Synology®

ActiveProtect Data Deduplication White Paper

Last updated: Jul 18, 2025

Table of Contents

Introduction

As businesses evolve and experience data growth, data management challenges become real. Companies need to implement a streamlined data protection solution and storage solutions to securely store and manage their data in this day and age. As traditional backup systems are usually not capable of handling large volumes of data, this results in inefficient backups and slow data recovery. This is why businesses need to rely on data reduction technologies when selecting an efficient and affordable backup solution.

Synology's ActiveProtect appliance is a purpose-built backup solution that comes pre-configured with hardware and runs ActiveProtect Manager, a built-in software that allows businesses to consolidate and manage all their workloads in a cluster, no matter where their data is located. With immutability and air-gapped backups, ActiveProtect protects against ransomware attacks and ensures that a clean copy of your data is isolated and available for recovery when needed.

As an advanced data protection solution, ActiveProtect also aims to boost backup performance with data reduction technologies. Not only does ActiveProtect reduce network traffic, but it also minimizes the storage space needed to store backed up data. This ensures that companies can maximize their storage capacity to increase the amount of data stored on their backup appliance.

In terms of data protection solutions, data reduction technology has increased in importance and is now widely adopted by industry-leading backup solutions.

This document aims to introduce data deduplication techniques, the deduplication ratio, and explain how they are implemented in Synology's ActiveProtect appliance.

Data reduction technology

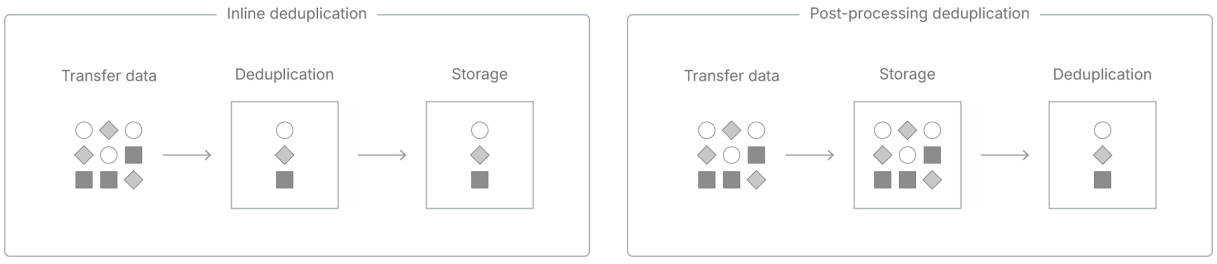
ActiveProtect utilizes data reduction technology to minimize backup storage consumption and optimize storage utilization so that users can back up multiple workloads efficiently. With advanced data reduction capabilities, ActiveProtect improves overall backup performance while ensuring the storage capacity has been utilized efficiently.

Data deduplication method explained

ActiveProtect uses inline deduplication technology. Inline deduplication removes redundant data before the data itself is even written to the storage.

With its integrated hardware and software design, the ActiveProtect appliance is pre-configured and already sized in advance for optimal performance. This means that the deduplication process itself does not impact backup efficiency. Instead, the deduplication algorithm is used before data is written to the disk, effectively reducing storage consumption while improving storage utilization.

By contrast, post-processing deduplication technology only removes redundant data after completing the backup process.



ActiveProtect uses a unique identification method to classify and manage backup data, as part of its deduplication method. The backup data is divided into 4 KB chunks. Each chunk is assigned a unique fingerprint which is generated with the SHA-256 algorithm to ensure data integrity and uniqueness.

Once the fingerprint has been created, the system will build an index. This index will be used to match incoming data chunks with existing chunks that are already stored in the system. If duplicate data is detected, the data deduplication engine can remove the redundant information to optimize storage utilization and boost overall backup performance.

To eliminate redundant copies of data across multiple backup targets, ActiveProtect also integrates global deduplication technology. This technology combines source-side and server-side deduplication to improve storage efficiency and network performance across the platform and off-site backups.

