opentext™

OpenText Retain CE 25.2

Planning

MAY 2025

Legal Notice

Copyright 2023-2025 Open Text

The only warranties for products and services of Open Text and its affiliates and licensors (“Open Text”) are as may be set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Open Text shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Contents

Preface 7
About Retain . . . 7
About This Guide . . . 7
For Documentation Feedback . . . 7
To Join the Retain Idea Exchange . . . 7
For Additional Documentation . . . 7
To Contact Technical Support . . . 7
For Sales . . . 7
For Professional Services . . . 7

1 Are You Creating a New Retain System? 9

2 Are You New to an Existing Retain System? 11

3 Retain Planning and Design Best Practices 13

Retain Architecture . . . 13
Server . . . 13
Worker . . . 14
Indexer . . . 14
Database . . . 14
Retain is Modular and Flexible . . . 14
All-in-One Systems . . . 14
Medium to Large Systems . . . 15
Worker Locations . . . 15
Post Office Servers . . . 16
A Separate Worker Server . . . 16
With the Retain Server . . . 17

4 System Requirements 19

VMware Support . . . 19
Retain Component Locations Dictate Resource Needs . . . 19
The Number of Users Affects Resource Needs . . . 19
Minimum Retain System Requirements . . . 19
RAM Requirements . . . 20
Single-Server RAM . . . 21
Separate Database Server RAM . . . 21
Separate Database Server and Retain Server . . . 21
Securing the Database Connection . . . 21
Software installed by Retain . . . 22
Linux Server . . . 22
Windows Server . . . 22
SQL Database . . . 23
Database Planning and Administration . . . 23
Database Platforms . . . 23

Browsers 23

5 Hardware Planning 25

Network Bandwidth 25
CPU Requirements 25
Planning for Disk Storage 25
Planning Your Archive Size - Archive Files (BLOBs) 26
Database Size 27
Choose XFS as the File System on Linux 27
Disk Options 27
Data Partitioning 29
Disk Performance 30
RAID Considerations 30
SAN / NAS Considerations 30
Measuring Disk Performance 31
RAM 31
Concepts 31
General Guidelines 32
Tomcat Memory Configuration 33
Database Memory Configuration 33
Virtual Memory 33
VM Configuration 33

6 Storage Planning 35

Introduction to Storage Requirements 35
Considerations for Storage Requirements 36
Operating System 36
Retain Program 36
Archive Storage Area 36
Indexes 37
Database 37
Logs 37
Storage Minimums 38
Index Size During Migration 38
Changing the Log File Location 39
Linux 39
Windows 39
About Storage Paths 40
Setting Custom Storage Path Rights on Linux 40

7 Tomcat Memory Requirements 43

General Recommendations 43
Tomcat Configuration Parameters 44
Checking memory 45

8 Ports Used 47

Retain Server 47
Retain Worker 48
Retain Message Router 48

Retain Reporting & Monitoring Server 48
Retain Stubbing Server 49
From RSM WAN IP to Untrusted, all TCP/UDP ports. 49
Database Management Systems 49
Blackberry 49
Exchange 50
Google Apps Module 50
GroupWise. 50
Mobile 50
Office 365 50
CellTrust Secureline. 51

9 Messaging Systems Supported for Retain Archiving 53

10 Getting Started with Retain Archiving 55

First, Plan Your Retain Archiving System 55
Next, Match Your Messaging Systems to Retain's Archiving methods 55
Where to Go from Here 57

11 System File Locations 59

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 is the planning documentation for Retain Unified Archiving version.

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 official website and join the conversations happening inside the Retain Idea Exchange.

For Additional Documentation

See the OpenText Knowledge Base website.

To Contact Technical Support

Browse to OpenText Support Page and begin typing *Retain Unified Archiving* in the **Search for a Product** field.

For Sales

Go to the OpenText Contact page.

For Professional Services

Contact OpenText Professional Services by calling (877) 772-4450.

1 Are You Creating a New Retain System?

If you are creating a new Retain system, you should:

- Read this section.
- Get a copy of your company's retention policy, so that you can design your Retain system to fulfill that policy.
- Review Chapter 3, "Retain Planning and Design Best Practices," on page 13.
- Read about the modules required for your messaging systems and note their requirements. See "Modules Overview" in *How Retain Works*.
- Complete the install checklist. See "Retain Deployment Checklist" in the *Installation and Upgrade* guide.
- Create a plan for backing up Retain. See "Backing Up Retain" in the *Installation and Upgrade* guide.
- Confirm that you are an authorized user for your company on the OpenText Customer Portal website. This lets you create service requests, get license keys, and perform other support functions.

2 Are You New to an Existing Retain System?

If you are a new Retain administrator, you should:

- Read this Overview section.
- Review the Best Practices section. See Chapter 3, “Retain Planning and Design Best Practices,” on page 13.
- Know how to access Retain’s management interface. See “Retain’s Browser-based Management Tool” in *Installation and Upgrade*.
- Read “Using Retain’s Archives” in the *User Guide* so that you can find data.
- Read about the modules that are installed on your system. See “Modules Overview” in *How Retain Works*.
- Check your license status. See “Licensing” in *Installation and Upgrade*.
- Make sure that your organization is backing up Retain. See “Backing Up Retain” in *Installation and Upgrade*.
- Read about ongoing maintenance requirements in “General Maintenance” in *Installation and Upgrade*.
- Learn how jobs work as described in “Understanding Retain Jobs” in *Archiving Guide*.
- Get and study a copy of your company’s retention policy so you can confirm Retain is doing what it is supposed to.
- Review Chapter 3, “Retain Planning and Design Best Practices,” on page 13.
- Confirm that you are an authorized user for your company on the OpenText Customer Portal website. This lets you create service requests, get license keys, and perform other support functions.

3 Retain Planning and Design Best Practices

Use and apply the concepts and principles in the following sections as they apply to your needs.

- "Retain Architecture" on page 13
- "Retain is Modular and Flexible" on page 14
- "Worker Locations" on page 15

Retain Architecture

Retain can run on

- Stand-alone server hardware

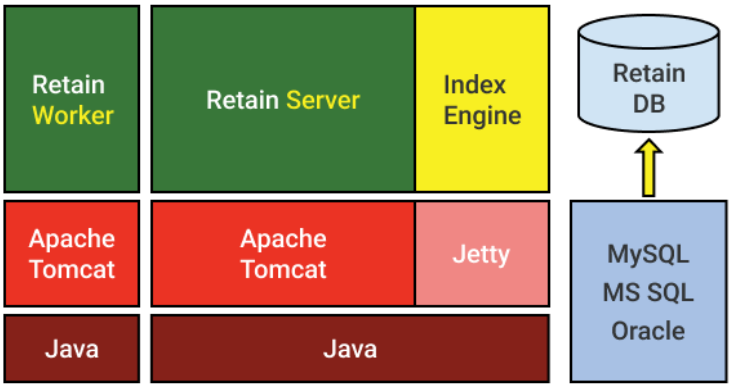
  Or

- A Virtual Machine hosted on a Windows or SuSE Linux supported hypervisor.

  This is the best-practice recommendation for backup purposes and flexibility.

Retain must have these four components:

- Server
- Worker
- Indexer
- Database



Server

This is where the archive system is configured and maintained. It coordinates and directs the storing, indexing, searching, and reading of archived items.

Worker

Workers interface with the messaging host/mail servers that contain the messages you are archiving. Workers retrieve the messages and hand them to the Retain Server.

IMPORTANT: A worker can handle only one job at a time. You can queue up more than one job for a worker, but make sure the worker can complete all of its jobs in less than 24 hours.

Indexer

This indexes each word (and some phrases) in all archived messages and attachments.

When someone searches in Retain, the indexer returns the list of hits by retrieving metadata from the database. (Some have incorrectly assumed that the database returns the list of hits.)

Retain's most memory-intensive process is indexing, not running the database.

Keep this in mind when dividing memory between Tomcat/the indexer and the database.

Database

This stores most of the Retain configuration and all the message metadata (subject, sender, recipients, links to attachments, indexed state of messages, folder-context of the message, and so on).

Everything displayed in a Retain mailbox is metadata that is retrieved through the index from the database, not from the message archive.

Retain is Modular and Flexible

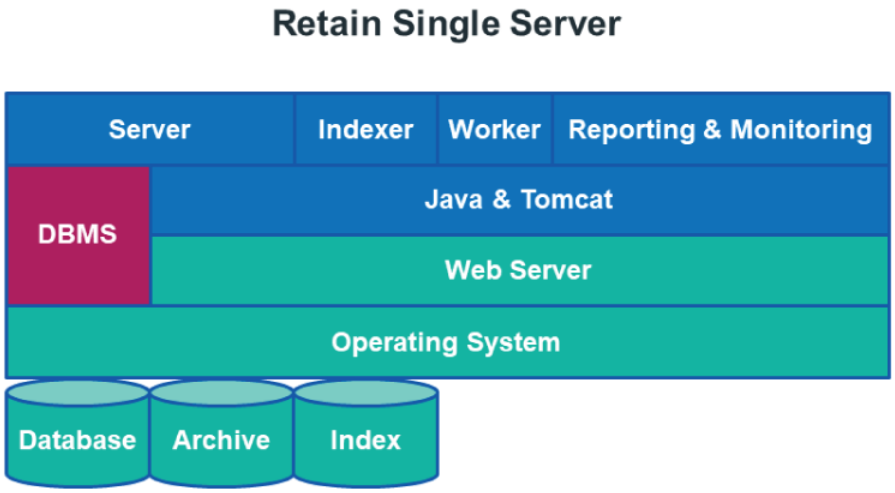
Because Retain is modular, component software can be installed on different servers, as the needs of your system dictate.

- "All-in-One Systems" on page 14
- "Medium to Large Systems" on page 15

All-in-One Systems

This is recommended for:

- Proof-of-concept systems.
- Small systems that aren't expected to grow beyond a few hundred users.

