

Data Deduplication White Paper

Based on Active Backup for Business 2.3.0



Table of Contents

Introduction	3
Data Reduction Technology	4
Deduplication processes	
Changed Block Tracking (CBT)	
Deduplication ratio	
Conclusion	9

Find your information

Synology publishes a wide range of supporting documentation.

In [Knowledge Center](#), you will find useful Help and FAQ articles, as well as video tutorials breaking up processes into handy steps. You can also find User's Guides, Solution Guides, brochures, and White Papers. Experienced users and administrators will find answers and guidance in technical Administrator's Guides and Developer Guides.

Got a problem and unable to find the solution in our official documentation? Search hundreds of answers by users and support staff in [Synology Community](#) or reach [Synology Support](#) through the web form, email, or phone.



Introduction

Data loss or system failure can be disastrous, regardless what storage device or system you're using. The devices most important to you – such as your work computers or servers – deserve a concrete and effective backup strategy that can help you avoid unexpected system failure in the event of natural disasters or other technical issues.

Active Backup for Business (ABB) allows you to manage all of your VMware, Hyper-V, Windows, Linux, and file server backups from a single, central console on your Synology NAS. With instant restore options for Windows devices, virtual machines, and individual files, you can keep critical services functioning, even in the face of sudden disaster. On top of that, Active Backup for Business utilizes a number of the latest techniques, such as **data deduplication** and **incremental backup**, to enhance backup performance and minimize the amount of physical space required by the backup data. Because these approaches are so commonly used among different backup services, our users frequently inquire about the technology and measurements utilized for ABB's deduplication ratio.

The aim of this document is to introduce some of the data deduplication techniques used and define the deduplication ratio implemented in Synology Active Backup for Business.

Data Reduction Technology

Data reduction technology is designed to make backup data consume less storage space, optimizing storage usage and allowing the user to backup as many devices as possible. Active Backup for Business has implemented cutting-edge technology to conduct data reduction and enhance overall backup performance.

Deduplication processes

Deduplication is an effective way to optimize storage space by removing any redundant data. The process of deduplication can be categorized into two types, “in-line” and “post-process”. **In-line deduplication** will remove any redundant data **before** the data is written to storage, whereas **post-processing deduplication** will only remove data **after** the backup process is complete.

In-line deduplication can be implemented on both the backup agent and the backup destination, depending on the backup solution. However, when deduplication is implemented via an agent, data scanning, deduplication and transmission will all occur at the same time, resulting in a more time-consuming and performance-impacting (backup destination) backup process.



Synology **Active Backup for Business** incorporates **in-line deduplication** into its design, however, data deduplication processes only occur on the destination before being written to storage. This is done in order to avoid the over-consumption of time and CPU resources on the backup source that is generally caused by source-side deduplication. Additionally, Active Backup for Business allows users to back up various systems and devices to a single destination that might contain redundant data.

In order to eliminate all redundant copies of the data across multiple targets, Active Backup for Business has also implemented **global deduplication**. Global deduplication allows Active Backup for Business to deduplicate the data across many backup tasks within a single shared folder, optimizing actual storage consumption without compromising the overall backup performance.

As for its deduplication mechanism, Active Backup for Business uses a unique identifier to classify each file and file chunk that are backed up. The Active Backup for Business deduplication engine does this by dividing backup data into individual **4KB** file chunks and assigning each one a unique fingerprint. This fingerprint is computed by a deduplication engine that uses the **SHA-256 algorithm** to identify each data chunk.

After a fingerprint is created for a file chunk, the system will create an index of the fingerprints for each stored file chunk. This index is used to match up incoming file chunks with their stored counterpart. Once identified, the deduplication engine will promptly remove the redundant data, optimizing the device's storage usage.

