White Paper

Drive Space Efficiency Using the Deduplication/Compression Function of the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series

The Deduplication/Compression function is provided by the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series.
This document provides an overview of the Deduplication/Compression function and points to consider for implementing this function. It also explores the benefits of this feature.

Table of Contents

List of Figures

List of Tables

Preface

In recent years, a large number of companies have been facing the challenge of managing rapidly increasing business data as represented by big data.
In addition, the initial investment and operating costs for storage tend to become higher due to increased amounts of data processed.
Reducing the IT cost regarding systems is an urgent need for many companies.
The compression/deduplication technology is increasingly being focused on as a possible solution to optimize system resources.
This advanced technology helps reduce a system's TCO by storing rapidly increasing data efficiently and compactly.

By providing test results, this document explains the extent of the storage space efficiency that can be achieved by using the Deduplication/Compression function of the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series.

■ Target Readers
This document targets the following readers.

- Readers who are looking to reduce implementation and operation costs of the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series
- Readers who are looking to understand how storage space efficiency is achieved using the Deduplication/Compression function of the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series

■ Abbreviations
The following abbreviations are used in this document.

- "ETERNUS AF series" for the FUJITSU Storage ETERNUS AF series
- "ETERNUS DX S4/S3 series" for the FUJITSU Storage ETERNUS DX S4/S3 series
  (excludes the ETERNUS DX60 S4/DX60 S3, the ETERNUS DX100 S4 /DX100 S3, and the ETERNUS DX8000 S3 series) *
- "ETERNUS AF/DX" for the FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series
- "TPP" for Thin Provisioning Pools
- "TPV" for Thin Provisioning Volumes
- "Deduplication/Compression function" for the Deduplication/Compression function of the ETERNUS AF/DX
- "Deduplication/Compression Volume" for a TPV with the Deduplication/Compression setting enabled in a TPP with the Deduplication/Compression setting enabled
- "non-Deduplication/Compression Volume" for a TPV with the Deduplication/Compression setting disabled in a TPP with the Deduplication/Compression setting enabled
- "HDD" for a Hard Disk Drive

* For the ETERNUS DX S3 series, the firmware version must be V10L60 or later; for the ETERNUS DX series, the Memory Extension must be installed.

■ Prerequisite
The product lineup and product information stated in this document are current as of December 2017.

1. Overview of the Deduplication/Compression Function

This chapter provides an overview of the Deduplication/Compression function while explaining the general concept of data deduplication and compression technologies.

1.1. What are "Deduplication" and "Compression"?

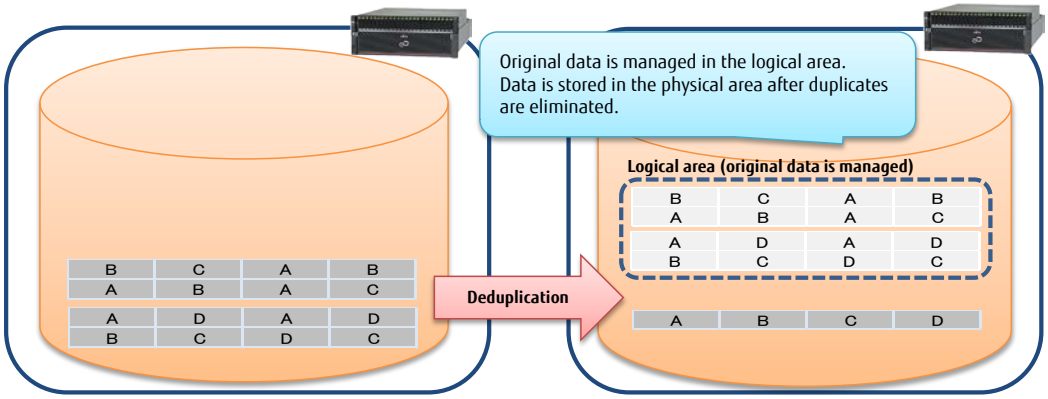
It is said that the amount of data in a typical company increases annually by 50% to 60%.
In addition, the need to protect data resources and to avoid risks place great importance on how to handle stored data.
However, backing up large amounts of data incurs high IT costs.
Under such circumstances, deduplication and compression technologies are attracting attention as a way to efficiently store data.

Both deduplication and compression are used for the same purpose of reducing the amount of data, but their mechanisms and operations are different. First, the general concept of the deduplication and compression technologies are explained below.

Deduplication is a technology that automatically analyzes and eliminates duplicate data with the same contents.
The deduplication function can significantly reduce the amount of transferred data and save storage space during backup operations.

Figure 1 shows a deduplication operation. Deduplication divides the target data into small data blocks and analyzes the blocks by comparing them to each other.
While updated blocks are stored in the physical area, the blocks that are recognized as duplicates are not stored in the physical area but are only associated with the source data in the logical area. As a result, storage space can be saved by storing data in the physical area after duplicates are eliminated.

Figure 1 Deduplication operation



"The size of the blocks" and "when and where the process is performed" are important factors for deduplication.
Depending on how these two factors are combined, deduplication efficiency changes significantly.

<Block sizes to detect duplicates>

■ Fixed-length block level method

Written data is divided into fixed size blocks (such as 4KB, 8KB, 16KB, and 128KB) and duplicates are eliminated.
Smaller block sizes enable a more granular comparison and provide significantly higher deduplication benefits.
However, the system load tends to be high because hash values are calculated when data is read or restored.

Table 1 Fixed-length block level method

Data							
A	B	C	D	E	F	G	H
4KB	4KB	4KB	4KB	4KB	4KB	4KB	4KB

Search by the unified block size (4KB)

■ Variable-length block level method

Written data is divided into variable size blocks, ranging from several KB to several hundred KB, and duplicates are eliminated.
Compared to the fixed-length block level method, this method can detect more duplicates.
However, depending on how the division of each variable-length block is detected, there may be specific cases where duplicates are not detected and the deduplication speed may become slow.

Table 2 Variable-length block level method

Data			
A	B	C	D
4KB	8KB	16KB	4KB

Search by multiple block sizes (4KB/8KB/16KB)

<Deduplication process>

■ Pre-process method

Duplicates are detected and eliminated in the server.
The amount of transferred data is small because it is transferred after the duplicates are eliminated. Therefore, the costs to secure bandwidth for transferring data can be reduced.
However, the server may experience heavy loads.

