.

Only for GroupWise 8.0.1 and later

1. Removes large messages from GroupWise storage.
2. Archives the messages.
3. Creates a database `stub` (link to the message in the archive).

Message-access experience is unchanged for GroupWise users.

◆ **Retain Router:** Installed in the network DMZ.

   ◆ Gathers phone message data using REST, from registered Android or Blackberry devices.

   ◆ Stores the data until it can forward it to the Retain Server for archiving as with other systems.
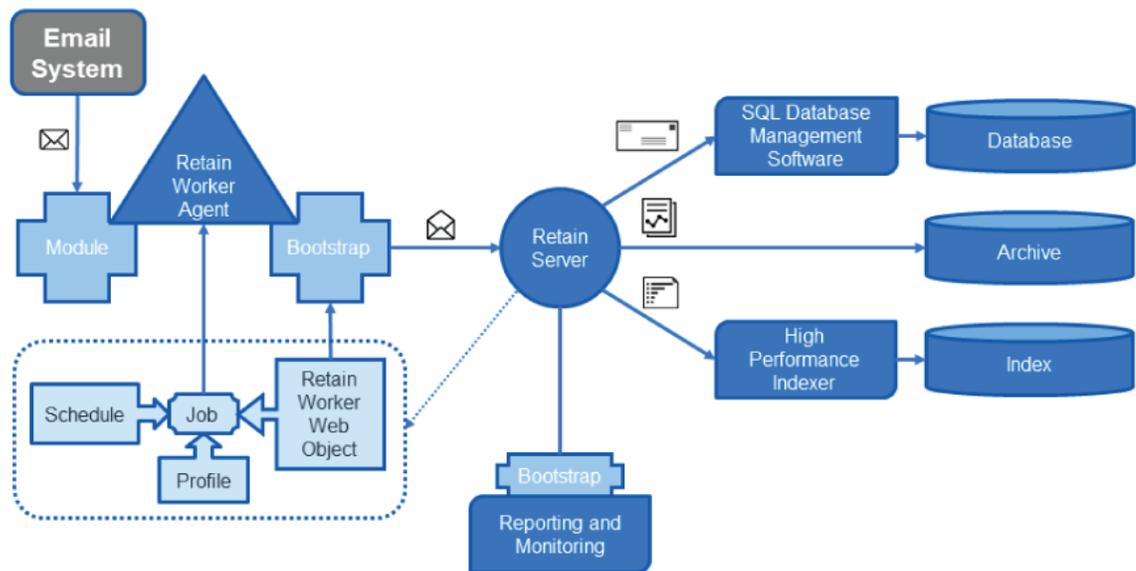
Modules Overview

The Module is how Retain connects to your messaging system.

Select the module that corresponds with your messaging system and configure the module.

Key Concepts

You must install a Module to connect to your messaging system, this reveals the Data Collection section in the console for setting up a job.

Archiving involves connecting Retain to your messaging system with a Module, setting up a schedule for when you want archives to happen, a profile that defines what should be archived, a worker agent needs to be installed, configured and connected to Retain, and finally a Job needs to be configured with all the previous components as well as who to archive.

To set up archiving, you need to configure:

Module: How to connect your messaging system.

Schedule: When the job runs.

Profile: What types of messages the job is to archive.

Worker: The Worker is a combination of the Retain Worker Agent that can sit on the Retain server, the messaging server or another server, and the Retain Worker Web Object in the Retain Server console where you create the worker bootstrap which tells the Worker Agent how to connect to Retain Server.