Changed Block Tracking (CBT)

Changed Block Tracking (CBT) is a type of incremental backup technology commonly used in a number of different backup solutions. **Incremental backup** is a type of backup that only backs up data that has been modified or newly created since the most recent backup version. CBT can help with incremental backups by conserving time and resources through only backing up altered blocks of data rather than completing a full backup every time. Because of this, the size of the data that will be transferred via the backup application, a.k.a., the **backup data transfer size**, ends up being smaller than the original data size on the backup source.

Active Backup for Business has included this feature in its architecture for backups to PC, physical servers, and virtual machines. ABB will perform forever-incremental backup for a backup task to maximize the number of available backup versions and minimize the amount of storage used for backup retention by only backing up modified data after a full backup has been executed.



Deduplication ratio

Definition

The **deduplication ratio** is a metric used to assess the effectiveness of the data deduplication performance of a backup solution. This ratio is often a deciding factor for users selecting a backup solution, as it allows them to analyze different features and performance across different providers. However, deduplication ratios can vary depending on the deduplication technology utilized and the general characteristics of the data.

A **deduplication ratio** or **compression ratio** is defined as the measurement of the data to be stored and the amount of actual storage spaced used for storing the backup data on the backup destination after any redundant data has been removed.

For example: You backed up 100 GB of data from your device, and the actual amount of space that was used on the backup storage server was 20 GB. Dividing the amount of backup data (100 GB) by the space used on the backup storage server (20 GB) will result in a deduplication ratio of 5:1, or an 80% deduplication rate.

The deduplication ratio is also widely used in marketing to advertise the strengths of a backup product. For example, promotions may say that a backup product has a 20 times higher deduplication ratio than others on the market. However, deduplication ratios vary depending on the deduplication technology utilized, the general characteristics of the data, the backup methods used, and how the deduplication ratio is calculated.

Because of this, the deduplication ratio may be not always be the most accurate way to compare different products. Therefore, it is important to keep in mind how each product's deduplication ratio is calculated, as well as the resulting amount of storage that is actually taken up on the backup destination.

Calculation

The deduplication ratio can be calculated using the following values:

- Total capacity of backup data before removing duplicates
- Backup data transfer size
- Actual capacity used on the backup destination

In general, a good method of calculating the ratio is to divide the **backup data transfer size** by the **actual capacity used on the backup destination**. This is the method that Synology uses.

On the other hand, some backup providers will divide the **total capacity of backup data before removing duplicates** by the **actual capacity used on the backup destination**, which may not necessarily produce an accurate deduplication ratio.

Consider the following scenario for example:

- Backup solution: A
- Total capacity of back up data before removing duplicates: 6 TB
- Backup data transfer size: 1.2 TB
- Actual capacity used on the backup destination: 0.83 TB
- Advertised deduplication ratio: 7.2:1

In this scenario, backup solution A is advertised as having a **7.2:1** (6 TB / 0.83 TB) deduplication ratio. However, this calculation is not entirely accurate. This is because only **1.2 TB of backup data** was transferred, the amount of redundant data was then reduced through CBT, and the data was further compressed to **0.83 TB** for storage on the backup destination.

Backup solution A divided the **total capacity of backup data** by the **actual capacity used on the backup destination** to obtain a higher deduplication ratio. But in reality, only the **backup data transfer size** shows the actual amount of data that will be deduplicated.

Dividing the **backup data transfer size** by the **actual capacity used on the backup destination** would produce an even more accurate result. Therefore, the deduplication calculations can be adjusted as follows:

Deduplication ratio

Backup data transfer size	÷	Actual capacity used on the backup destination	=	Data deduplication ratio
1.2 TB		0.83 TB		1.44:1

Deduplication rate

(Backup data transfer size	-	Actual capacity used on the backup destination)	÷	Backup data transfer size	x 100%	=	Deduplication rate
	1.2 TB		0.83 TB			1.2 TB			30.8%

Backup method

Another element that can influence the deduplication ratio is the **backup method**, particularly whether or not backups are all **full backups** or if **incremental backup** is being used. Backup systems that only perform full backups may appear to have a better deduplication ratio than those that use incremental backups. This is because incremental backups only transfer changed blocks of data and remove any redundant data before transferring the data, resulting in less actual capacity used on the backup destination. Full backups, on the other hand, include a large amount of redundant data in comparison to the original or previous backup version.