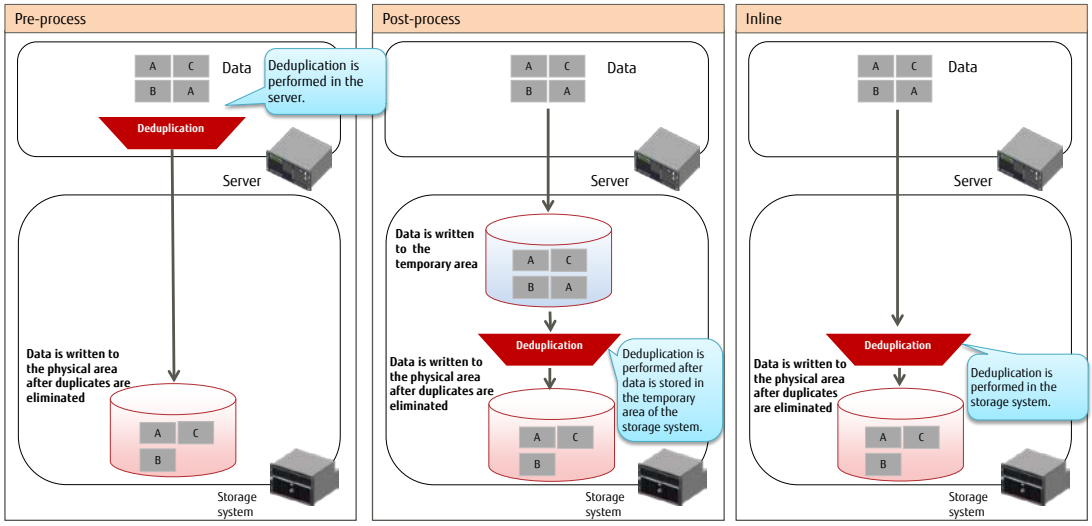
■ Post-process method

Duplicates are detected and eliminated in the storage system.
Storage system performance is less affected because duplicates are detected and eliminated after data is stored in the temporary area and the process is performed asynchronously with the I/Os.
However, the amount of transferred data is large and the I/O load is also high because data is transferred before duplicates are eliminated.
In addition, the initial investment is increased because a temporary area is necessary, therefore larger drive space and more drives are needed.
In designing a solution based on this method, the start and end times of the deduplication process and the drive space for the temporary area must be taken into consideration.

■ Inline method

Duplicates are detected and eliminated in the storage system when data is transferred from the server.
This method does not use a temporary area that is necessary for the post-process method, so implementation and operation of deduplication is easy to design.
However, the amount of transferred data is large and the I/O load is also high because data is transferred before duplicates are eliminated.

Figure 2 Pre-process, post-process, and inline methods



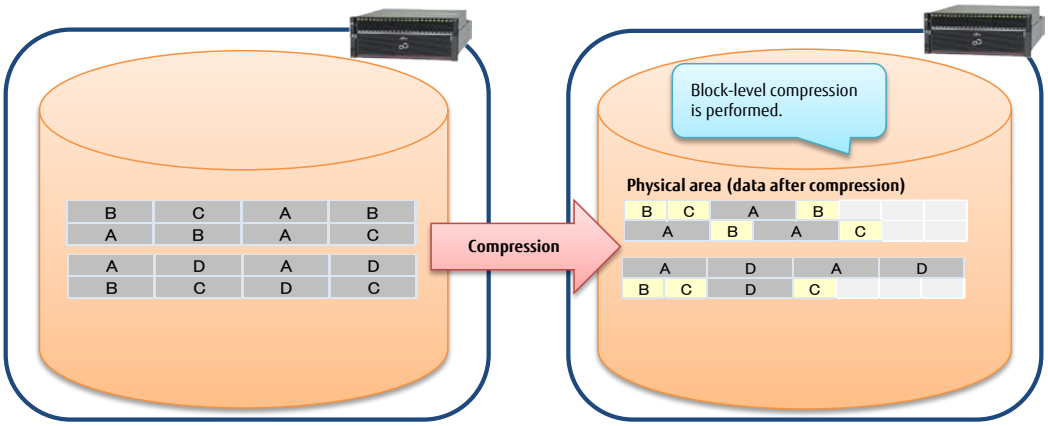
The following table shows the deduplication methods that are adopted by various vendor's products.

Table 3 List of deduplication methods that are used by various vendor's products

Product	Block size to detect duplications	Deduplication process
ETERNUS AF/DX	Fixed-length block level method (4KB)	Inline method
NetApp FAS2500 series	Fixed-length block level method (4KB)	Post-process method
EMC VNX series	Fixed-length block level method (8KB)	Post-process method
EMC XtremIO	Fixed-length block level method (4KB)	Inline method
Pure Storage FlashArray//m series	Fixed-length block level method (512byte)	Inline method
HP 3PAR StoreServ 7450 Storage	Fixed-length block level method (16KB)	Inline method

Next, compression is explained below.
Compression is a technology that reduces the size of data and converts it to different data while maintaining its nature.
Compression functions are associated with "ZIP" and "LHA" which are typical file compression functions, but the compression function in storage systems is completely different.
The compression function in storage systems compresses data after dividing it into small block sizes.
By reducing the size of the data that is to be saved, the compression function provides storage space efficiency.

