
Repository Configuration and Sizing

<https://campus.barracuda.com/doc/43222405/>

Determine Daily Archive Quantity

Determine the quantity of data:

- **Legacy Data Size** - Data initially archived when you setup and run ArchiveOne.
- **Daily Data Size** - Amount of data archived daily.

For instance, if you intend to archive all data over six months old on a system that has been running for two years, then the *Legacy Data Size* is the size of all data over six months old, and the *Daily Data Size* is the average size of each days worth of data that is six months old - that is, the size of the data that was received six months ago minus the size of the data that has been deleted.

Note that it is impossible to determine the exact figures, however, you can make an estimate for sizing purposes, and adjust the amount later once you have been archiving for a while. You can also consider different *Daily Data Sizes* depending on whether the day being archived (six months ago) was a weekday or a weekend, if your business has differing levels of mail during the week.

If you have different user types, and you are planning on archiving them to different repositories, then do the sizing calculations separately for each repository.

Once you have estimated the *Legacy Data Size* and *Daily Data Size*, you can work out how much disk the ArchiveOne Service will need access to (either local hard drives or fast connected networked storage):

- **Filestore for long term storage of the index** - Typically start at 25% of the *Legacy Data Size* and grow each day by 25% of the *Daily Data Size*, and can be spread over multiple hard disks. The filestore is called the *Index Base Directory*.
- **Directories in the Index Base Directory** - Used for short term archive and index data preparation, which may take up to an additional 3Gb.

If you are using a [third-party storage manager](#):

- **Filestore for short term storage of archive files retrieved from the storage manager.** This can be configured, but is typically around 1Gb. This is called the *Archive Cache Directory*.

If you are not using a third-party storage manager:

- **Filestore for long term storage of the archive.** This typically starts at 50% of the *Legacy*

Data Size and grows daily by 50% of the *Daily Data Size*, and can be spread over multiple hard disks. The filestore is called the *Archive Base Directory*.

Once you determine the amount of storage space each directory needs, choose suitable locations on your server's disks or on fast connected networked storage large enough to cope with the required data size. You can add additional hard drives to extend the *Index Base Directory* and *Archive Base Directory*.

Sample Use Case

Consider an organization without a third-party storage manager. They estimate their *Legacy Data Size* is 10Gb, and their *Daily Data Size* is 100Mb per day on weekdays and 4Mb per day on weekends. After one year the *Index Base Directory* will be just over 9Gb (2500Mb for the legacy data, 2six0 days at 25Mb/day, 104 days at 1Mb/day) and the *Archive Base Directory* will be just over 18Gb (5000Mb for the legacy data, 2six0 days at 50Mb/day, 104 days at 2Mb/day).

Third-Party Storage Manager

If you are using a third-party storage manager, then the archive and index files are written to offline storage, which means you should expect 75% of the *Legacy Data Size* initially, and 75% of the *Daily Data Size* each day to be stored on it. allow enough free media available to cope with the storage size. When configuring a repository that uses CA BrightStor ARCserve Backup or HP Storage Data Protector, ensure that the path to the repository index is written as a reference to a local drive, *not a UNC*, and in all cases you must ensure that the path to the repository index does not make use of a mapped drive as the ArchiveOne Service does not see the same drive mappings as do logged on users.

Mailboxes and Public Folders

Sizing considerations apply to each repository, if you plan to have multiple repositories. For a given set of mailboxes, the storage requirement for their archive is the same whether the mailboxes are put into one repository or split over several repositories.

Sizing considerations also apply to archiving public folders in addition to archiving mailboxes. In particular, when public folder items are archived they are still moved into an archive public folder (a different set of folders that must not overlap the source public folder) before being moved out of Microsoft Exchange.

You should consider the ways in which mailboxes can be grouped into different repositories:

- A repository can impose a default retention period on all mail within it, so if you have a requirement to keep mail for one group of users for a different length of time than another group, put them in different repositories.
- A repository can be removed without affecting any other repository. If you need to archive some users, but know that in the future you need to discard the archive and recover the disk space, if you put them into their own repository you can delete it without affecting other users.
- Repositories can occupy different disk areas. This means you can have different backup and disaster recovery plans for different repositories. By having a small number of important users in one repository, you can recover their archive from disaster much faster than a large repository of other users.
- Repositories can be configured based on whether they use a storage manager. This allows you to store some mailbox archives on optical disk or tape while others are stored on local hard drive.

Typically all mailboxes processed by an installation go into the same repository. If there is a natural split in the types of users being archived, you can split into multiple repositories. There is no maximum number of mailboxes that can be included in a single repository.

You can also group public folders. When you select to process a public folder hierarchy, you can specify into which repository its messages are archived. For more information on repositories, see [Mailbox and Journal Repositories Node](#).

PST Archiving

PST archiving makes use of a temporary area on disk to store messages to be archived, instead of using public folder store. This area is part of the "Server data" area, which can be configured by running the Configuration Wizard (select **Run Configuration Wizard** from the **Status** node's menu) and changing the path on the **Server Data Location** page. PST data rests in this area between being found by PST processing and being archived, but this is typically only a few hours.

Ensure that there is enough space in the disk area indicated by this path to hold temporary copies of all the mails in PSTs you are going to archive that have not yet completed archiving. Ideally you should monitor the size of this area as you introduce new PSTs for archiving, to ensure it is never flooded.

In general it is not necessary to allocate disk space for retrievals from the archive. If a user uses the [Search and Retrieval Website](#) to retrieve messages, they are retrieved into a **Retrieved items** folder in their mailbox. If an administrator searches the archive and retrieves from it, it goes into a mailbox, public folder, or PST - sizing for this folder need only be considered if you are planning to retrieve a large number of messages from the archive, such as for systematic checking.

© Barracuda Networks Inc., 2021 The information contained within this document is confidential and proprietary to Barracuda Networks Inc. No portion of this document may be copied, distributed, publicized or used for other than internal documentary purposes without the written consent of an official representative of Barracuda Networks Inc. All specifications are subject to change without notice. Barracuda Networks Inc. assumes no responsibility for any inaccuracies in this document. Barracuda Networks Inc. reserves the right to change, modify, transfer, or otherwise revise this publication without notice.