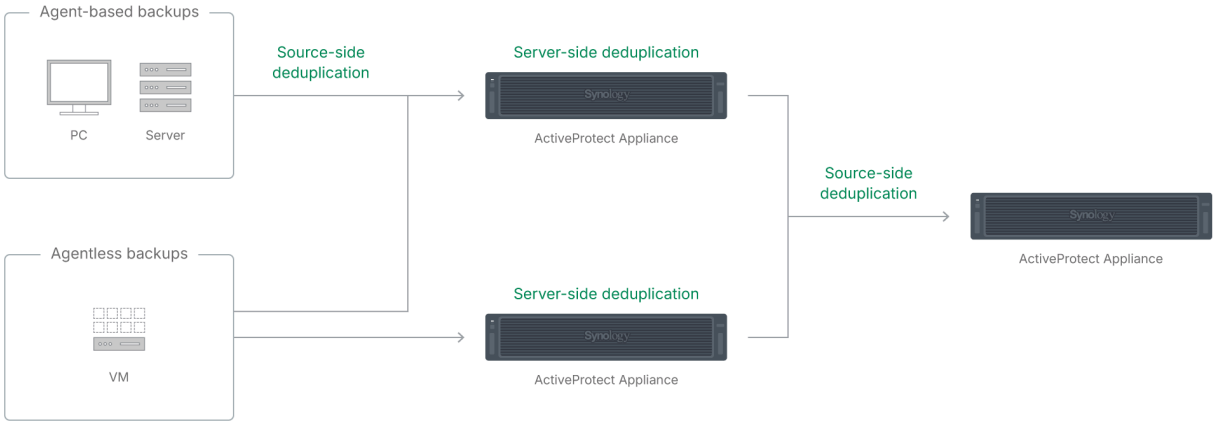
ActiveProtect implements server-side global deduplication when backing up various platforms—a crucial feature modern-day enterprises require as employees' data may be stored across multiple environments, such as on-prem file servers or SaaS applications, such as Microsoft 365. Even though this data may be stored on separate platforms, they might contain identical data, which leads to unnecessary storage consumption and an increase in expenses.

ActiveProtect uses intelligent data recognition to identify and eliminate duplicate data stored across platforms while backing up data. This reduces storage requirements and meets global server-side deduplication requirements.

ActiveProtect's global deduplication technology goes beyond improving storage efficiency within a single site. It also plays an important role for companies looking to adopt the off-site backup strategy and strengthen their 3-2-1-1-0 backup strategy.

Businesses tend to replicate their backup data to off-site locations to meet their data protection needs. However, large backup data volumes and varying network infrastructure could lead to inefficiencies and slow backup performance, especially across geographically dispersed sites.

To address these challenges, ActiveProtect uses global-source-side deduplication to eliminate redundant data before the data is even transmitted to the appliance. This significantly reduces network traffic and optimizes off-site backup performance.



ActiveProtect is a smart and comprehensive data protection solution that helps businesses minimize enterprise storage demands and boost cross-platform and off-site backup efficiency through global deduplication technology.