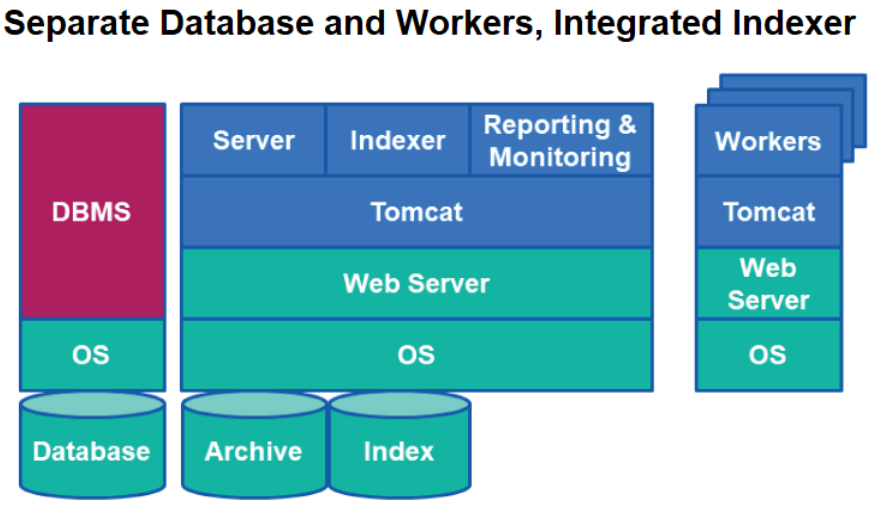


Medium to Large Systems

This is recommended if

- Your organization already has a dedicated database server.
- Your users are assigned to several different post offices and/or messaging systems.

For medium to large systems, OpenText recommends assigning one worker agent per post office/ messaging system (as applicable).



Worker Locations

Dredging large email systems can take a long time, especially when it involves multiple mail servers.

Because Worker Agents do the heavy lifting when it comes to dredging, their placement is critical to overall system efficiency.

You can install Worker Agents in three locations as your deployment needs dictate.

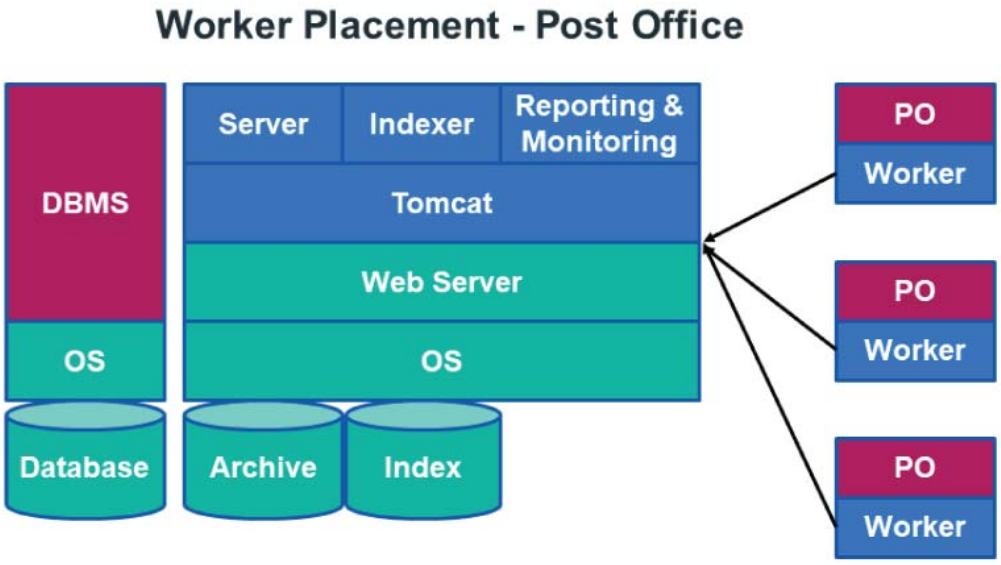
By order of recommendation, these locations are:

- “Post Office Servers” on page 16
- “A Separate Worker Server” on page 16
- “With the Retain Server” on page 17

Post Office Servers

OpenText recommends installing one worker on each mail server wherever possible because

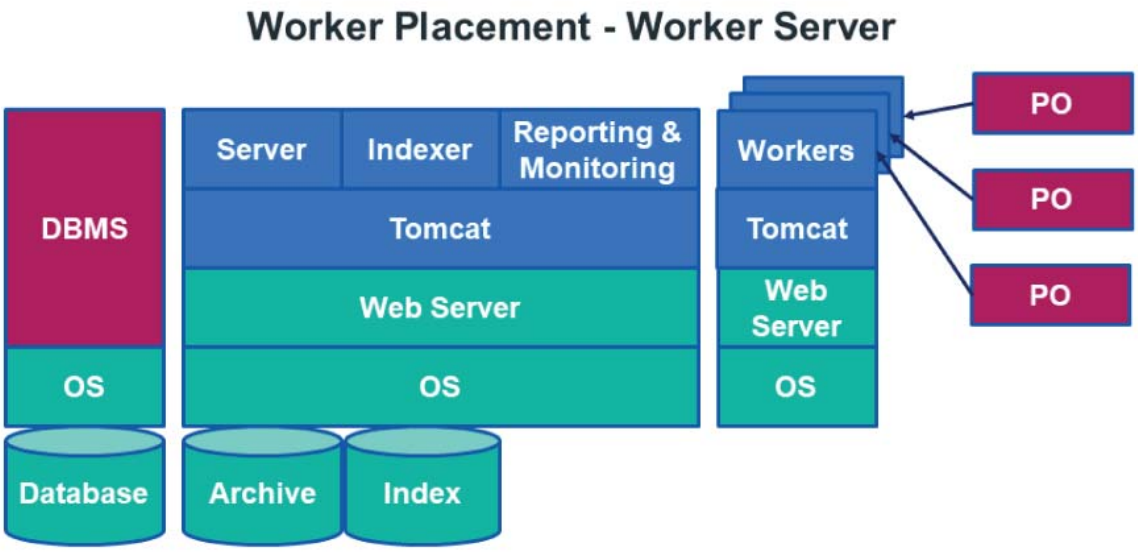
- Item retrieval happens quickly and efficiently within the server itself.
- Processing happens locally.
- Only new messages are transmitted over the network to the datastore, conserving considerable bandwidth. See “How Archive Jobs Work” in the *How Retain Works* guide.



A Separate Worker Server

If installing on post office servers is not an option for whatever reason, you can install up to 10 Worker Agents on a dedicated Worker Server.

Keep in mind that the network must have sufficient bandwidth to handle the load of all items that require processing, traveling from the post office servers to their respective workers. See “How Archive Jobs Work” in the *How Retain Works* guide.



With the Retain Server

For troubleshooting purposes, OpenText recommends always installing one worker on the Retain server.

However, having this Worker Agent function as the system worker is only recommended for small, proof-of-concept, all-in-one deployments. This configuration would rarely, if ever, be an effective solution for actual production workloads.

4 System Requirements

First decide where everything should be installed. Once the support and core components are installed, using Retain is the same no matter what platform it is installed on.

- "VMware Support" on page 19
- "Retain Component Locations Dictate Resource Needs" on page 19
- "The Number of Users Affects Resource Needs" on page 19
- "Minimum Retain System Requirements" on page 19
- "Single-Server RAM" on page 21
- "Separate Database Server RAM" on page 21
- "Separate Database Server and Retain Server" on page 21
- "Securing the Database Connection" on page 21
- "Software installed by Retain" on page 22
- "Linux Server" on page 22
- "Windows Server" on page 22
- "SQL Database" on page 23
- "Browsers" on page 23

VMware Support

VMWare ESX and ESXi are supported. The same disk space, RAM, and other requirements apply as for physical hardware deployments.

Retain Component Locations Dictate Resource Needs

Retain components can be deployed in various configurations, which can impact the requirements of the Retain server, indexing engine, database server, and so on.

The Number of Users Affects Resource Needs

Ultimately the number of active users in the system determines the resource needs of the system. See "Retain Planning and Design Best Practices" on page 13 for details.

Minimum Retain System Requirements

- **Processor:** Dual-Core 3 GHz or better Processor

NOTE: 8 cores provide optimal performance during initial archiving and re-indexing processes

- **Web server:** Apache 2.4.x or IIS, installed and configured by a system administrator
- **Database:** Supported SQL Database, installed and configured by a system or database administrator
- **RAM:** Retain Server alone requires 8 GB RAM. Adding a worker increases the requirement to 12 GB RAM. The installation aborts if less is detected in either case. At least 16 GB is recommended.
- **Storage Space:** As much storage space as messaging system requires. See "Storage Planning" on page 35
- **Swap Space (SLES/Linux):** At least 10 GB.
- **Standalone Worker server:** 4 GB RAM for server and one worker. 2 GB RAM for each additional worker.
- **Standalone Reporting and Monitoring server:** 8 GB RAM
- **Standalone Router:** 8 GB RAM

RAM Requirements

Memory needs vary, depending on the load and use of the Retain system. However, Java automatically tunes for the memory that is available on the server. Tuning depends on what is selected for installation.

The Retain Install Automatically Allocates RAM

The installation detects the amount of RAM and automatically assigns memory values as shown

Less than 16 GB system memory:

- 8 GB: Server and Worker
- 6 GB: Server only

Less than 20 GB system memory:

- 12 GB: Server and Worker
- 10 GB: Server only

More than 20 GB system memory:

- 16 GB: Server and Worker
- 14 GB: Server only

In all cases where only the Worker is installed, the memory is tuned to 3 GB.

Single-Server RAM

Assuming that the Retain Server, Retain Worker, Database, and Index engine are housed on the same machine and archiving less than 500 active mailboxes.

The Retain Message Router should be installed on a dedicated server in the DMZ.

- 16-24 GB RAM

Separate Database Server RAM

Assuming that the Retain Server, Retain Worker, and Indexing engine are on one server, and the Database is on its own server.

500-5,000 active mailboxes:

- 12-16 GB RAM

5,000-30,000 active mailboxes:

- 16-32 GB RAM

Separate Database Server and Retain Server

50,000 Active mailboxes

- 16+ GB RAM for Retain Server (depending on environment)
- RAM as required for Database on separate server
- Retain Workers installed remotely on each post office in the message system
- Disk Space is dependent on the amount of email being archived and length of retention time. It is recommended to estimate two years of storage and adding additional space as needed. Storage speed is the determining factor for performance.
- SAN and local storage preferred.
- NAS is not supported.

Securing the Database Connection

OpenText recommends that you consider securing the connection between Retain and the database server.

To prepare for securing the connection, make sure that you have identified the certificate files described below and made them available to install on the Retain server while running the initial configuration wizard:

- All the supported database types (MariaDB, MySQL, Oracle, PostgreSQL, and MSSQL) support secure database connections with Retain. We support certificates from industry-recognized Certificate Authorities and Self-signed certificates.
- For the Retain server, we support industry-recognized CA issued certificates, and Self-signed certificates.