Figure 3 Compression operation

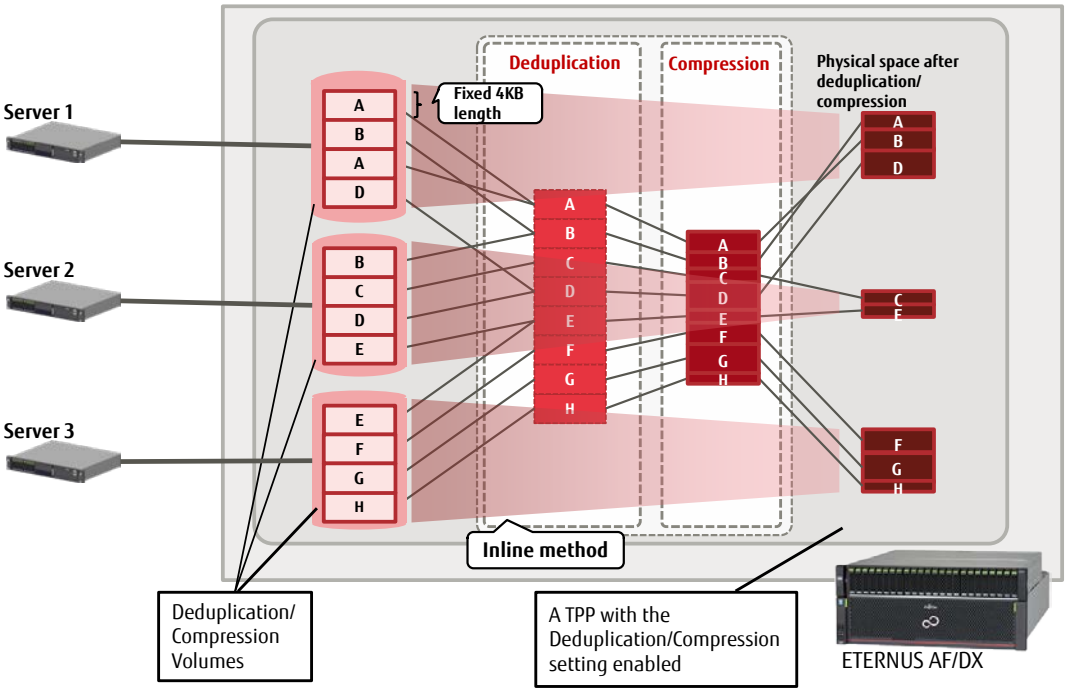


A big difference between compression and deduplication is that while deduplication is a function that deletes duplicated data blocks in the storage system, compression is a function that compresses different data blocks all at once. The data size can be further reduced by effectively combining the deduplication and compression functions since the effectiveness of data size reduction depends on the type of data.

1.2. Features of the ETERNUS AF/DX Deduplication/Compression Function

The ETERNUS AF/DX adopts the "fixed-length block level" method that uses small blocks of 4KB and the "inline" method for deduplication.
Because the inline method does not require a temporary area, implementation and operation of this function are easy to design and the initial investment cost can be reduced.
The ETERNUS AF/DX provides both the deduplication function and the compression function to achieve greater drive space efficiency.
The Deduplication/Compression function can further improve operational efficiency by being used to perform either a deduplication or a compression, depending on the data to be stored.

Figure 4 Deduplication/Compression function



Because the Deduplication/Compression function is intended to save storage space, the efficiency of the physical space is enhanced in a TPP based internal structure.
The data of all Deduplication/Compression Volumes in the TPP with the Deduplication/Compression setting enabled is scanned for duplicates.
The "SHA-1" hash algorithm is used to detect duplicates.
SHA-1 is a global standard hash function. Compared to MD5 and other hash functions, SHA-1 takes longer to scan but the detection accuracy is higher.
As a first step for deduplication, data is divided into blocks and a hash value is calculated for each block. If blocks have the same value, they are confirmed to have the exact same block information. If the blocks have already been stored in the physical area, the duplicate blocks are not stored in the physical area and only their configuration information is managed.
When the deduplicated data is invoked, data is restored by collecting the necessary blocks according to the configuration information.

■ Configuration management of the Deduplication/Compression function

Configuration information of the Deduplication/Compression function is managed for each pool.
Therefore, ETERNUS Web GUI and ETERNUS CLI also display the space saving ratio for each TPP.
(Hereinafter, the space saving ratio for each TPP with the Deduplication/Compression setting enabled is referred to as "space reduction rate".)
If 10TB of data is compressed to 3TB and then written, "70%" is displayed as the space reduction rate.

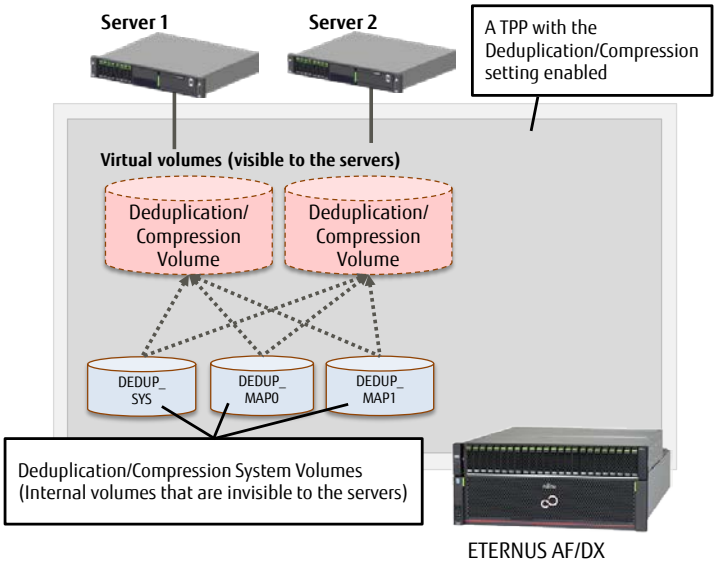
When the Deduplication/Compression setting is enabled, internal volumes that are invisible to the server (hereinafter referred to as "Deduplication/Compression System Volumes") are automatically created.
Deduplication/Compression System Volumes are internal volumes that manage the physical information and the configuration information after deduplication/compression is performed.
Deduplication/Compression System Volumes include DEDUP_SYS Volumes and DEDUP_MAP Volumes.

- DEDUP_SYS Volume
  A volume that stores the metadata and the actual data for which the Deduplication/Compression function is performed.

- DEDUP_MAP Volume
  A volume that stores the mapping table that is used to identify the physical location of the DEDUP_SYS Volume from Deduplication/Compression Volume numbers (or LBAs).

After the Deduplication/Compression System Volumes are created, create Deduplication/Compression Volumes (which are virtual volumes visible to the server).
Deduplication/Compression Volumes are virtual volumes used for storing data before a deduplication/compression is performed.

Figure 5 Configuration management of the Deduplication/Compression function



If volumes are created in a TPP with the Deduplication/Compression setting enabled and the Deduplication/Compression setting is enabled for those volumes, they become Deduplication/Compression Volumes.
Deduplication/Compression Volumes and non-Deduplication/Compression Volumes can coexist.