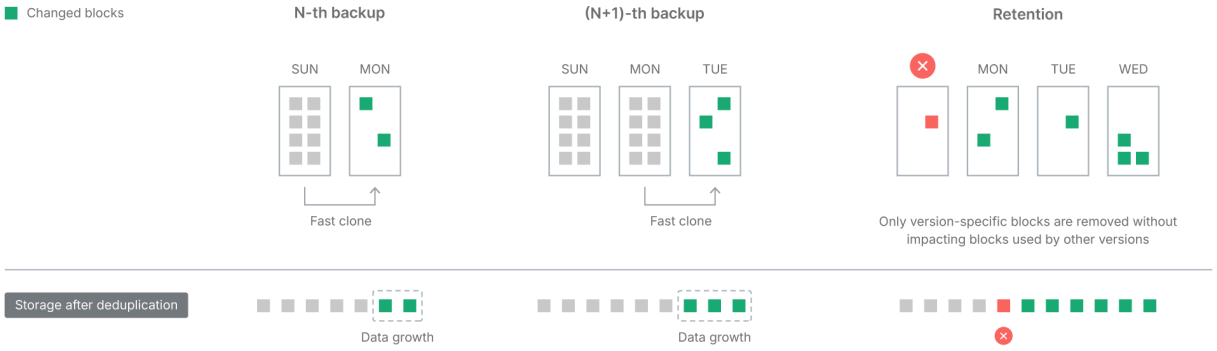
Modern backup chain explained

Changed Block Tracking (CBT) technology tracks changes made to your data. This is especially important when performing incremental backups or regular full backups. Traditional forever-incremental backups may impact subsequent backup chains if a backup version becomes corrupt. This is why traditional backup vendors usually recommend performing full backups on a regular basis to reduce the risk of this happening.

Synology's ActiveProtect uses a modern backup system, so users will no longer have this issue and can perform backups efficiently without the limitations associated with traditional backup methods.

ActiveProtect uses a unique backup chain technology and integrates multiple modern backup technologies into a single solution. This includes Changed Block Tracking (CBT), source-side deduplication, global server-side deduplication, and a snapshot-like backup version replication technology. Each backup version is created individually and does not have to rely on previous versions. This means that even if previous backup versions are deleted, the integrity of the newer backup versions is not affected.

The snapshot-like backup version replication technology quickly copies the disk image of the previous version on the server side. Synology ActiveProtect's modern backup system combines CBT and source-side deduplication technology to ensure only changed blocks are written to the new backup version's disk image. The new version's disk image only needs to store the new data blocks via global deduplication, significantly reducing storage space.



Data reduction ratios

Understanding data reduction ratios

The data deduplication ratio is a metric used to evaluate the deduplication performance of backup solutions. Using this metric, users can assess the backup performance of multiple vendors, making it crucial for companies looking to select a backup solution. However, data deduplication ratios can vary according to which data deduplication technology was used and the type of data that has been backed up.

The **deduplication ratio** or **compression ratio** refers to the ratio between the original data that will be stored and the actual storage space used to store the backed up data at the backup destination after duplicate data has been removed.

As an example, let's imagine if 200 GB of data is backed up from a device and takes up 20 GB on the backup server. When we divide the backed up data (200 GB) with the actual amount of space taken up on the backup server (20 GB), then the deduplication ratio is 10:1, meaning that the deduplication rate is 90%.

Vendors tend to use data deduplication ratios in marketing to promote backup capabilities. Terms such as "deduplication ratio" or "average deduplication rate" are widely used to promote the effectiveness of data reduction technology. Some vendors might even claim that their products have a deduplication ratio of 50:1 or 60:1, which makes them seem much better than many competing products in the market.

However, note that data deduplication ratios depend on the type of deduplication technology implemented, the type of data backed up, which backup method was used, and how the data deduplication ratio itself has been calculated. Based on these criteria, comparing backup products across different vendors on the basis of data deduplication ratios might not be the best approach.

It is important to note how each backup vendor calculates the data deduplication ratio for each product and how much storage space is used at the backup destination.

Calculating data reduction ratios

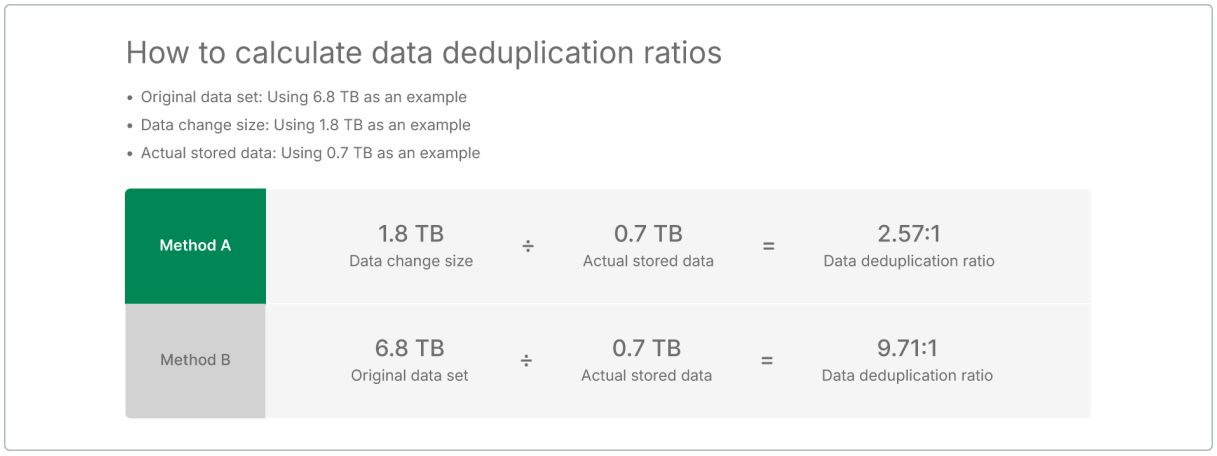
Even though each backup vendor calculates data deduplication ratios differently, it can be summarized into these 3 key points:

1. Total data capacity before data deduplication
2. Data change size
3. Actual amount of data stored at the backup destination

Backup vendors tend to use the original data size as their baseline for calculating data reduction ratios. This is typically expressed through the usage of terms such as scanned data size, disk data size, or application data size. However, these values usually include both **processed** and **unprocessed** data, meaning that the calculations might not be as accurate.

By contrast, Synology uses the data change size as the baseline for calculating data reduction ratios. In this method, only the actual changed size of your data is included while previously processed backup data is excluded, providing a more accurate reflection of the effect of data deduplication.

Take a look at the following example:



With **Method A**, only 1.8 TB of data is changed in size as the redundant data is removed through CBT and then further compressed so that there is a total of 0.7 TB of data that is to be stored at the backup destination.

With **Method B**, the total data set is divided by the actual stored data at the backup destination. This tends to yield a higher data deduplication ratio. In reality, only the changed backup data size is able to show the actual amount of data after data deduplication.

Synology recommends using the data change size as the primary metric for measuring the effectiveness of data deduplication. This is because using the total data set contains both the “old” and “new” data from the backup source and by dividing this number with the final amount of retained data, it may deliberately inflate the data reduction ratio. It also does not clearly demonstrate a backup vendor’s ability to deduplicate data.

ActiveProtect’s data reduction ratio

Calculating the data deduplication ratio with the total backup data size before deduplication can yield higher deduplication ratios. However, it may not accurately reflect the actual size of your data after deduplication. ActiveProtect uses the data change size to calculate the exact size of your data after deduplication.

As shown in the chart below, backup vendors use various calculation methods to calculate deduplication ratios. However, their customer stories do not reflect those deduplication ratios.[1]

Vendor	Formula	Announcement	Customer stories
Synology	Data change size / Storage usage	80%	80%
Vendor D	Source data size / Storage usage	65:1	25:1 ~ 55:1
Vendor V	Scanned size / Stored size	50:1	70%
Vendor C	Application size / Data size on Disk	90:1	80%

In short, users should not solely rely on data deduplication ratios provided by vendors. Instead, they should focus on comparing deduplication efficiency across backup products with consistent standards and evaluating benchmarks accordingly. This will allow users to have a more accurate assessment of the effectiveness of each product when looking to optimize storage capacity.

Notes:

1. The competitive intelligence presented in this document was based on publicly available information. This information is subject to change. Synology makes no representations or warranties, express or implied, regarding the accuracy, completeness, or reliability of the information herein. Any features, functionalities, or comparisons mentioned are subject to change without notice.

Conclusion

Data deduplication technology is a key feature of modern-day backup solutions. It reduces duplicate data, saves storage space, and reduces load on network transmissions. However, backup vendors use various methods to calculate deduplication ratios, making it difficult to compare them directly.

Synology's ActiveProtect uses the data change size as the primary metric for calculating deduplication ratios as it better reflects real-world results compared to other backup vendors' methods of calculating data deduplication ratios based on the total amount of data. By implementing advanced technologies such as Changed Block Tracking (CBT), source-side, and global server-side deduplication, ActiveProtect provides high-performance data backup capabilities and ensures the integrity of your backed up data.