- Following are the supported and certified files formats:
  - MSSQL - `.pfx`
  - Postgres - `.pem and .crt`
  - MariaDB -`.pem` and `.crt`
  - MySQL - `.pem` and `.crt`
  - Oracle - `.p12`

Initial setup instructions are in the *Installation and Upgrade* guide, starting with "Message Store Database."

After the initial setup, you manage the certificates and keys using the "SSL Certificates Management Tab", as documented in the *Configuration and Administration* guide.

Software installed by Retain

Retain downloads and installs

- Java Development Kit (JDK) 17 64-bit
- Tomcat 10, the Java webserver that sits on top of Apache or IIS.

  Installation on an existing Retain system running a standard Tomcat installation removes the current Tomcat connection and installs the embedded Tomcat that comes with Retain.

Linux Server

- SUSE Linux Enterprise Server 15 SP6 (64-bit)
- SUSE Linux Enterprise Server 15 SP5 (64-bit)
- SUSE Linux Enterprise Server 12 SP5 (64-bit)
- Open Enterprise Server 25.1 (64-bit)
- Open Enterprise Server 24.4 (64-bit)
- Open Enterprise Server 23.4 (64-bit)

IMPORTANT: Adequate disk space, by itself, is not sufficient because Retain stores many small files, especially when archiving mobile message data.

Only the XFS or EXT4 file systems can supply sufficient inodes to accommodate the large number of files.

Windows Server

- MS Windows Server 2025
- MS Windows Server 2022
- MS Windows Server 2019
- MS Windows Server 2016

SQL Database

- "Database Planning and Administration" on page 23
- "Database Platforms" on page 23

Database Planning and Administration

IMPORTANT: The basic information included in "Creating and Maintaining an SQL Database" in *Installation and Upgrade* is only informational and outlines Retain's database requirements at a high level.

OpenText does not provide support for database administration for legal and other reasons.

Your organization is entirely responsible for database installation, configuration, tuning, and maintenance. You should always enlist the help of a qualified database administrator for database administration tasks.

Database Platforms

- MariaDB 10.*x* and 11.x
- Microsoft SQL Server 2022, 2019, 2017, and 2016
- MySQL 8.0+
- Oracle 23ai, 21c, 19c
- PostgreSQL 16, 15, and 14

Browsers

- Firefox 67 or later
- Google Chrome 74 or later
- Microsoft Edge 121 or later

5 Hardware Planning

There are four major considerations you need to take into account when designing the hardware for a Retain system:

- "Network Bandwidth" on page 25
- "CPU Requirements" on page 25
- "Planning for Disk Storage" on page 25
- "RAM" on page 31
- "VM Configuration" on page 33

Network Bandwidth

The Worker queries your messaging system for messages and receives all of them. However, not all items are subsequently sent to the Retain Server.

If the link between the Worker and the messaging system is slow, consider placing the Worker on the messaging system's server or on a server that has a fast link to the messaging system.

The downside to this strategy is software updates.

When upgrading Retain software, you must update each Worker. Workers running on the Retain Server or on a separate server are upgraded together.

CPU Requirements

Retain is multi-threaded and can make use of multiple CPU cores. The base server uses 4 threads, and the indexer starts with 4 threads.

It is recommend to have 25 indexing threads for every 2 CPU cores. A system with 16 cores is tested with 200 threads. It never maxed out the CPU, which stayed around 75% utilization while indexing. The bottleneck will be disk IO once you hit 150 to 200 threads no matter how many CPU cores you have.

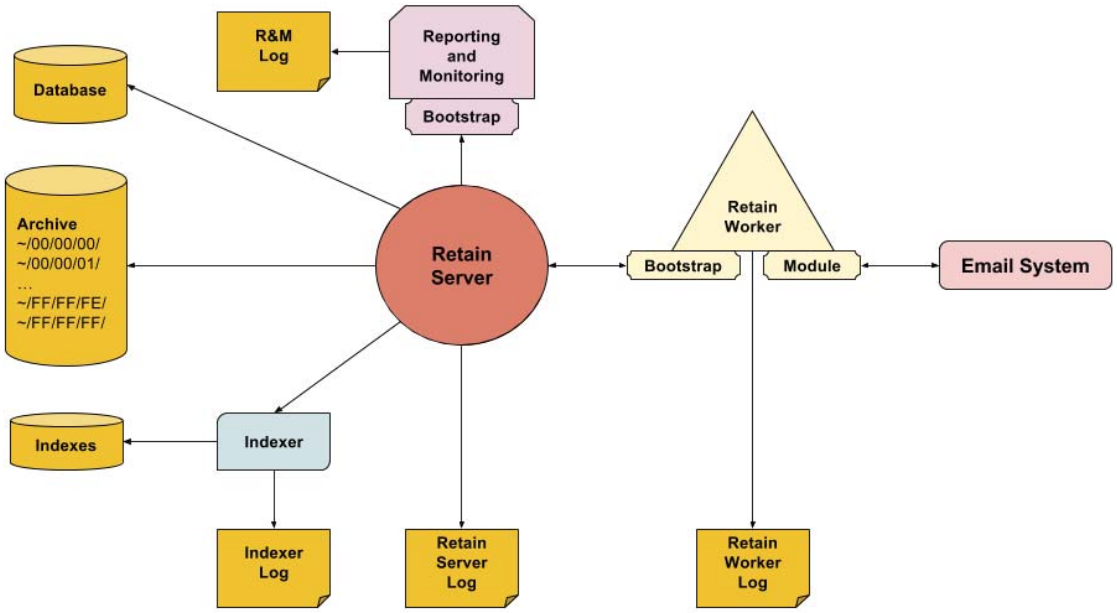
Planning for Disk Storage

If not monitored, Retain can completely fill its allocated archive storage.

Although Retain warns of disk-full conditions, you are responsible to keep the storage from filling up completely.

Once storage is full, recovery is difficult because server performance is heavily impacted.

It is critical that you design your system so that you can easily add storage as the system grows.



- “Planning Your Archive Size - Archive Files (BLOBs)” on page 26
- “Database Size” on page 27
- “Choose XFS as the File System on Linux” on page 27
- “Disk Options” on page 27
- “Data Partitioning” on page 29
- “Disk Performance” on page 30
- “RAID Considerations” on page 30
- “SAN / NAS Considerations” on page 30
- “Measuring Disk Performance” on page 31

Retain’s success depends on a robust storage design.

Install the OS on its own partition so that it’s easier to recover from a disk-full condition.

Make sure you have a comprehensive backup strategy for Retain. See “Backing Up Retain” in *Installation and Upgrade*.

Planning Your Archive Size - Archive Files (BLOBs)

As you begin planning your Retain archive, we recommend that you start with the current size of your post offices and other systems, then multiply that by your system’s yearly growth rate and add that amount to cover at least one year, it not two.

It isn’t possible to predict how much archive space requirements will increase over time, but at least this sets a good starting point for your initial archive and growth in the near term.

If you have a virtualized environment, you can allocate more space than you think you will be used and thin provision the disks.

Retain archiving is designed so that only one copy of a message or attachment is archived no matter how many users receive it, or which post office they belong to.

Retain lets you expire and delete messages from the archive after a specified time period.

Database Size

For cloud deployments, we typically set the db partition to 500 GB and go from there.

If a partition runs low on disk space at any point, support can direct you on the proper steps to move the data to another partition if necessary.

The numbers provided in the following table are representations of three different systems. Two customers with the same number of messages in their system may have vastly different database sizes due to the difference in the message metadata.

For example, Customer A may have short distribution lists while Customer B has a lot of emails with hundreds if not thousands of recipients associated with the messages. The purpose of providing sample data is to illustrate differences.

Example Systems	Deployment A	Deployment B	Deployment C
Message Count	104,976,966	18,261,383	2,699,654
Archive Size	5.3 TB	1 TB	115 GB
File Size per Message in the archive	4.54 KB	4.71 KB	6.21 KB
Database Size	455 GB	82 GB	16 GB
File Size per Message in the database	56.27 KB	64.02 KB	45.06 KB

Choose XFS as the File System on Linux

OpenText recommends choosing XFS for Linux servers because it creates iNodes dynamically and performs well.

OpenText does not recommend ReiserFS (poor performance with Retain), or Ext3 (iNode inflexibility).

Disk Options

- “Physical ("bare metal") Server” on page 28
- “VM Guest on Host With Local Disks” on page 28
- “NAS or SAN” on page 28
- “Recommendations” on page 28

Retain archive jobs are disk-I/O intensive and includes:

- Storing message content in the archive

- Indexing each message
- Updating the database with each message’s metadata
- Updating various logs continually

In light of this, here are a few recommendations.

Physical ("bare metal") Server

Physical servers have their own locally attached disks. If there is just one disk, then disk I/O contention negatively impacts performance, especially while jobs are running.

VM Guest on Host With Local Disks

If your VM host has only local disks (NAS or SAN),make sure that you create multiple disks and that each one is on a different datastore if possible.

NAS or SAN

This could be physical server where the storage is mounted/mapped to a NAS or SAN; or, this could be a VM guest where:

- The VM guest itself is stored on a NAS/SAN; thus, the VM guest's "local disks" are also sitting on a NAS/SAN; or,
- The VM guest itself is stored on the hosts local disks but the "local disks" of the VM guest are on datastores residing on a NAS/SAN; or,
- The VM guest is mounting volumes stored on a NAS/SAN.

If the Retain storage is on a NAS/SAN and if the volumes are expandable on the fly, there are so many configurations that recommendations aren’t possible, except to understand what Retain is trying to do and then see what can be done on the hardware end to facilitate best performance.

If it is a NAS/SAN, consider the pipe speed to the storage: 1 gigabit/sec is very slow. On top of that, consider how many disks are in the array, their RAID configuration, and the speed of the disks themselves.

Recommendations

If all the Retain storage is located on the same volume and you run out of space, Retain provides the ability to create additional storage volumes for the archive files. After an additional logical storage volume is created within Retain, all archive files go to the new location.

However, the indexes continue to grow and Retain doesn't have the ability to partition indexes. Some customers have run out of disk space, created new logical storage partitions that point to another volume, but then run into problems with their archive jobs because they are still out of disk space for the indexes. Thus, for logical reasons, you want to have your archive files on a separate volume to begin with, unless the volume containing the archive is expandable on demand.