Table 4 List of the specifications for the Deduplication/Compression function

	ETERNUS AF/DX
Block length method	Fixed-length method (4KB)
Process method	Inline method
Setting target	Every pool
Target for a duplicate check	Data on all Deduplication/Compression Volumes in a TPP with the Deduplication/Compression setting enabled

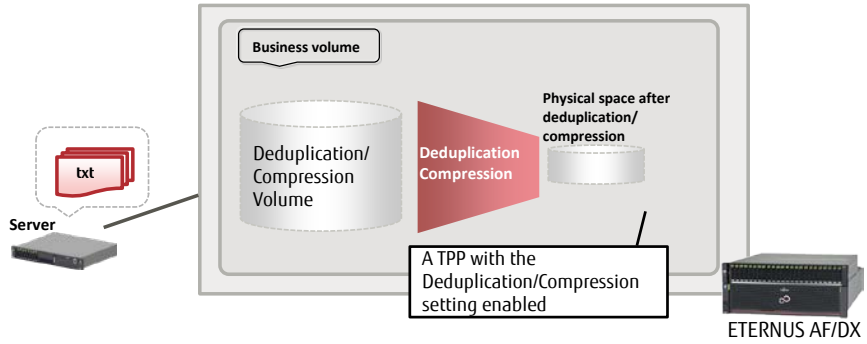
1.2.1. Operation Examples Using the Deduplication/Compression Function

The Deduplication/Compression function can efficiently save storage space in the following environments.

- Environments where document data and text data are frequently used
- Environments where many generations of full backups are created using functions such as OPC.
- Environments where the target volume for a full backup is updated infrequently

[Operation example 1 Saving storage space when the data size can be efficiently reduced]
In businesses where document data and text data are frequently used, the Deduplication/Compression function is highly effective for saving a significant amount of storage space.
In businesses where data is accessed mainly for reference, large amounts of reference data can be efficiently stored because the data reading of the Deduplication/Compression function affects the system performance less.

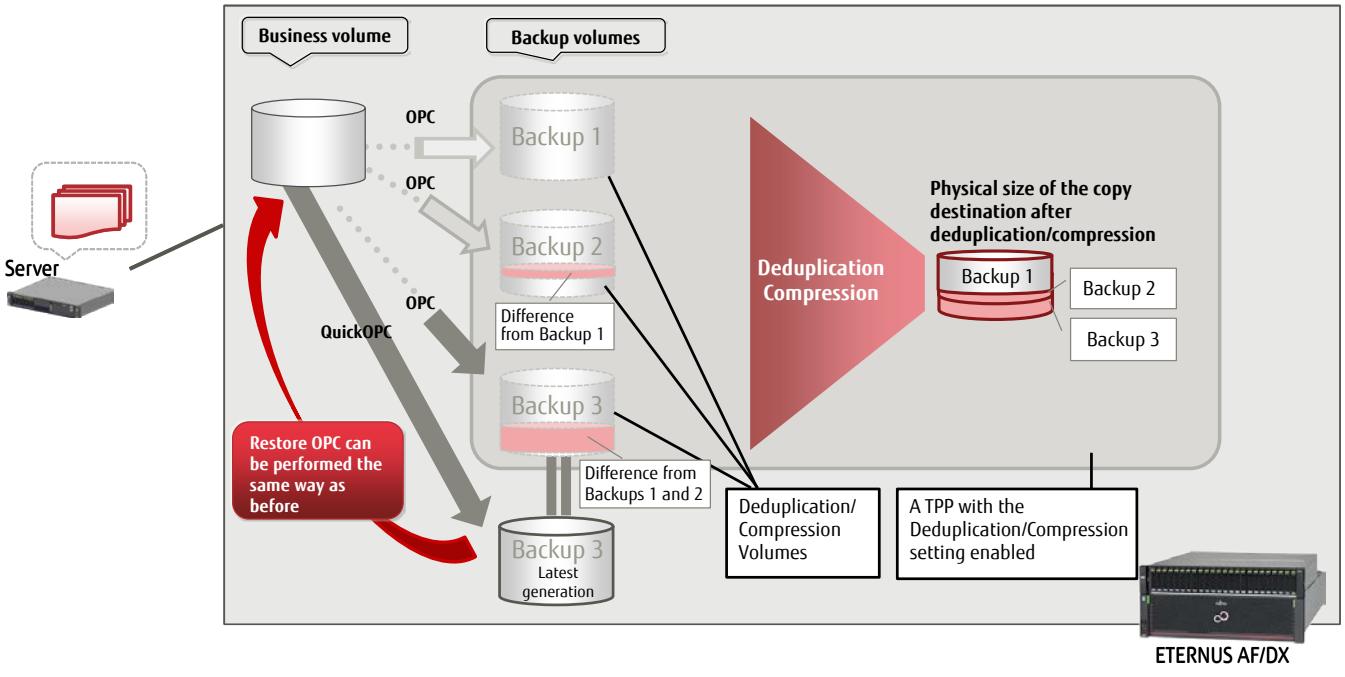
Figure 6 Reducing the physical size when the data size can be efficiently reduced



[Operation example 2 Saving storage space in an environment where backups are performed using OPC]
By using the Deduplication/Compression function in an environment where backups are performed using OPC, the amount of data to be stored can be significantly reduced.

Figure 7 is based on an OPC backup operation that saves three generations. Conventionally, the copy destination normally requires space for the number of generations that are to be saved. However, if the Deduplication/Compression function is used, the required space can be significantly reduced.
By using QuickOPC to perform a differential copy to the latest backup destination, not only is the response time of Restore OPC maintained the same as before but the restore time of the physical data is also maintained and data can be quickly restored.

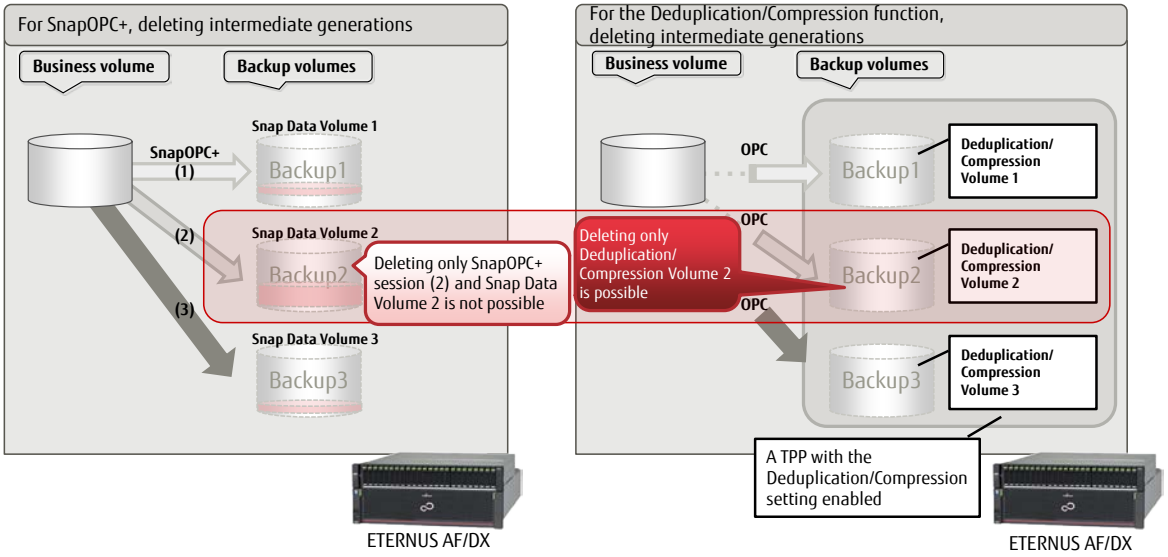
Figure 7 Reducing the physical size of the backup data in a backup environment



Backups using OPC with the Deduplication/Compression function also have the following advantages over generation backups using SnapOPC+.