Job: Who to archive, what expiration date to create, what schedule, profile, and worker to use.
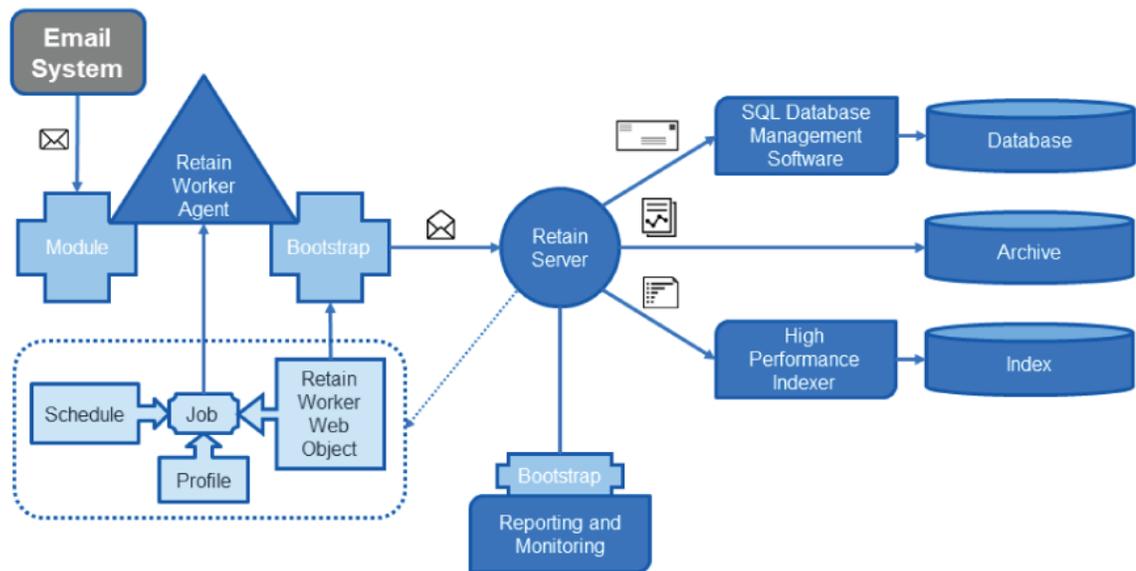
Modules

Modules are how Retain connects to a messaging system and jobs specify what Retain archives.

Modules are where you provide the data needed for Retain to connect to the messaging system. This includes the address of the messaging system server and credentials to enter the system, such that Retain can access all the mailboxes.

This is used by the Retain Worker to connect to the messaging system. It is the Retain Worker that does all the work of bringing the data into Retain.

Since the Worker Agent software can be installed on a server separate from the Retain server itself we need to provide it with a bootstrap so it knows how to connect tot he Retain server.

In Retain 4.2 and above, multiple modules of the same type can be enabled. This allows Retain to connect to multiple messaging systems of the same type at the same time that do not share common access. The modules can be given different names to make it easy to distinguish.

Multiple modules is enabled for GroupWise. Exchange, Google Apps (G Suite) and Mobile modules. It does not apply to the other modules.

Module Configuration

After configuring a module, it is recommended you manually refresh the Address Book.

| Refresh Address Book | Sync job is not running at this time. |

| Mailbox Mapping Options |

Configured Modules

GroupWise4.214	GroupWise4.160	Exchange4.213	GBS Notes
Configure	Configure	Configure	Configure
Install Date 16-Nov-2016 10:09:26	Install Date 22-Mar-2017 15:19:47	Install Date 16-Nov-2016 10:09:25	Install Date 08-Dec-2016 10:22:24
Last Address Book Cache 23-Mar-2017 01:02:10	Last Address Book Cache 23-Mar-2017 01:02:13	Last Address Book Cache 23-Mar-2017 01:02:14	Status Configured.
Status Configured.	Status Configured.	Status Configured.	

You enter the Module name in the Module.

For example, the GroupWise4.214 module has a different domain and trusted application key compared to the GroupWise4.160 module.

Profiles

Once you have configured a module you can configure a Profile for the module type.

Profiles describe what to archive.

Not all modules have profiles.

How Retain Stores the Archives

Retain uses a hybrid data-storage approach.

- **SQL Database:** Stores meta data, folder structures, attachment information, and links to messages in the archive file.
- **Archive File System:** Stores message text and attachments in a single-instance storage scheme that is designed to protect against tampering.