If it makes sense to do so (based on all the concepts previously discussed), you'll want to separate your archive files from your indexes and from your database, which means two to three other partitions on your Retain Server in addition to your OS partition. If your database is on a separate server from Retain, then only two other partitions are needed; otherwise, you'll want three additional partitions.

Data Partitioning

We recommend dividing up your storage directories onto separate disks, so beyond the OS disk there should be:

- Disk 1: Archive
- Disk 2: Index (250G start). For best search performance, consider making this a solid state drive.
- Disk 3: Logs, xml, ebdb, export, backup, and license (150 - 200G)
- Disk 4: Database (if on-board)

Disk 2 should be expandable and you'll want to give it room for the indexes to grow; but, if you cannot do that, then when it runs out of space, you'll simply need to move your index files to another volume with more disk space in the future. For disk 2 - as mentioned previously - you may want to consider an SSD, as that would increase the search performance.

If disk 1 and disk 2 can literally be on different physical disks, then you get some performance gains from that because an archive job writes simultaneously to the archive directory, the index directory, and to the database. If each of those are on different physical disks, then this eliminates disk contention bottlenecks. Smaller systems may not need to be concerned with performance while larger systems that have archive jobs running for hours may want the performance gains.

Using disk 3 for logs is especially helpful for larger systems. If you have 6 Workers averaging 5 - 10 messages per second, expect a RetainServer log of around 60G unzipped. Plan for 150 - 200G for your logs directory. For the initial archive job, the rule of thumb is 10G per day per Worker. If you do not use a third disk, then the logs are written on the OS partition and that could spell trouble. Also, if users access their archives often and perform PDF exports, that can grow as well. The xml, ebdb, and license directories are pretty much static with minimal to no growth. The backup directory is a backup of the index directory and other important items. However, if the disk begins to run out of room, you can copy this data over to a larger disk at some future time and point retain to that new disk.

Finally, if your database is on the Retain Server, you'll want a third or fourth disk for it (depending on whether you decide to dedicate a disk for your Retain logs).

If performance is an issue, you should place all three partitions on different physical disks (or at least a NAS/SAN with many disks that it can swipe across). You should also put the indexes and the database on high speed drives. Your archive directory does not need the performance and can be on less expensive disk media.

Make sure to set the permissions of the new disks correctly in Linux, or the installation fails.

Disk Performance

Knowing that disk I/O is the top issue with archive job performance, it is best to plan out your disk storage accordingly.

Storage design and disk I/O has everything to do with Retain performance as archive jobs are I/O intensive. You have the following processes writing to disk simultaneously:

- The indexer to the [storage path]/index
- The database (if on the Retain server)
- The Retain Server to [storage path]/archive
- The Retain Server to the logs directory:
  - Linux: /var/logs/retain-tomcat
  - Windows: [drive]:\Program Files\Beginfinite\Retain\Tomcat\logs

With all of that disk activity, if a single drive is having to handle all of it, then you can see that the performance bottleneck would be disk I/O. However, many modern disk systems involve multiple disks using (i.e., RAID 5 or RAID 10) that write the data across multiple disks. The more disks involved, the more you spread the load and the faster the overall performance. You also have a difference in drives (SATA/SAS/SSD). In those cases, you now are looking at whether the disks are local to the server or in a SAN/NAS.

RAID Considerations

Let's say your server employs RAID 5, which provides better redundancy than, say, RAID 10. If there were 4 disks. As you know, RAID 5 uses an extra parity bit that consumes an entire disk, which leaves it with 3 drives on which to stripe across. If one of those drives becomes unavailable, that leaves you with 2. Striping across 2 or 3 drives doesn't lend for great speed, especially if the disks are lower-end SATA drives.

SAN / NAS Considerations

If on a SAN/NAS, now you are looking at the network link speed as well. You could have very fast drives, but if your link speed is 1 Gb/s, your bottleneck is going to be your link.

The 1 GB/s network link is slower than a SATA 2 or 3 connection (AKA SATA 3 Gb/s and SATA 6 Gb/s.) Your SATA 2 connection (which is now getting to be a pretty old standard) is 3x faster than a 1000 Mb/s network link (or 1 Gb/s network connection). A fast single HDD can saturate a 1 Gb/s connection but not quite a 3 Gb/s connection (SATA 2.0, or SATA 3 Gb/s) with a sequential read/ write. 7,200 RPM platter drives usually top out around 160-170 MB/s (or 1.28-1.36 Gb/s).

Measuring Disk Performance

It really comes down to IOPS. Here is a very simple IOPS calculator: http://www.thecloudcalculator.com/calculators/disk-raid-and-iops.html (http://www.thecloudcalculator.com/calculators/disk-raid-and-iops.html) or you can find one of your own.

So, it really comes down to you understanding your underlying disk storage. This article just gives food for thought. If you are running Retain on a VM guest server like most customers do, then you need to also understand your VM host and VM infrastructure. Is the Retain storage viewed by the server OS running on the VM guest as "local" storage? If so, what type of disk system is holding your VM's datastore? If it is not local storage but the server is connecting to external storage, then you need to take a look at the external system's configuration.

Bottom line: Disk I/O performance is key to Retain's performance and there are several areas to investigate where the bottlenecks could be.

In addition to partition considerations, make sure that your storage is reliable. NFS mounts can be problematic, so you may want to shy away from those. NSS volumes are not supported, so do not use them.

RAM

The amount of memory depends on the number of active mailboxes you are archiving, the mail volume, your underlying hardware, and how your Retain system is used.

Let’s discuss the concepts and general guidelines. In most instances, you should experiment with various memory configurations until you find what works best in your environment.

- “Concepts” on page 31
- “General Guidelines” on page 32
- “Tomcat Memory Configuration” on page 33
- “Database Memory Configuration” on page 33
- “Virtual Memory” on page 33

Concepts

Retain runs under Tomcat as shown at the beginning of this article and Tomcat runs on Java. The Retain Server uses the Java "heap" for its memory and the indexer uses the OS memory as well as virtual memory (see the Virtual Memory subsection below). For this reason, you should configure Tomcat/Java with the bare minimum to have it run in an acceptable fashion for you. If logins or Retain in general seems sluggish when in the mailbox or using the web admin tool, you may need more heap. The sweet spot for most systems with a single Worker installed on the local Retain server is 8 GB minimum (xms) and maximum (xmx). You want to leave as much RAM as possible for the Indexer, which uses non-heap RAM.

The amount of Java heap you set depends on the total RAM on your system and the number of Workers you install in addition to the default single Worker. As we grow in customer experience with Retain 4, we adjust this article's memory recommendations accordingly.

Right now, development has suggested 1 - 2GB per additional Worker beyond the 8 GB you normally would give to the Java heap for a system with a single Worker local to the Retain server; however, we've had a customer with 110 million messages with 7 Workers local to the Retain server get away with 8 - 10 GB of RAM, but that is really pushing it. They didn't run under that configuration for more than 24 hours, so we cannot tell whether it would have been successful in the long run.

The installer for Retain 4.0.1 and later tunes Tomcat/Java memory based on total RAM and which Retain components are installed. See the online manual's topic, "Tomcat Memory tuning" (note: that link goes to the 4.0.1 documentation, so if the link doesn't exist in the future, go to the online manual and find that topic). Again, as we learn more from customer experience, the installer's default RAM configuration is subject to change.

If you really want the fastest search performance, load it up with RAM, like 64GB or more. Systems with large numbers of messages (100 million or more) seem to be needing 64 GB of RAM or more. If you have a database system running on your Retain Server along with multiple local Workers, then those decrease the available RAM for the indexer, so you need to take that into account. The indexer wants to cache indexing data into RAM and memory access is much quicker than disk.

General Guidelines

All of this really depends on the priority you place on Retain performance. If a customer is only interested in getting data into Retain and it doesn't matter how long the archive jobs take (as long as they finish within a 24-hour timeframe) nor does the customer care how long it takes to search for messages (because they do not do it that often), then none of this matters.

The key test is how quickly tomcat shuts down and how much memory the OS is sending to swap. If tomcat is shutting down slowly, that's probably an indication that it has code in swap memory that it is having to call off of disk in order to close out. Reserving more memory for the OS should alleviate that problem; thus, reserve a minimum of 4G for the server OS right up front. On some systems, we have had to allocate more, on others, less. So, the key is to try different configurations on your system to see what makes the difference.

Once you have subtracted the OS memory from your total memory, give 2 - 4G of RAM to the database (if the database is on the same server; otherwise, the remainder can go to Tomcat). Note that Tomcat needs a minimum of 2G.

For small systems (1 - 250 mailboxes), 8G of RAM might deliver acceptable performance if that's all you can afford to allocate. Small Retain system can theoretically run on 4G, but performance is unacceptably low in most cases. You really should not go lower that 8G unless you are a very small business and have 0 - 50 mailboxes. You might even want to consider trying 12 to 16G and weigh the performance improvement against the cost. For some, it can make a big difference. For others, it might make no difference because the performance bottleneck is elsewhere.

For medium sized systems (250 - 750 mailboxes), 12 - 16G of RAM should be considered.

For larger systems, 16G should be considered a minimum. Many large systems range from 24 - 48G of RAM. The more mailboxes and mail volume, the more RAM you might consider giving your Retain server. But, again, we have to emphasize that every system is unique and RAM may not be the biggest performance factor for them.

Case in point: We have a customer with 700 users that found allocating 24G of RAM made a big difference. In another case, a customer that had 1,500 users needed only 12G. We have systems with thousands of mailboxes and those systems do benefit from increased memory allocation, but their needs vary.

Tomcat Memory Configuration

Tomcat memory is manually configured. The latest version of Retain sets it to 8G by default. It is an industry best practice to set the minimum and maximum memory values to the same value.

In Linux

You set the Tomcat memory parameters in a file called j2ee found at `/etc/opt/beginfinite/retain/tomcat`. See "Tomcat Memory Requirements" on page 43 for more detail. Tomcat must be restarted after configuring it.

In Windows

You can set Tomcat parameters by running Programs | Tomcat 10.1 | Configure Tomcat. Go to the "Java" tab to set them. Note, we also recommend setting the stack size to 256k (it defaults to 160k in Windows).

Database Memory Configuration

Since most organizations employing Oracle or MS SQL have someone designated as a database administrator (DBA), they typically understand memory configuration. What they need to know is that archiving speed and user mailbox browsing performance is affected by the amount of memory given to the Retain database.