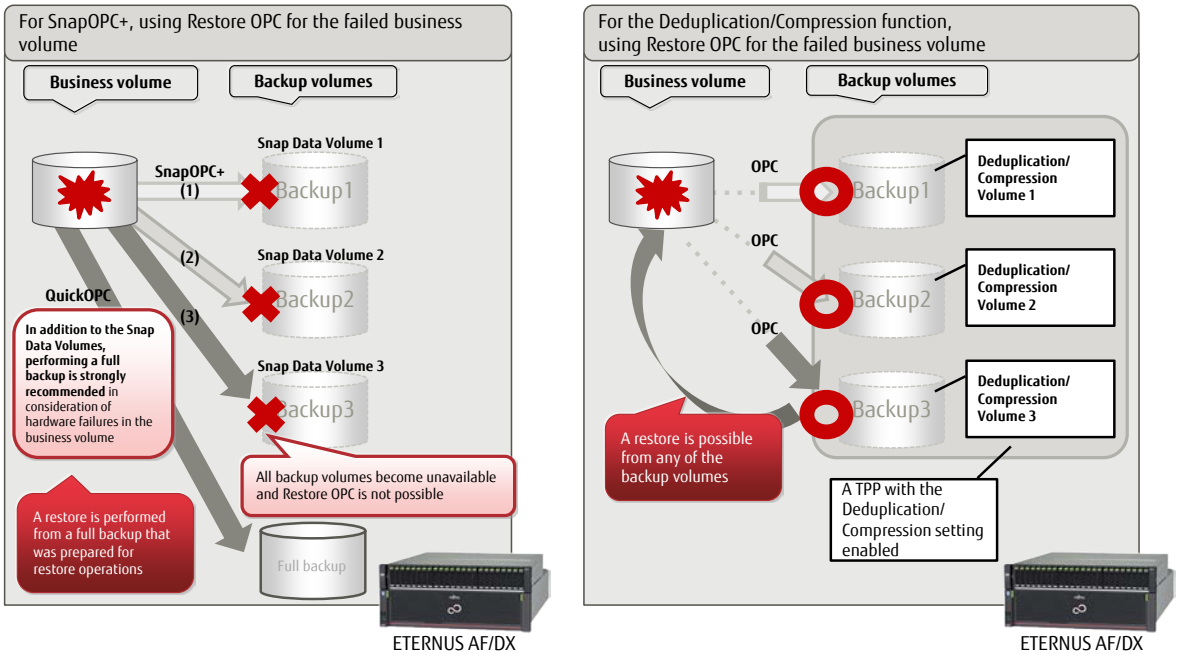
- SnapOPC+ backs up and manages differential data in chronological order. Therefore, even if users want to delete some of the generations due to various reasons such as drive space shortage, limited number of SnapOPC+ generations, or resource organization, deleting only the unnecessary generations is not available. The only option is either deleting all the generations prior to the target generation or keeping all generations.
  On the other hand, OPC with the Deduplication/Compression function always takes a full backup of all generations. Therefore, only the unnecessary generations can be deleted. If a generation is no longer needed due to drive space shortage or resource organization, users can delete just the target generation and manage backup volumes more flexibly.

Figure 8 Deleting intermediate generations of backup volumes



- For SnapOPC+, if a business volume fails, the backup volumes also become unavailable. Therefore, using QuickOPC, a separate full backup must also be kept to restore the failed business volume from the full backup using Restore OPC.
  On the other hand, for OPC with the Deduplication/Compression function, even if a business volume fails, backup volumes are still available. Because of that, a bull backup can be restored from any backup destination volume. This eliminates the need to design an operation that takes business volume failures into consideration.

Figure 9 Restoring failed business volumes



1.3. Items to Consider before Implementing the Deduplication/Compression Function

The Deduplication/Compression function reduces the performance of the storage system while efficiently saving storage space.
It is important to balance these two conflicting needs before enabling the Deduplication/Compression function.
Implement the function after determining whether the benefit of saving space with a degradation in performance of the target system is acceptable.
The following items affect the Deduplication/Compression function.

- Data

Types of data change (such as file addition, data update, and data deletion), percentage of data change, file size, and available file compression ratio.
Drive images of multiple virtual machine systems and mail data have a lot of redundant data because they generally have a specific structure and format. As a result, more duplicates can be detected and more storage space can be saved.
Data in an uncompressed format (such as text, static images, and audio) is a type of data that has a high compression ratio and the size can be efficiently reduced.
However, the Deduplication/Compression function is not effective when data is frequently updated or in environments where data is used frequently in a compressed format (such as LZH, ZIP, and CAB).

- Performance

Data access frequency and I/Os (sequential or random, size, and pattern).
The Deduplication/Compression function is effective in backup environments where generations are managed because the storage system performance is less affected.
However, the Deduplication/Compression function may not be suitable for batch processes and online business environments because the storage system performance is greatly affected.

- Hardware

The CPU use rate, cache memory size, and type and speed of the drive.
Since the Deduplication/Compression function places a heavy load on the performance of the storage system, refer to the following requirements to estimate an optimal design configuration and applicable model.

1.4. Requirements

There are several requirements for the Deduplication/Compression function.
The following conditions need to be met in order to achieve an optimal configuration.

- Deduplication/Compression mode

To use the Deduplication/Compression function, it must be set in the ETERNUS AF/DX.
Enable the Deduplication/Compression mode in the ETERNUS AF/DX.

- Memory Extension

To use the Deduplication/Compression function, Memory Extension must be installed.

Refer to Table 5 to check the Memory Extension capacity.