Retain's Archive Data Organization

Retain organizes data streams in the archive by

- The user who created or used them

- The time they were created.
- The data source.
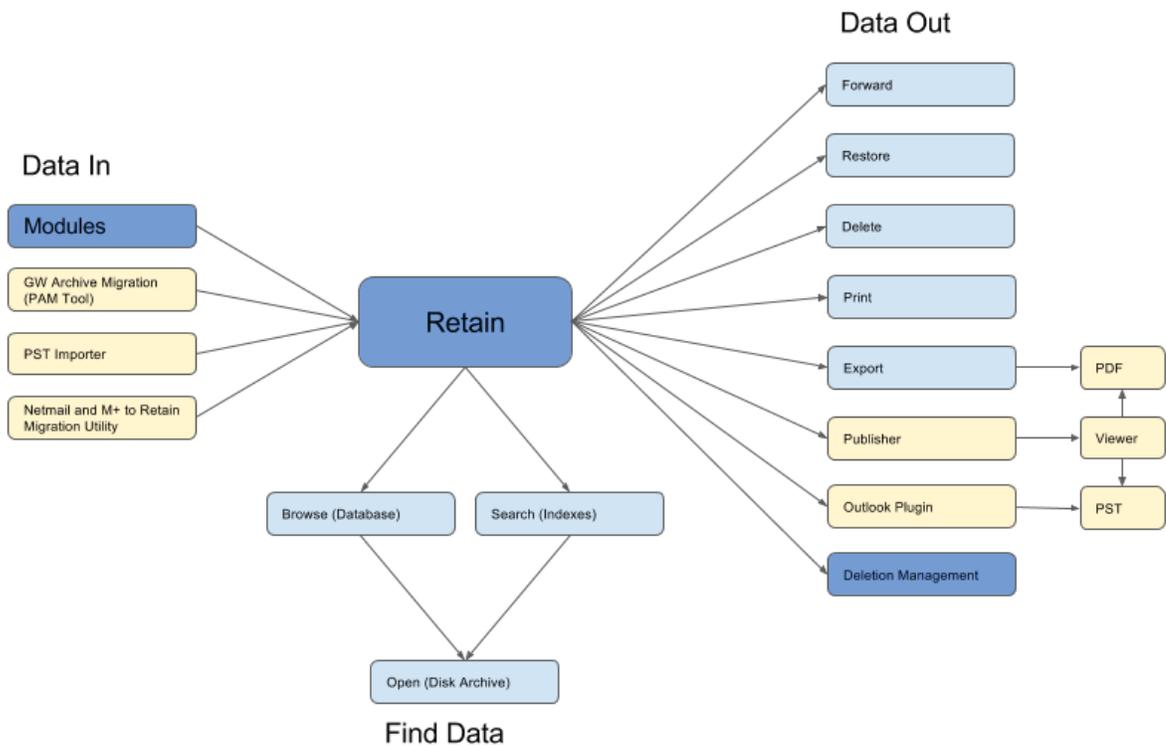
Retain then indexes items for searching, exporting, and publishing from the archive.

2 Retain Functional Overview

Retain provides three functions:

- Importing and archiving data
- Finding specific data in the archives
- Exporting data from the archives

Importing and Archiving Data

Live Data

Modules dredge data from live messaging systems.

Offline Data

Workstation tools migrate off-line data into Retain. In the top-right corner of the Retain web console, click the Tools drop-down.

Offline data migration tools include

- The GroupWise Archive Migration Tool
- The PST Importer
- The Netmail and M+ to Retain Migration Utility.

Metadata Vs. Message Data

Metadata is used for listing the content of mailboxes and delivering search results.

When you search in Retain, you are leveraging the indexes and metadata that Retain uses to find things quickly.

When you open a specific message, you are viewing the message data from the archive.

Exporting Data from Retain

Using the Web Interface

Using the web interface, click the checkbox for each message you want to export, then in the list of actions, click the action you want to take.

By default, users can view and save attachments, as well as forward and print messages.

Using the Outlook Plugin to Export Messages

The Outlook plugin lets users search and download messages from Outlook.