Virtual Memory

If you have the available disk space, we recommend increasing the virtual memory to at least 50GB. In Linux, this is known as swap. In Windows, this is called the page file. Ideally, this swap or page file should be placed on a fast storage for performance reasons.

VM Configuration

VM (Virtual Machine) NIC Settings

We have found that using VMXNET3 for the network adapter in VMs helps performance.

Virtual Machine SnapShots

We have found that VM snapshots can reduce performance of the Retain Server. Keeping the number of snapshots to a minimum is highly recommended.

6 Storage Planning

- “Introduction to Storage Requirements” on page 35
- “Considerations for Storage Requirements” on page 36
- “Storage Minimums” on page 38
- “Index Size During Migration” on page 38
- “Changing the Log File Location” on page 39
- “About Storage Paths” on page 40
- “Setting Custom Storage Path Rights on Linux” on page 40

Introduction to Storage Requirements

It must be understood that no system storage requirement estimation can be expected to maintain or have any kind of reliable accuracy. Future mail use, litigation requirements, and compliance standards all may change and are unpredictable at best. OpenText bears no responsibility to accurately define or recommend storage needs for various messaging systems. Different messaging systems have different storage characteristics, and individual implementation renders general calculations invalid.

Keep in mind, however, that storage-system performance dictates mail view, indexing, and data performance. If the storage system is housed on slow hardware employing a file system designed for any other consideration than speed, (e.g. compact data storage), performance may be impacted. Consider what types of files are to be archived and access frequency when choosing the file system. For example, a Retain system dealing with mainly or only archived mobile data (SMS, pin, MMS, &etc.) performs best with different file system settings than a Retain system archiving a message system with large attachments.

For best results; pair current mail storage needs against projected future needs with the ability to easily add extra storage to the Retain system as needed. The ability to freely add additional storage space grants control and freedom over the messaging system and should be of paramount consideration. This practice is the only course which can be relied on with any confidence. Due to the challenges and circumstances involved with each different system, (and even certain versions of different systems), only individual consideration provides a reliable baseline for storage needs.

The simplest way to check current disk usage and storage requirement size, is to monitor disk space usage on the mail servers and create a projection for the near future for needs. As a rule of thumb, we have noticed that message systems increase by about 11% per day.

However, the different options and variables between messaging systems make disk storage estimations so unpredictable that anything other than specific system monitoring cannot determine real disk usage. It is best to create a storage system where additional space may be added as required when existing space is consumed.

In addition, Exchange 2010 has abandoned single-instance storage in favor of highly available performance, possibly causing multiple Exchange servers in the system to all have copies of the same data. Retain utilizes single-instance storage and may vastly decrease the storage size of a system that heavily utilizes this feature. Due to the differences between the storage and main messaging system, it is nearly impossible to establish a baseline for Retain storage needs. Retain may tremendously decrease the needed size to archive an Exchange 2010 system, or, depending on system size and implementation, it may not significantly decrease the needs of the current system. Though Retain requires additional space to continue archiving mail, the initial archive job will not exceed the size of the current messaging system.

Consulting with the Retain Sales representative offers the best tailored information for each system and each implementation of the different platforms available.

Considerations for Storage Requirements

- "Operating System" on page 36
- "Retain Program" on page 36
- "Archive Storage Area" on page 36
- "Indexes" on page 37
- "Database" on page 37
- "Logs" on page 37

Operating System

The OS receives a number of updates over the life of a Retain server, so that should be provisioned for.

Retain Program

The Retain program is updated often as well. This is often placed on the same volume as the OS.

Archive Storage Area

The Archive contains the item bodies and attachments. This is generally the largest percentage of the storage requirements on a Retain server. The bodies and attachments are stored as BLOB (Binary Large OBject) files in a folder structure that starts with /00/00/00 and grows to /FF/FF/FF for 16.7 million directories. BLOB files are stored only once, Retain implements a single-instance storage system, only a single copy of identical items is stored. BLOB files never change, they are only created or removed.

The files are accessed when a message is opened in the Retain Search Message interface or downloaded by Retain Publisher.

Indexes

The indexes allow for fast search of the data in Retain. When searching for items in Retain the indexes are used to return the results.

For best search performance, the indexes should be placed on a fast disk and optimized regularly.

The Index requires periodic optimization. This is set under *Server Configuration* | *Maintenance*. The Indexer requires as much free disk space as currently used index space for optimization. It requires three times as much if optimizing during an archive job.

IMPORTANT: Prior to Retain 4.11, the Zookeeper runtime folder was embedded in the index folder under the Retain's base path. However, from Retain 4.11 onwards, as the zookeeper service supports Solr7 and Solr8, the Zookeeper folder is decoupled from the index folder and is placed inside Retain's base path.

Solr stores configuration files in the runtime ZooKeeper directory to manage nodes, replica allocations and schemas.

With these changes in Retain 4.11, the Tomcat user should be the owner of the retain base directory, otherwise, the runtime creation of the Zookeeper directory within this base directory will fail which will prevent the Retain Server from starting.

To fetch the base path from ASConfig.cfg file, run the following command for Linux:

```
cat /opt/beginfinite/retain/RetainServer/WEB-INF/cfg/ASConfig.cfg |grep -i basepath

<basePath>/home/retain</basePath>
```

To set the ownership of a base path to `tomcat:www` execute the following command:

```
chown tomcat:www /home/retain
```

Database

The database contains the header information for each item. For example: Sender, recipient, date received, and so on as well as pointers to the message body and attached files stored in the archive. Each item may be quite small but may be replicated many times in the database if there are many recipients.

When browsing messages, the items in the database is being viewed. On larger systems, the database is often placed on a dedicated database server.

Logs

The logs track the actions that Retain takes. They can become quite large but are compressed at the end of each day and removed after 10 days, by default. This can be changed under *Server Configuration* | *Logging*. Logs can be moved to another volume as described below.

Storage Minimums

Retain is very disk intensive. Running out of disk space can be very challenging to recover from and can potentially result in data loss. To minimize the chances of this happening certain safeguards have been implemented.

- Retain issues warnings when there is less than 10 GB free space for the storage, index, and system volumes.
- If the Retain system reaches 5GB of free disk space, it enters maintenance mode.

These minimums are configurable, but it is not recommended to permanently change these settings:

1. Change to the configuration file directory, by default in:

   Linux: */opt/beginfinite/retain/RetainServer/WEB-INF/classes/config/*

   Windows: *C:\Program Files\Beginfinite\Retain\RetainServer\WEB-INF\classes\config*
2. Edit the *misc.properties* file
3. Change the following settings (in gigabytes) as desired, defaults being:

   discspace.warn.gb=10

   diskspace.error.gb=5

Index Size During Migration

When a migration begins, Retain creates a copy of the current index, thus temporarily doubling the space consumed by the index. After the migration completes, the old index can be removed and space reclaimed.

Also, the new index may be larger than the current index, due to the increased power and abilities of the new indexer. How much change there is, depends on the composition of the current archive, any limits set in indexing attachments, and how many attachments there are in the archive. If there are a lot of attachments, or very large attachments, then the increase in size is significantly larger than if there are small or a limited number of attachments in the archive. With no limits on indexing attachments and a lot of large attachments in the system, an increase of up to 3.5 times the current index size is possible. With limits included, the new index may be smaller after upgrading. For the upgrade and a worst-case scenario, ensure that up to 5 times the current index size is available before starting the index migration. (With a current index size of 5 GB, ensure that the volume has a minimum of 20 GB free before starting the migration.) The average increase is expected to be around 20%.

If there is insufficient space to perform the migration, Retain moves into maintenance mode until more space is provided. Retain 4 checks for free disk space on the storage, index, and system locations. If there is less than 20 GB free space, Retain sends a warning message every 6 hours. At 10 GB of free space left, Retain enters maintenance mode and all jobs are disabled. To exit maintenance mode, you must provide more free space on the volume.

NOTE: It is important to recognize that migration time and performance depend mostly on the performance of the storage system. External storage systems, such as appliances, may 'prioritize' seldom-used data to low performing storage and have a negative effect on migration performance.

In addition, file systems of existing storage largely dictate how fast indexing, migration, and message view can be performed. To ensure high performance, house the storage on performance hardware with a performance file system.

Changing the Log File Location

- "Linux" on page 39
- "Windows" on page 39

Linux

To change the location of the log files, create a symbolic link (somewhat synonymous with Windows shortcuts). Retain, by default, gives you the option during installation to store the logs at `/var/log/retain-tomcat` or at `/opt/beginfinite/retain/tomcat`. If stored at `/var/log/retain-tomcat`, then Retain creates a symbolic link at `/opt/beginfinite/retain/tomcat` directory called 'logs" that points to the /var/... location.

1. Create the directory in the location you wish to use.
2. Stop tomcat as described in "Starting and Stopping the Retain Server" in *Installation and Upgrade*.
3. Move the current logs to the new location: `mv /var/log/retain-tomcat/* /[path to new directory]`
4. Make tomcat the owner of the new directory path: `chown -R tomcat:tomcat /[path to new directory]`
5. Set the appropriate file *permissions for the tomcat user and group: `chmod -R 664 /[path to new directory]`

   * If you are moving them to a separate volume, the file permissions must be 774.
6. Create a symbolic link in the parent directory of the default logs directory and point it to the new location:
   a. Change to the /var/log directory
   b. Remove the current log directory: `rm -r retain-tomcat`
   c. Create a new symbolic link called retain-tomcat that points to your new log location: `ln -s [path to new directory] ./retain-tomcat`
7. Start tomcat.

Windows

1. Ensure no archive jobs are running and stop Tomcat.
2. Configure Tomcat's default log location.
   a. Click on Start.
   b. In the "Search programs and files" box, type: configure tomcat
   c. Click on Configure Tomcat
   d. Click on the Logging tab.

e. Type in the new log path using the standard Windows path utilizing backslashes "\" (i.e., d:\retain\logs) or browse to it by clicking on the button with "..." on it.

f. Click OK.

3. Make a backup copy of the existing *log4j.properties* file.
4. Edit the log4j.properties located at *[drive]:\Program Files\Beginfinite\Retain\[RetainServer, RetainWorker, RetainWorker1, &etc]\WEB-INF\classes*.
5. Do a search on *${catalina.base}/logs/* and replace it with *[desired path using forward slashes "/ "]/logs/ (i.e., D:/retain/logs/)*. An easy way to do this is to use the text editor's (i.e., Notepad") "Replace" function.
6. Repeat steps 2 - 4 for every log4j.properties file (server, worker, stub server, stats server). The only log that is created at the default log location is the localhost.[date].log, but it is a very small log.