Table 5 List of Memory Extension capacities for each model

Model	Memory Extension capacity
ETERNUS AF250 ETERNUS AF250 S2	64GB system memory/storage system
ETERNUS AF650	256GB system memory/storage system
ETERNUS AF650 S2	512GB system memory/storage system
ETERNUS DX200 S3	24GB/CM if the cache memory capacity is 8GB/CM Maximum 48GB/storage system
ETERNUS DX200 S4	16GB/CM if the cache memory capacity is 16GB/CM Maximum 32GB/storage system
ETERNUS DX500 S3	24GB/CM if the cache memory capacity is 8GB/CM 24GB/CM if the cache memory capacity is 16GB/CM 32GB/CM if the cache memory capacity is 32GB/CM Maximum 64GB/storage system
ETERNUS DX500 S4	32GB/CM if the cache memory capacity is 8GB/CM or 16GB/CM 64GB/CM if the cache memory capacity is 32GB/CM 128GB/CM if the cache memory capacity is 64GB/CM Maximum 256GB/storage system

ETERNUS DX600 S3	48GB/CM if the cache memory capacity is 16GB/CM 48GB/CM if the cache memory capacity is 32GB/CM 64GB/CM if the cache memory capacity is 64GB/CM Maximum 128GB/storage system
ETERNUS DX600 S4	64GB/CM if the cache memory capacity is 16GB/CM or 32GB/CM 128GB/CM if the cache memory capacity is 64GB/CM 256GB/CM if the cache memory capacity is 128GB/CM Maximum 512GB/storage system

For the ETERNUS AF series, the above values indicate the fixed capacity of the system memory (including the size equivalent to the Memory Extension capacity).

■ Supported models

ETERNUS AF series
ETERNUS DX S4/S3 series
(However, the ETERNUS DX60 S4/DX60 S3, the ETERNUS DX100 S4/DX100 S3, and the ETERNUS DX8000 S3 series do not support this function.)

■ Supported volume types

TPV (only SAN volumes are supported. NAS volumes are not supported.)

■ Maximum number of TPPs where the Deduplication/Compression setting can be enabled

The maximum number varies depending on the model.
Refer to Table 6 to check the maximum number.

Table 6 List of the maximum number of TPPs where the Deduplication/Compression setting can be enabled

	ETERNUS AF250 S2/AF250	ETERNUS AF650 S2/AF650	ETERNUS DX200 S4/S3	ETERNUS DX500 S4/S3	ETERNUS DX600 S4/S3
Maximum number of TPPs where the Deduplication/Compression setting can be enabled	4	8	4	4	8

1.5. Notes

There are several notes for the Deduplication/Compression function.

■ Changing existing volumes to Deduplication/Compression Volumes

Existing volumes (Standard/TPV/FTV/WSV) can be changed to and from Deduplication/Compression Volumes with RAID Migration.
When changing existing volumes to Deduplication/Compression Volumes, specify a Deduplication/Compression Volume as the migration destination. Deduplication/Compression Volumes in the same TPP or a different TPP can be specified.

■ Backup

When a remote copy (REC) or a backup/restore is performed by setting a Deduplication/Compression Volume as the copy source, data is transferred without being deduplicated and compressed.
If the copy destination is a Deduplication/Compression Volume, data is stored after the Deduplication/Compression function is performed.

■ Deduplication/Compression System Volumes and the total logical capacity of the Deduplication/Compression Volumes in a TPP

The following table shows the specifications of the Deduplication/Compression System Volumes.
In addition, the total logical capacity of the Deduplication/Compression Volumes is limited by the logical capacity of the DEDUP_SYS Volume in the same TPP.

- If the TPP consists of two or more RAID groups, the upper limit is 10 times the logical capacity of the DEDUP_SYS Volume.
- If the TPP consists of a single RAID group, the upper limit is 5 times the logical capacity of the DEDUP_SYS Volume.

Refer to Table 7 to check the specifications of the Deduplication/Compression System Volumes and the upper limit of the total logical capacity of the Deduplication/Compression Volumes.

Table 7 List of Deduplication/Compression System Volume specifications

	DEDUP_SYS Volume	DEDUP_MAP Volume
Volume type	TPV	TPV
Number of volumes	One in each TPP with the Deduplication/Compression setting enabled	The number varies depending on the number of RAID groups in the TPP with the Deduplication/Compression setting enabled. * One when the TPP only has one RAID group * Two when the TPP has two or more RAID groups
Logical capacity of a volume	8TB (default) to 128TB	5.38TB (fixed)

2. Test

This chapter provides two test results to show how the Deduplication/Compression function is effective in saving storage space.

2.1. Testing the Access Performance when the Deduplication/Compression Function is Used

This test assumes that the data partition of the Windows OS (or a partition other than the C drive), which is located in the ETERNUS AF/DX, is accessed.
This test verifies the access performance when the Deduplication/Compression function is used in the partition.

2.1.1. Purpose

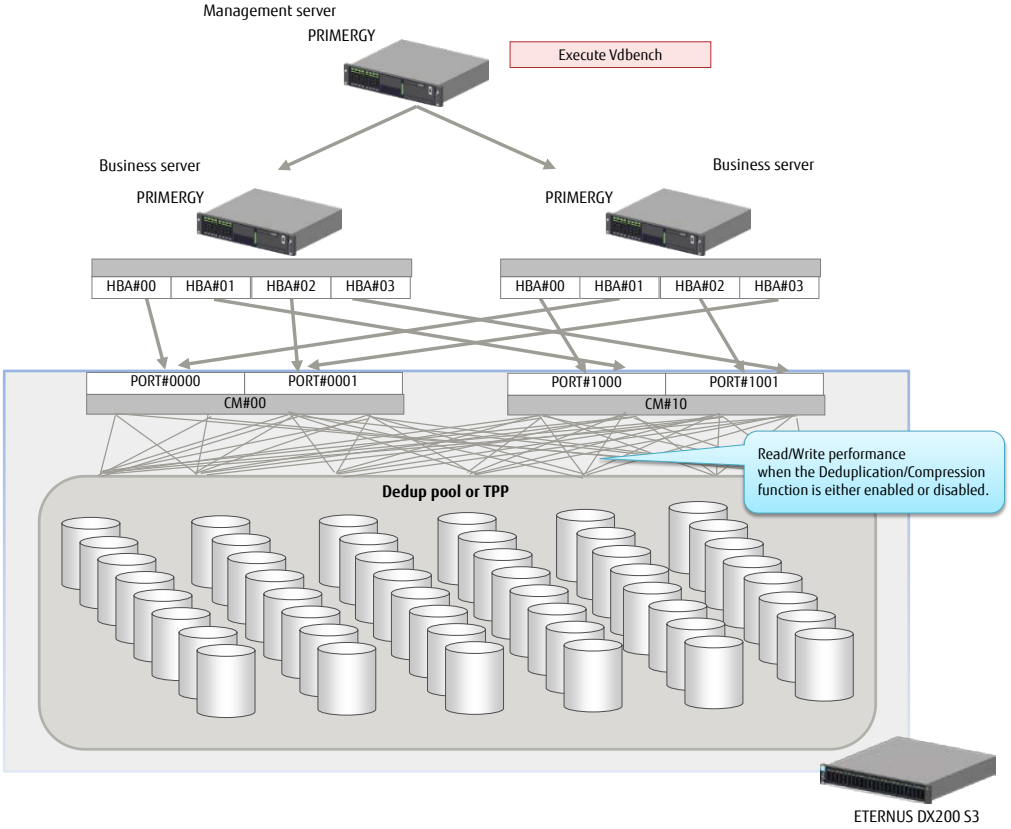
The purpose of this test is to measure the performance when the Deduplication/Compression function is implemented.