Dealing with Large Quantities of Data

For large quantities of data, use the a workstation tool found under the Tools menu. See "Retain Publisher and Viewer" in *Retain 4.9.2: User Guide*.

Removing Data from Retain

When older messages have reached the end of the data retention policy, use Data Removal as described in "Deleting Data" in *Retain 4.9.2: Configuration and Administration*. to permanently remove messages from the archive.

3 Target Systems and Data Streams

The systems and data streams which Retain can archive are listed below:

- "Smart Phone Targets" on page 17
- "Email System Targets" on page 17
- "Search Engine Targets" on page 17

Smart Phone Targets

Retain can archive PIN, SMS, and phone call data, as configured.

- Blackberry (BES Server)
- BBM Enterprise (on all platforms)
- Android
- IOS and Android (via CellTrust Secureline)

Email System Targets

Retain archives all specified data, which can include: email, notes, appointments, meetings, reminders, and tasks, from the following email systems.

- Exchange
- Office 365
- GroupWise
- Gmail
- Bloomberg
- GBS Notes

Search Engine Targets

For Search Engines, Retain can archive the search criteria specified by search-engine users.

4 Jobs

How Archive Jobs Work

When an archive dredge job is running, the work flow follows this pattern.
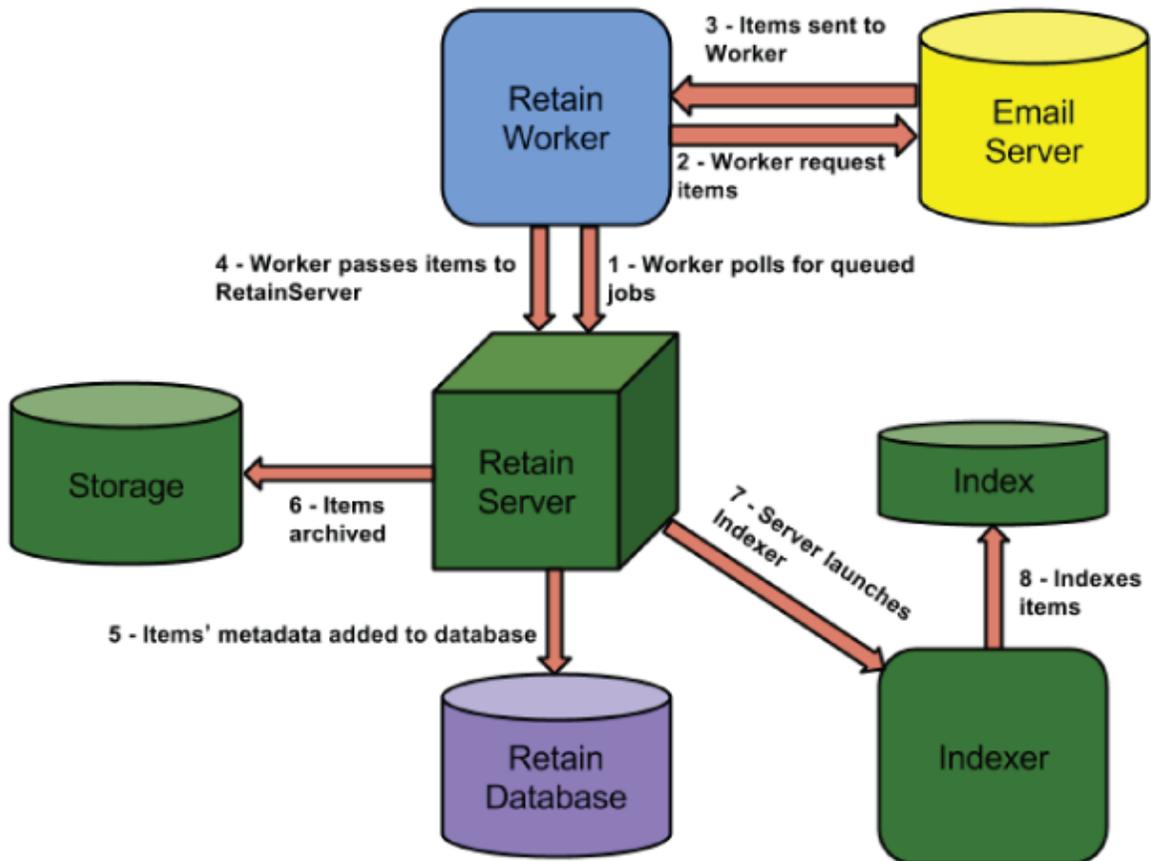
1. The Worker polls the Server every 10 minutes (default) for new queued jobs it may need to run. It launches the job.