About Storage Paths

Retain was designed to be very flexible in allowing you to choose where the data is stored. Choosing the right hardware and software storage combination is crucial. Storage itself is a function of the operating system and hardware you choose.

As storage needs approach enterprise levels, the need to specify your data storage locations becomes more critical. Some hardware and operating systems have a hard limit of 2 Terabytes, for instance. In situations like these, a SAN might be a good choice for consideration. Either way, it's imperative that you take the time to investigate your storage needs BEFORE you begin. Retain monitors the storage space available, and issues email warnings at 20 GB, and locks down Retain to a maintenance mode when the storage space drops to 10 GB. This is to avoid possible corruption to the storage and index which may occur when space runs out.

When choosing your storage system, bear in mind the following considerations:

- Storage size limits. Do you expect storage needs to ever exceed this?
- Ability to add additional storage capacity
- OS Version and kernel to support the storage system of your choice
- File system and partitioning
- Is a SAN right for you?
- Speed and reliability
- Backup/restore is your responsibility

Setting Custom Storage Path Rights on Linux

During the initial Retain Server configuration process, you must specify where archived data will be stored.

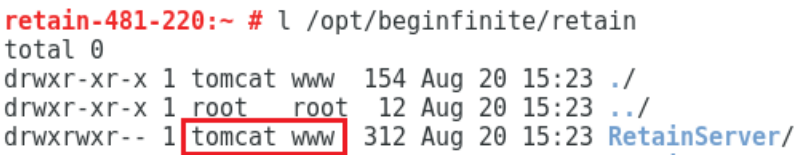
If you plan to specify a path other than `/var/opt/beginfinite/retain` (the default), then, before running the configuration wizard, you must grant the Retain Server read and write access permissions and ownership for the storage paths that you specify.

Do the following before running the Retain Server Configuration Wizard.

1. After installing the Retain server, determine the user and group that Retain uses to access the file system by opening a terminal prompt on the Retain server and entering the following command, which is a lower-case L:

   `l /opt/beginfinite/retain`

   In the following screenshot, Retain is running as `tomcat:www`.

   
   ```
   retain-481-220:~ # l /opt/beginfinite/retain
   total 0
   drwxr-xr-x 1 tomcat www  154 Aug 20 15:23 ./
   drwxr-xr-x 1 root   root  12 Aug 20 15:23 ../
   drwxrwxr-- 1 tomcat www 312 Aug 20 15:23 RetainServer/
   ```

2. If your customized storage locations don't already exist, you can create them using the `mkdir` command:

   `mkdir /mount-point/directory-structure`

   For example, we have previously created a mount-point for our data storage device on the Retain Server and named it

   `/datastore`

   We now create directories on the device for storing message attachments and index files. We also create a subdirectory under the `/attachments` directory for storing PDF files, as follows:

   ```
   mkdir /datastore/attachments
   mkdir /datastore/attachments/PDF
   mkdir /datastore/index
   ```

3. After you create your customized storage locations, make sure to grant ownership and access permissions at the top level of each directory structure you create, using the `chown` and `chmod` commands.

   ```
   chown –R user:group /mount-point/directory-structure
   chmod –R 744 /mount-point/directory-structure
   ```

   Continuing the example from Step 2 on page 41, we grant ownership and access permissions to the top-level directories as follows:

   ```
   chown –R tomcat:www /datastore/attachments
   chown -R tomcat:www /datastore/index
   chmod –R 744 /datastore/attachments
   chmod -R 744 /datastore/index
   ```

4. Finally, you should verify the changes you made to your new storage directories using:

   `l /mount-point`

   Continuing with our example:

```
retain-481-220:/ # l /datastore
total 0
drwxr-xr-x 5 root   root  54 Aug 21 13:04 ./
drwxr-xr-x 1 root   root 188 Aug 21 12:30 ../
drwxr--r-- 3 tomcat www   17 Aug 21 13:04 attachments/
drwxr--r-- 2 tomcat www    6 Aug 21 13:04 index/
retain-481-220:/ #
```

5 You can now specify your customized storage locations when you run the Configuration Wizard. See "Running the Retain Server Configuration Wizard" in *Installation and Upgrade*.

7 Tomcat Memory Requirements

- "General Recommendations" on page 43
- "Tomcat Configuration Parameters" on page 44
- "Checking memory" on page 45

General Recommendations

Retain automatically tunes on installation, depending on available system memory. Further tuning should not be necessary except in extreme conditions.

By default, Retain detects installed system memory and tunes the maximum memory value according to the following levels:

In general, a minimal install requires 4GB for the OS, the tomcat heap XMS setting (default 4GB), 2GB for Retain and any memory the database requires if installed.

Less than 16 GB system memory:

8 GB: Server and Worker

6 GB: Server only

Less than 20 GB system memory:

12 GB: Server and Worker

10 GB: Server only

More than 20 GB system memory:

16 GB: Server and Worker

14 GB: Server only

In all cases, when only the worker is installed the memory is automatically tuned to 3GB.

To manually tune the memory, carefully consider the following information.

Tomcat Configuration Parameters

There are three significant parameters:

1. Memory allocated upon Tomcat startup
    - Indicates how much memory is immediately allocated and reserved to Tomcat upon startup. This memory is used by Tomcat and is never available to the other server processes.
    - Typically, this is set to 50%-100% of the maximum memory parameter discusses below. It can be lower, but pre-allocating a sizable percentage of memory enhances performance and reduces memory fragmentation.
2. Maximum Memory available to Tomcat
    - If the memory usage grows beyond the startup allocation, Tomcat requisitions additional blocks of memory in chunks as needed up to this limit. It never returns this memory to the general server memory pool. (Although memory internally is freed and reused for Tomcat applications).
    - Tomcat guarantees to never exceed this memory allocation parameter. If Tomcat runs out of memory, it tries to reclaim unused memory via garbage collection. If this is insufficient, the web application doesn't receive the memory allocation. Unpredictable (but invariably unpleasant) results then occur.
    - This is the most critical parameter to tune, and is the value which is automatically tuned by the installer.
3. Stack Size
    - For each thread (which includes each and every concurrent user request), Tomcat allocates stack space.
    - This value is typically measured in KB, and defaults to 512KB.
    - It is allocated per concurrent users, and is in fact far too generous a number in general. 1000 users for example would take 500 MB of RAM just for stack space, before the program even allocates memory to run!
    - Generally, we recommend reducing this number to 160k. Increase conservatively, in 64 KB chunks, if you see out-of-stack-space errors in the logs.

The parameters are abbreviated in the Catalina Opts line as follows:

`Stack Size="-Xss<value>k"`

`Memory Heap="-Xms<value>g"`

`Maximum Memory=" -Xmx<value>g"`

On Linux: edit the following configuration file:

Retain 4.9.x or higher

`/etc/opt/beginfinite/retain/tomcat/j2ee`

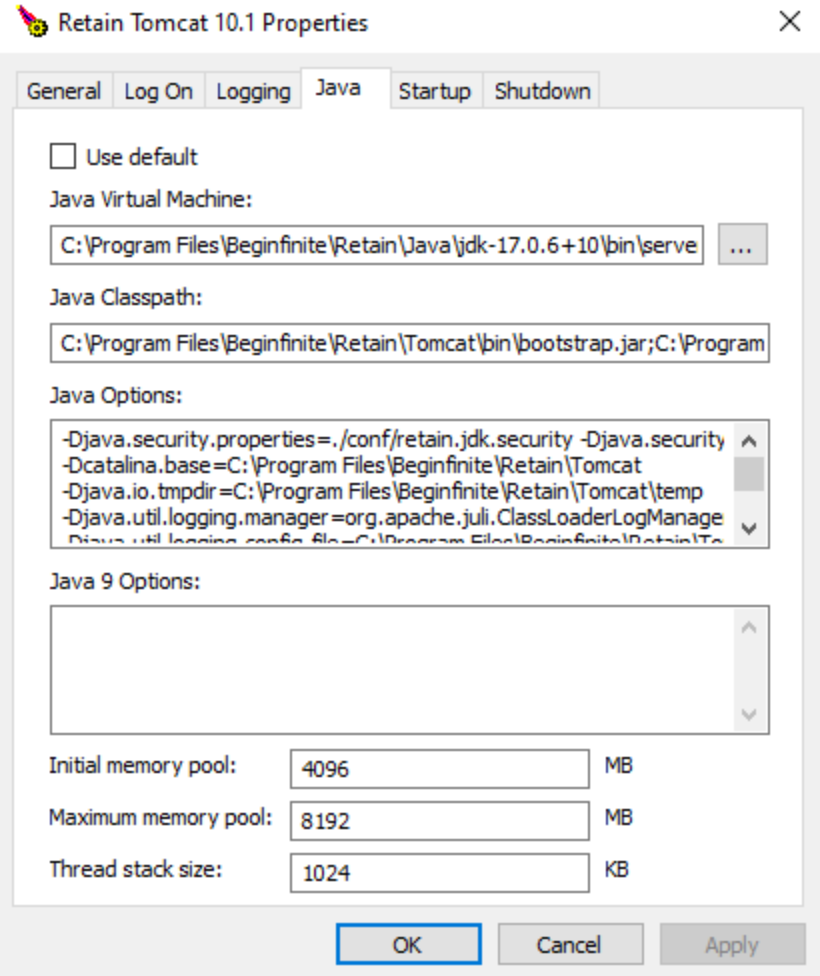
Retain-Tomcat is automatically tuned depending on system memory. To tune the system manually, edit the following lines with the appropriate parameters for your system. For example:

`CATALINA_OPTS="-Xms4g -Xmx12g -Xss-256k"`

```
JAVA_OPTS="$JAVA_OPTS -XX:MaxmetaspaceSize=1024m"
```