2.1.2. Details

In this test, the partition is accessed with a Read:Write ratio of 3:1 to mainly reference files and in some cases, to update files.
This test measures the performance of two systems: one that performs typical access to the partition and the other that mainly accesses small data.
The average access size is 8KB for the system that performs typical access to the partition and 4KB for the system that mainly accesses small data.
While the Deduplication/Compression function can save storage space, it may reduce the performance depending on the I/0 load. Therefore, the test results, which are described later, are provided for cases where the Deduplication/Compression function is either enabled or disabled.
vdbench is used as the tool to access the partition.

2.1.3. Environment

Figure 10 Performance test when the Deduplication/Compression function is either enabled or disabled



2.1.4. Prerequisites

The following table shows the storage system settings.

Table 8 Prerequisites of the access performance test

Target storage system	ETERNUS DX200 S3
Deduplication mode	Enabled
Pool	Enable the Deduplication/Compression function: Create Deduplication/Compression Volumes in a TPP with the Deduplication/Compression setting enabled. Disable the Deduplication/Compression function: Create TPVs in a TPP with the Deduplication/Compression setting disabled.
RAID	Create four RAID5(4+1) groups (each group consists of five HDDs).
Logical volume	Create 16 volumes.
Path	Connect each CM to two business servers in a four path configuration (eight paths in total).

2.1.5. Results

For the system that performs typical access to the partition

Table 9 Results for the typical system (Read: Write=3:1)

Deduplication/Compression	Mean access size	Throughput	IOPS	Space reduction rate
Enabled	8K	14.1MB/s	1803	60%
Disabled	8K	26.8MB/s	3425	0%

For the system that mainly accesses small data

Table 10 Results for the system that mainly accesses small data (Read: Write=3:1)

Deduplication/Compression	Mean access size	Throughput	IOPS	Space reduction rate
Enabled	4K	9.6MB/s	2457	60%
Disabled	4K	15.2MB/s	3890	0%

2.1.6. Summary

As the results show, adequate performance is ensured considering the effectiveness of space reduction even if the Deduplication/Compression function is used.
This function is useful for users who place a higher priority on space efficiency in systems for storing large amounts of files with a high space reduction rate.
Before using the Deduplication/Compression function, consider the balance between space efficiency and performance.

2.2. Testing the Space Reduction Rate for the Data Partition

This test verifies the space reduction rate for the various types of data files to be accessed. The data partition of the Windows OS (or a partition other than the C drive) that stores the data files is located in the ETERNUS AF/DX.

2.2.1. Purpose

The purpose of this test is to measure the space reduction rate for various types of data files when the Deduplication/Compression function is used.

2.2.2. Details

This test measures the space reduction rate for two cases: one where the following various types of data files are stored in a single directory of the data partition (there are no duplicate files), and the other where the same files are stored in two directories (there are duplicate files because the directory is copied).

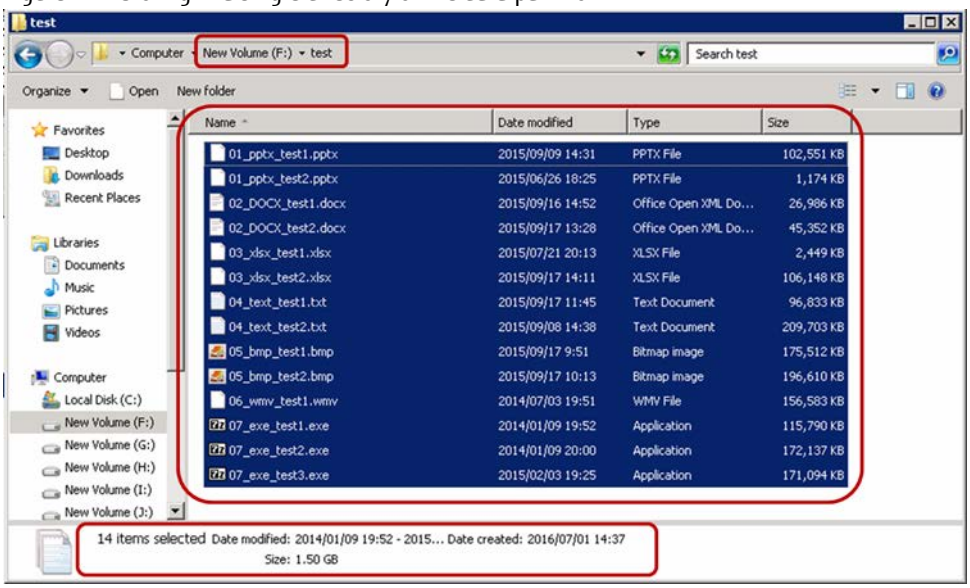
Seven types of data files with a total capacity of 1,539MB are used for the test. The details of the data types and file sizes are as follows.