2. The Worker connects to the messaging system, logs in to each mailbox, and requests the items in that mailbox based on the settings in the profile.

3. The messaging system responds by sending the items to the Worker.

4. The Worker sends smaller items to the Server. For large items, it sends the item's metadata and awaits instructions from the Server as to whether the item already exists.

   ◆ If it already exists, the Retain Server notifies the Worker that it does not need to send the item over.

   ◆ If it does not exist, the Retain Server notifies the Worker to send the item.

5. The Retain Server updates the Retain database with a record of the item's metadata if a record does not already exist.

6. The Server adds the item to the storage area on disk.

7. The Server launches the indexing process (if it is not already running) to begin the indexing process.

8. The Indexer indexes any items that need to be indexed.

An Archive Job Example

1. User A sends message 1 to User B.

2. When a Worker processes User A's mailbox, Retain archives Message 1 in the Retain archive and creates a record in the database that points to the archived message and associates it with User A's mailbox.

3. When a Worker processes User B's mailbox, Retain notes that message 1 is already archived and that a database record already exists.

4. Therefore, Retain only needs to update the database record so that the message is also associated with User B's mailbox.

5 Retain Users and the System Address Book

As Retain gathers mail from users of the targeted message system, it stores the users in the System Address Book, which is cached locally on the Retain server.

As new users have messages archived, they are added to the address book.

User Removal Not Allowed

The address book contains all users (current and past) who have archived data.

As long as users have archived messages, they cannot be removed from the address book.

Retain Handles Users with the Same Name

Retain distinguishes between multiple users with the same name. For example, "John Smith" added today, is a different user from "John Smith" who began working at the company six months ago, and from "John Smith" who left the company last year.

6 Retention Services and Item Store Flags

Retain keeps an "item store flag" to ensure that no item gets left behind.

With Exchange and O365 Holds and the Recoverable Items folder can be used for retention compliance. With On-Premise Exchange a journaling mailbox can be used but it is not recommended.

Gmail, by default, does not have a retention service.

GroupWise, on the other hand, has its own built-in feature called "Retention Services" that prevents items from being emptied from the mailbox until they have been successfully archived.

The following sections explain Retain's support of the GroupWise Retention Services, followed by a discussion of how Retain ensures that all items get archived in all other email systems.

How Retain Works with GroupWise Retention Services

GroupWise has a feature that can be enabled in its GroupWise Administration option called Retention Services.

When enabled, GroupWise prevents a user from emptying an item from Trash that has not yet been confirmed to have been archived. The way it does this is through a date/time field in each user database called the "digest retention time". It relies on third party archiving solutions like Retain to set that date/time, but GroupWise is the one that enforces it when set. What this does is it prevents any item newer than the date/time set in the "digest retention time" field from being emptied from Trash. This "digest retention time" is known in Retain as the "retention flag".

When Retain runs an archive job on a mailbox, it sets the digest retention time to the date/time of the newest/latest message it archived. However, if an error occurs on any item during that job which prevents Retain from archiving it or its attachment, Retain sets the digest retention time in the GroupWise user database for that mailbox to the date/time of the item that could not be archived due to an error.

And, even though Retain encounters an error on an item and cannot archive it, it moves beyond that item and continues to archive all other mailbox items; however, again, it doesn't advance the retention flag past the date/time of the FIRST error it encountered. Thus, when the next archive job gets run on that mailbox, Retain checks the item store time set in its database of the user and uses that date/time as its starting time for the new job, minus one hour.