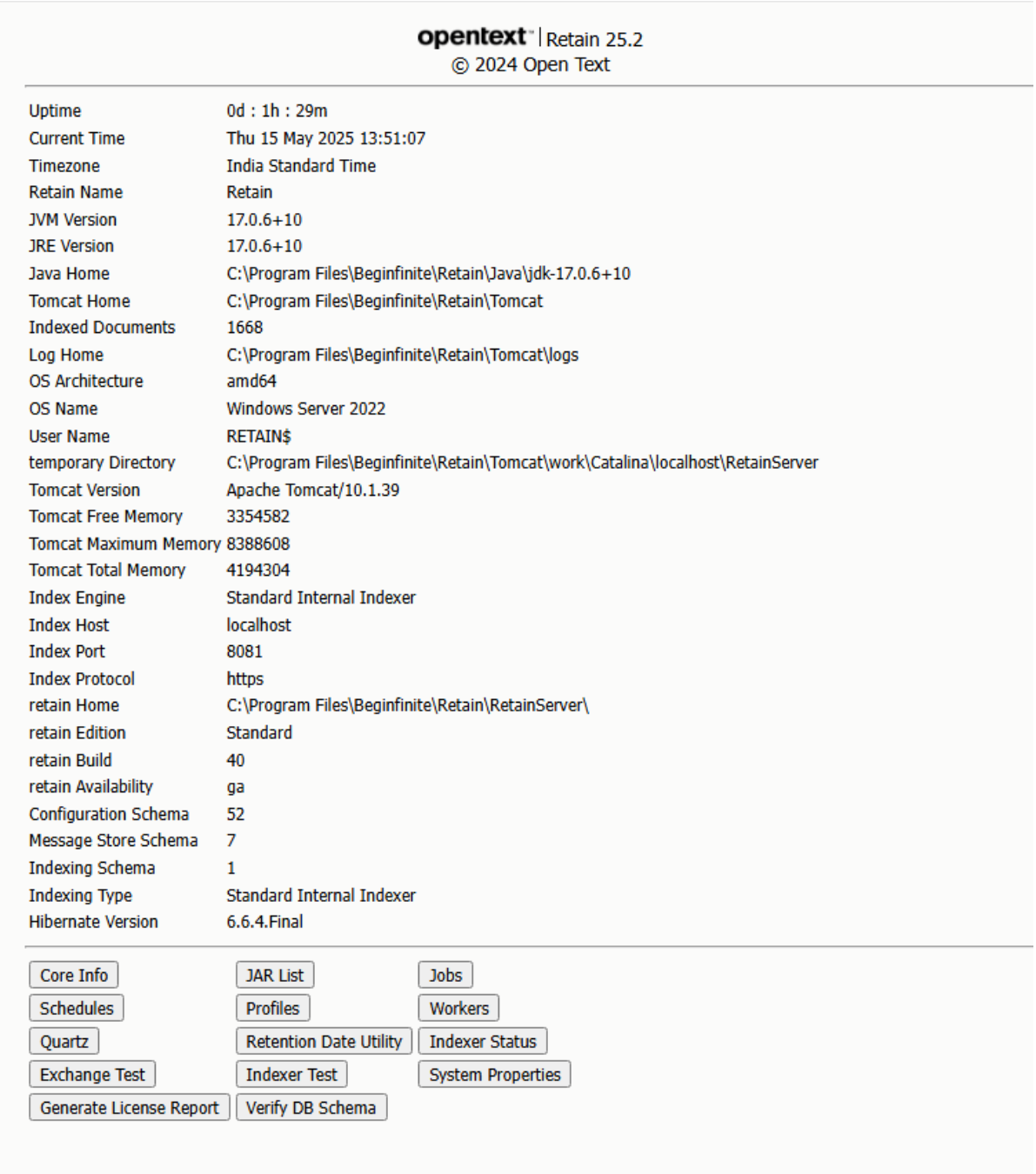
On Windows: run the Configure Tomcat utility to configure and modify the appropriate settings.

Retain 4.9.x or higher.



Checking memory

After logging into Retain Server, click the **About** link in the upper right of the interface. Among other stats, you'll see the tomcat Memory statistics.



Page File

NOTE: The page file guide is included here as a courtesy. It is your responsibility to find the proper procedures and documentation from your chosen operating system vendor. The entire responsibility for installation, care, and maintenance of the operating system lies with the customer. We do not provide any kind of support for the operating system.

In some environments, tomcat runs out of resources and crashes. Increasing the page file might solve this issue. In some cases doubling the page file is sufficient, in other cases increasing it to 64GB is sufficient.

8 Ports Used

- “Retain Server” on page 47
- “Retain Worker” on page 48
- “Retain Message Router” on page 48
- “Retain Reporting & Monitoring Server” on page 48
- “Retain Stubbing Server” on page 49
- “From RSM WAN IP to Untrusted, all TCP/UDP ports.” on page 49
- “Database Management Systems” on page 49
- “Blackberry” on page 49
- “Exchange” on page 50
- “Google Apps Module” on page 50
- “GroupWise” on page 50
- “Mobile” on page 50
- “Office 365” on page 50
- “CellTrust Secureline” on page 51

Retain utilizes many ports to facilitate communication between Retain components themselves and with the different systems being archives.

By default, Retain uses the following ports for the following services:

Retain Server

The Retain Server is the key component in the Retain system. Other Retain processes communicate with the Server through port 48080 by default. The Server is always listening on that port regardless of how other components might be configured to communicate with it (i.e., SSL port 443).

- 48080 (TCP) Requires incoming access if any Retain processes are running on a server external to the server hosting the Retain Server.
- The Server Requires outgoing access if the Reporting & Monitoring Server component has been installed on a server external the server hosting the Retain Server. See also the Reporting & Monitoring component in this article.
- 48009 (TCP) The AJP (Apache JServ Protocol) port is used by for communication between the web server and Tomcat. Since both should reside on the same server, there are no external port access requirements.
- 48005 (TCP) Port is used by Tomcat to accept shutdown requests.

- 80 / 443 (TCP - HTTP / HTTPS) Requires incoming access to reach the Server web interface.
- 25 (TCP) Requires outgoing access so that the Retain Server can send email notifications on server errors, job statuses, and job errors. Outgoing access to Database Management System (DBMS) port. This depends on the database system you are using. See the “Database Management Systems” section of this page.

Other ports must be opened on the server hosting the Retain Server, depending on the modules being used.

Retain Worker

The Retain Worker is the component that pulls the data from the messaging source, whether that be an email system or a mobile device.

- 48080 (TCP) Requires outgoing access if on a server external to the Retain Server.
- 80 or 443 (TCP HTTP or HTTPS) Requires incoming access to reach the Worker web interface

Other ports must be opened on the server hosting the Retain Worker, depending on the modules being used; and, in some cases, on servers hosting the messaging system Retain is archiving.

Retain Message Router

The Message Router is for customers of the Retain for Mobile module. The Message Router connects with mobile devices to handle SMS message log forwarding, BBMP device configuration. It typically would sit inside a DMZ.

- 443 (TCP) Requires both incoming from and outgoing access to the Internet as well as incoming from and outgoing access to the Retain Server.
- 111/2049 (UDP / TCP) for NFS Client services. (Only required if not using REST) Requires outgoing access if the Worker is not on the Message Router server so that it can place the logs on the Worker's server. Samba can be used, but NFS seems to be more reliable.

Other ports may need to be considered for NFS to work properly (http://serverfault.com/questions/377170/which-ports-do-i-need-to-open-in-the-firewall-to-use-nfs).

It is up to the customer to do this research to get NFS services to work properly.

Retain Reporting & Monitoring Server

New to Retain v3.4, this process provides archive job and server reporting and monitoring services. It is not installed by default unless specifically selected during the installation process. With Retain 4.2 it is a required install to allow Mailbox Error Monitoring to occur.

- 48080 (TCP) Requires both incoming and outgoing access if on a server external to the Retain Server.
- 80 / 443 (TCP) Requires Incoming access to reach the R&M Server's web interface.
- 25 (TCP) If running on a server external to the Retain Server, then it requires outgoing access.

Retain Stubbing Server

The Retain Stubbing Server is the component that provides stubbing services to the Retain Server. It is rarely used or installed by customers. See the Administration and Users Guide for a listing of its advantages and disadvantages.

- 48080 (TCP) Requires outgoing access if on a server external to the Retain Server.
- 80 / 443 (TCP) Requires incoming access to reach the Stubbing Server web interface.

From RSM WAN IP to Untrusted, all TCP/UDP ports.

If the RSM WAN IP is a private IP, it needs be NATed to an appropriate routable IP address. The LAN IP address does not need a corresponding inbound NAT rule.

Database Management Systems

The following are the default ports these database management systems use, but they are configurable within those systems. Requires incoming access for the database server and outgoing access on the Retain Server.

- MySQL: 3306
- MS SQL: 1433
- Oracle: 1521
- Postgres: 5432

Blackberry

BES Web Services (SOAP) only supports the secure ports via TLS. Retain initiates contact with BES Web Services if you are syncing the Address Book with the BES. Outgoing access is required for the Retain Server and incoming access for the BES server for the following ports:

- BES 12: 18084 (TCP)
- BES 10: 38443 (TCP)
- BES 5: 443 (TCP)
- 111/2049 (UDP / TCP) on the Blackberry Enterprise Server (BES) for NFS Server services.

The Blackberry module requires incoming access if the Worker is not on the BES server so that it can retrieve the BES logs. Samba can be used, but NFS seems to be more reliable.

Other ports may need to be considered for NFS to work properly (http://serverfault.com/questions/377170/which-ports-do-i-need-to-open-in-the-firewall-to-use-nfs).

It is up to the customer to do this research to get NFS services to work properly.

Exchange

- 80 / 443 (TCP HTTP / HTTPS) on the Worker server. Requires outgoing access. The autodiscover process attempts https to the CAS servers for connecting to Exchange mailboxes. If that fails, it uses port 80 as a last resort. If Exchange / autodiscover / EWS are set up properly, only port 443 should be necessary.
- 3268 or 3269 (TCP). Requires outgoing access. It uses one of these ports for LDAP lookups to the global catalog host, which is the primary database server for Active Directory.
- Port 3268 for plaintext and 3269 for SSL (recommended).
- 53 (UPD). Requires outgoing access. It's the port used by DNS. Retain utilizes DNS lookups during its autodiscover process.

Google Apps Module

- 443 (TCP) Requires outgoing access for the Server (address book sync) and the Worker (it attempts to use the Gmail API for archiving; if that fails, it reverts to IMAP, thus the need for port 993 as described below).
- 993 (TCP). Requires outgoing access for the Workers only. Under certain circumstances, the Worker may switch to using IMAP over SSL when requesting email from Gmail.

GroupWise

- 7191 (TCP) Requires outgoing access so that the Retain Server can download the Address Book. This is the default SOAP port the GroupWise POAs use, but this is configurable and is dependent upon the POA agent setting in GroupWise.