In summary, although incremental backup solutions may not display a better deduplication ratio than full backup solutions, they still consume less actual storage space on the backup destination.

Active Backup for Business deduplication ratio

As mentioned previously, using the **total capacity of back up data before removing duplicates** to calculate the deduplication ratio can result in better deduplication ratios, but they may not reflect the actual amount of data that was deduplicated. Active Backup for Business utilizes the **backup data transfer size** instead, since this represents the actual amount of data that is going to be deduplicated. As shown in the table below, Active Backup for Business's deduplication techniques resulted in less **actual capacity used on the backup destination**, proving that it saves users more storage space overall.

In conclusion, instead of relying on products' deduplication ratios, users can focus more on the actual amount of storage that was taken up in relation to the backup transfer size to get a better idea of how that product helps you save storage space.

The following table provides the actual deduplication ratio for **Active Backup for Business** and compares it to the deduplication ratios of other backup providers.

Provider	Backup data transfer size	Actual capacity used on the backup destination	Deduplication Ratio	Deduplication Rate
Synology Active Backup for Business	1.2 TB	637.2 GB	1.2 TB / 0.62 TB = 1.93:1	48.18%
Provider A	1.2 TB	819.8 GB	1.2 TB / 0.8 TB = 1.5:1	33.33%
Provider B	1.2 TB	847.7 GB	1.2 TB / 0.82 TB = 1.45:1	31.08%
Backup source contents: Windows PC x 9, Windows Server on VM x 1				



Conclusion

Data deduplication is a complex process that allows users to reduce the amount of redundant data in order to more efficiently manage backups and optimize storage space. Since deduplication is a common technique used in a number of different backup solutions, many users want to know the deduplication ratios to compare different products. However, with a number of measurement methods available and different factors that may impact deduplication measurement techniques, the provided ratios for each product may not always be reliable. Thus, when considering the deduplication ratio for different products, it is more important to focus on the actual amount of physical storage space that is being used to store backup data after deduplication.

Synology Active Backup for Business not only provides a comprehensive backup solution that can be utilized across various platforms, but also has an architecture that enables quick, reliable, and effective data deduplication, allowing you to save time and get the most of your storage space.



**SYNOLOGY
INC.**

9F, No. 1, Yuan Dong Rd.
Banqiao, New Taipei 220545
Taiwan
Tel: +886 2 2955 1814

**SYNOLOGY
AMERICA CORP.**

3535 Factoria Blvd SE, Suite #200,
Bellevue, WA 98006
USA
Tel: +1 425 818 1587

**SYNOLOGY
UK LTD.**

Unit 5 Danbury Court, Linford Wood,
Milton Keynes, MK14 6PL,
United Kingdom
Tel.: +44 (0)1908048029

**SYNOLOGY
FRANCE SARL**

102 Terrasse Boieldieu (TOUR W)
92800 Puteaux
France
Tel: +33 147 176288

**SYNOLOGY
GMBH**

Grafenberger Allee 295
40237 Düsseldorf
Deutschland
Tel: +49 211 9666 9666

**SYNOLOGY
SHANGHAI**

200070, Room 201,
No. 511 Tianmu W. Rd.,
Jingan Dist., Shanghai,
China

**SYNOLOGY
JAPAN CO., LTD.**

4F, No. 3-1-2, Higashikanda
Chiyoda-ku, Tokyo, 101-0031
Japan

Synology®



synology.com

Synology may make changes to specifications and product descriptions at any time, without notice. Copyright © 2022 Synology Inc. All rights reserved. ® Synology and other names of Synology Products are proprietary marks or registered trademarks of Synology Inc. Other products and company names mentioned herein are trademarks of their respective holders.