Example: If today is September 17, 2014 but an item in the previous job produced an error, could not be archived because of that error, and had a delivered date/time of September 15, 2014 09:15, then when today's job runs, it asks GroupWise for all items beginning with September 15, 2014 08:15 and on.

Now let's say that a month has passed and the problematic mail message has not been properly dealt with and we run a job. Even though Retain may have archived all items in the user's mailbox up to - let's say October 15th - it still starts the query with the item store time of September 15, 2014 08:15 because it could not advance the retention flag. If it were to do so, then the problem message would never get archived because Retain starts the query for items beginning with the digest retention time. Thus, if Retain were to advance the flag to the date/time of the newest/latest item it archived, then the problematic message would fail to fit within the query range and GroupWise would never send it to Retain.

How Retain Works with Exchange and Office 365

These email systems do not have a built-in retention service similar to GroupWise, there is no "digest retention time" field in any of their mail system databases that Retain can use; thus, Retain uses its own field in the "retain" database to keep track of its job starting point. This "item store flag" works just like the "retention flag" with GroupWise jobs. That date/time gets set to the date/time of the newest/latest item archived for a given mailbox; or, if an error(s) occurred during a job, the item store flag gets set to the date/time of the first item that had an error. That way, when the next archive job runs, it starts with the date/time of the item store flag, ensuring that Retain tracks the item until it is properly archived. However, it is important to note that not advancing the item store flag does not prevent the user from emptying the item from their Trash in these email systems because they do not have a retention feature similar to GroupWise.

Placing a Hold Prevents Loss of Unarchived Messages

To prevent items from being deleted from Exchange/O365 a hold must be placed on the mailboxes. This can be an In-Place or Litigation hold. When a user deletes a message from Outlook the message is moved to the Trash, the user can then empty the trash. Exchange/O365 then moves the message to a Recoverable Items folder for 14 days before removing it from disk. However, a user can right-click on the trash and attempt to recover a deleted item, and at this point can purge an item immediately to remove it completely. This may be against your data retention policy, so to prevent the deletion, a hold then moves the item to the hidden Purged folder, where the user cannot remove it but Retain can still archive it.

Journaling Mailbox, an Alternative to the Item Store Flag but Not Recommended

Alternatively, a journaling mailbox may be used on On-Premise Exchange. When a journaling mailbox is set up in Exchange, it can be configured in a way that redirects a copy of each message that is either sent or received throughout the entire mail system into they journaling mailbox. Retain can be configured to include the journaling mailbox in its archive job. Thus, even if a user empties an item from Trash, a copy of that item already exists in the journaling mailbox and remains in that mailbox until it is archived by Retain. If configured properly, Retain removes that item from the journaling mailbox upon successfully archiving it. Items emptied from a user's Exchange mailbox but archived from the journaling mailbox do not appear in the user's Retain mailbox; however, they are searchable using the Retain search feature.

Because of the fact that duplicates of all email messages system wide get placed in the journaling mailbox, it can fill up fast. For this reason, we recommend that you not use the journaling mailbox feature and go with the Recoverable Items feature instead. If the journaling mailbox gets too big, Exchange is no longer able to serve the mailbox. Thus, when Retain tries to run an archive job against it, it fails because Exchange never responds back. This is why it is no longer recommended.

How Retain Works with Gmail

Gmail does not have retention services, by default. That requires the purchase of their Vault service.

7 Retain Unified Archiving

Retain software versions are incremented as follows:

`major-version.minor-version.service-pack.patch-release`

Each number in the version string is 1 or 2 digits (0-99).

Examples include (in chronological order):

- **4.0** The initial release of Retain 4
- **4.8** The eighth minor- version release of Retain 4
- **4.8.0.1** The first patch release for Retain 4.8
- **4.8.1** The first service pack release for Retain 4.8

Patches and service packs are generally developed for the current version only.