Mobile

- 80 / 443 (TCP) Requires both incoming and outgoing access so that the Retain Server and the Message Router can communicate device configuration information with each other. See also the Retain Message Router component in this article as well as the Mobile subsection under Retain Worker. See also the Retain Message Router component in this article.
- 111/2049 (UDP / TCP) for NFS Server services Requires incoming access if the Worker is not on the Message Router server so that the Message Router can place the logs on the Worker's server. Samba can be used, but NFS seems to be more reliable.

Other ports may need to be considered for NFS to work properly (http://serverfault.com/questions/377170/which-ports-do-i-need-to-open-in-the-firewall-to-use-nfs).

It is up to the customer to do this research to get NFS services to work properly.

Office 365

- 443 (TCP) Requires outgoing access. Retain uses SSL to connect with Office 365 to authenticate users logging in to Retain.

CellTrust Secureline

- 111/2049 (UDP / TCP) on the CellTrust Secureline server for NFS Server services.

Requires incoming access if the Worker is not on the CellTrust Secureline server so that it can retrieve the CSV logs. Samba can be used, but NFS seems to be more reliable.

Other ports may need to be considered for NFS to work properly (http://serverfault.com/questions/377170/which-ports-do-i-need-to-open-in-the-firewall-to-use-nfs).

It is up to the customer to do this research to get NFS services to work properly.

9 Messaging Systems Supported for Retain Archiving

Microsoft Exchange

- Exchange Server 2019
- Exchange Server 2016
- Windows Authentication for EWS and Autodiscover must be enabled
- Basic Authentication for EWS and Autodiscover must be enabled

Microsoft Office 365

- Office 365 Enterprise E1, E3, or E5 plan
- Office 365 Business, Business Premium plan
- Microsoft 365 Business Basic
  Microsoft 365 Business Standard
- Microsoft 365 Business Premium

OpenText GroupWise

- SOAP port must be enabled on all POAs to be scanned for messages.
- SSL is supported, but significantly impacts performance.
- GroupWise 24.4
- GroupWise 24.3
- GroupWise 24.2
- GroupWise 23.4
- GroupWise 18

Android (RetainService app)

- Version 5.0 (Lollipop) and later

Blackberry

- BES 10, 10.2, 12, 12.2
- BBM Enterprise

Bloomberg

- CCNS account

CellTrust

- CellTrust SL2
- CellTrust Gateway 4.2.0 or later

GroupWise Messenger

- Retain Server URL
- Retain Application Key and Secret

Ionlake MyRepChat

- Retain Server URL
- Retain Application Key and Secret

TeleMessage

- Retain Server URL
- Retain Application Key and Secret

UM Labs

- Retain Server URL
- Retain Application Key and Secret

Wireless Carrier AT&T

- AT&T Message Archiving Service agreement
- Retain Server URL
- Retain Application Key and Secret

Google Applications

- Google Workspace (Formerly G Suite)

10 Getting Started with Retain Archiving

First, Plan Your Retain Archiving System

As a best practice, we recommend that you start by creating a written plan. Obviously, this plan will evolve and expand over time, but beginning with a plan is the most effective and efficient way to ensure a successful Retain deployment.

Do the following:

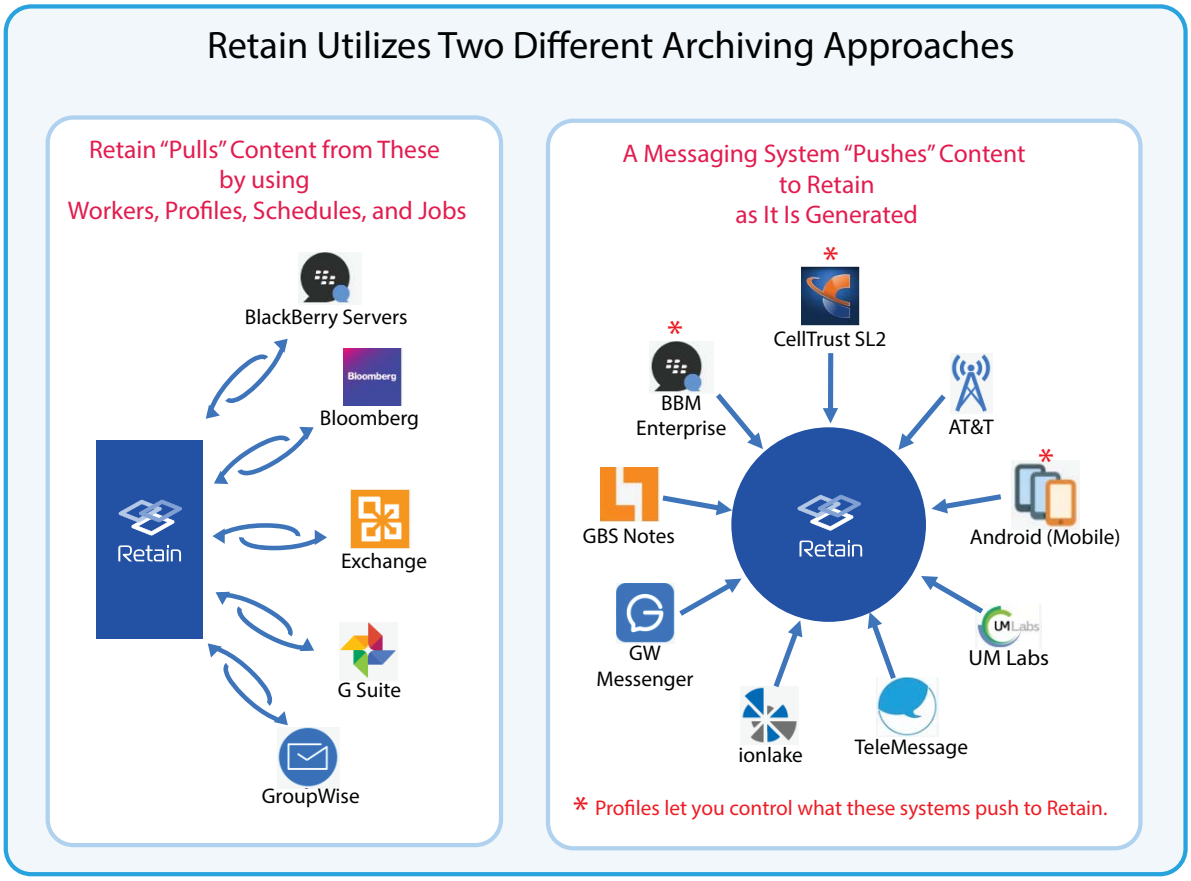
1. Identify the stakeholders in your organization who have a vested interest in a successful Retain deployment.
2. With the stakeholders, decide on a convenient and accessible way to propose, review, and finalize your Retain deployment plan.
3. Create your plan by doing at least the following:
    a. Identify the messaging systems that Retain will service.
    b. Record each system's DNS and IP address information, physical location, administrator information, and any other information relative to the system's future interaction with Retain.
    c. Identify the users and groups who will access Retain, including:
        - Administrators
        - Regular users and groups
    d. Identify the platforms that users and groups will use to access Retain: desktops, mobile devices.
    e. Identify how users will authenticate to Retain: Username/Password, OpenID Connect, or Multi-factor Authentication through NetIQ Advanced Authentication.
    f. Identify those who will need offline access to Retain
    g. Save a space to record details identified during the deployment, such as system credentials, licenses, and other information that Retain and third-party systems will require as you connect them together.
    h. Determine the times when systems and network traffic are slow and best suited for archiving operations to take place. This will feed into schedule creation as proposed in "Creating Schedules" in the *Archiving Guide*.
    i. Articulate the scope of what needs to be archived for each system: users, groups, message types, date ranges, attachments, and so on.
    j. Avoid redundant data collection.

Next, Match Your Messaging Systems to Retain's Archiving methods

Retain's archiving methods are designed to match the messaging systems being archived.

Find your messaging platforms in Figure 10-1 to determine which of Retain's archiving models your organization requires.

Figure 10-1 *Retain's Archiving Models Match Platform Needs*



Retain Pulls Data to Its Message Store	Messaging System Pushes to Retain's Message Store
◆ Traditional email systems store messages in their message stores. ◆ You create jobs that run according to schedules you set and "pull content" that matches your profile-specified parameters from those message stores. ◆ Retain provides practically unlimited search and retrieval capabilities for administrators and users with sufficient archive-access privileges.	**IMPORTANT:** These messaging systems require one or more additional licenses to connect with Retain. ◆ More recent messaging providers let users communicate in real time as they exchange voicemail, text messages, and so on. ◆ These providers don't offer a centralized message store. At best, data is temporarily kept in device-based storage. ◆ Retain works with these providers so that as message data is generated, it is also archived by Retain. ◆ Retain automatically organizes archived message data for search and retrieval. ◆ As the graphic shows, for three of the systems, profiles let you control what is archived.

Where to Go from Here

- **Pull-based:** For your messaging platforms that require Retain-initiated archiving, go to "Understanding Retain Jobs" in the *Archiving Guide*.
- **Push-based:** For your messaging platforms that support push-based archiving, skip to "Setting Up Push-Based Archiving" in the *Archiving Guide*..

11 System File Locations

Retain Server components reside:

- Linux: /opt/beginfinite/retain/
- Windows: [Drive]:/Program Files/Beginfinite/Retain

In this directory you find subdirectories named for the Retain components and support systems:

- java
- tomcat
- RetainServer
- RetainWebUI
- RetainStubServer, if installed
- RetainStatServer, if installed (on this server), this component may be installed on a separate server.
- RetainWorker if installed (on this server), this component may be installed on a separate server. *N* being the worker number if more than one worker is installed.
- Logs are found in the /logs folder within the Tomcat folder.

To find your Retain system's storage path, log in to the Retain Server administration web interface (http://[ipaddress/DNS hostname]/RetainServer). Starting at the navigation pane on the left side under "Configuration", click on Server Configuration | Storage. Most customers leave it at the default setting under "Advanced Settings" of "Derive all file locations from the above base path", but you must click that checkbox to have it expand to show all the paths.

Other file locations can be found in the ASConfig file. ASConfig.cfg is stored in a directory off of your Retain installation:

- Linux default: /opt/beginfinite/retain/RetainServer/WEB-INF/cfg
- Windows default: [Drive]:\Program Files\Beginfinite\Retain\RetainServer\WEB-INF\cfg