Table 11 List of data types and files

	Data type	File	Extension	No. of files	Total file size
1	Document data	Power point file	pptx	2	101MB
2		Word file	docx	2	70MB
3		Excel file	xlsx	2	106MB
4		Text file	txt	2	299MB
5	Image data	Bitmap file	bmp	2	363MB
6	Video data	Video file	wmv	1	152MB
7	Program data	Executable file	exe	3	448MB
			Total	14	1,539MB

<Case when storing in a single directory of the data partition (without duplicates)>

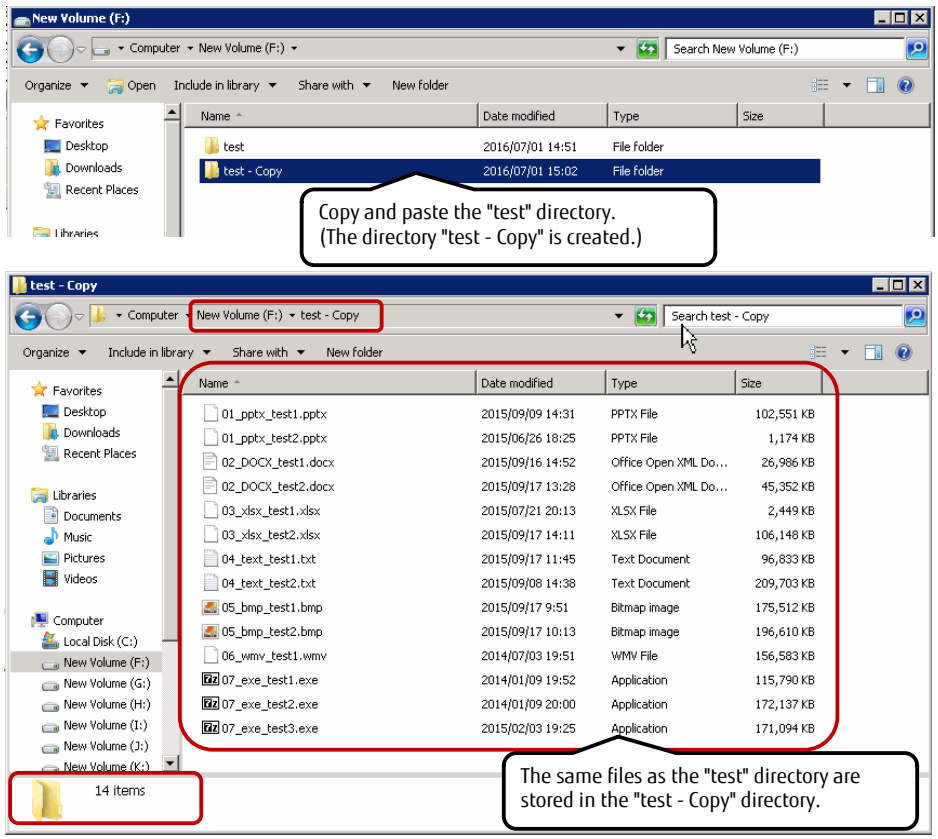
Figure 11 Storing in a single directory of the data partition



The above figure shows a screenshot of when files are stored in the "test" directory of the data partition (F:\).
(A selection of 14 files with a total file size of 1,539MB [1.50GB])

<Case when storing the same files in two directories (with duplicates)>

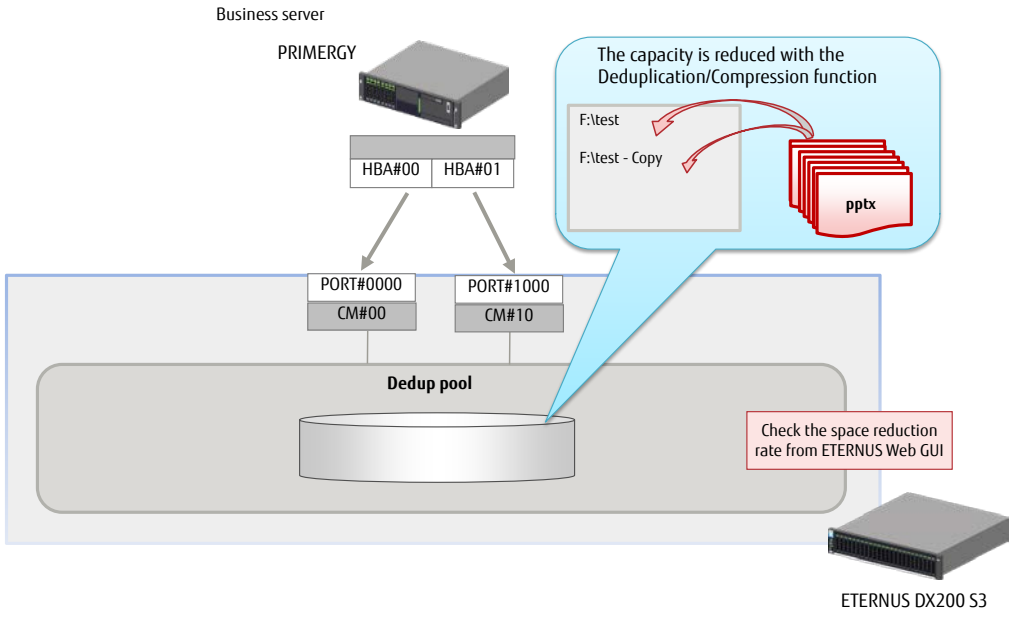
Figure 12 Storing the same files in two directories of the data partition



The above figure shows a screenshot of when the same files are stored in the "test" directory and in the "test - Copy" directory of the data partition (F:\).

2.2.3. Environment

Figure 13 Verifying the space reduction rate for the data partition in the test configuration



2.2.4. Prerequisites

The following table shows the storage system settings.

Table 12 Prerequisites of the test for the space reduction rate

Target storage system	ETERNUS DX200 S3
Deduplication mode	Enabled
Pool	Create a 512GB TPP with the Deduplication/Compression function enabled.
RAID	Create a RAID1(1+0) group that consists of four HDDs.
Logical Volume	Create a 256GB Deduplication/Compression Volume with the Deduplication/Compression function enabled. Make the server recognized this Deduplication/Compression Volume and perform a quick format. Create a basic partition of 262,014MB (255.87GB) in the volume. Assign "F:\" as the drive letter.
Space reduction rate	Refer to the values displayed when "Enable" is set for [Deduplication/Compression] in [Thin Provisioning Pool] with ETERNUS Web GUI.

- The space reduction rate is a ratio of the data capacity that is deleted by the Deduplication/Compression function to the logical data capacity that is written.
- Each measurement is conducted by recreating the TPP with the Deduplication/Compression setting enabled to initialize the data area and after the Deduplication/Compression Volume is re-recognized by the server.

2.2.5. Results

Case when storing in a single directory (F:\test) of the partition (without duplicates)

Table 13 Results when files are stored in a single directory (F:\test) of the partition

Size of the TPP with the Deduplication/Compression setting enabled	Deduplication/Compression Volume size	Partition size	Destination location and stored file size	Space reduction rate	[Stored file size x (1 - space reduction rate/100)] (*1)
512GB	256GB	255.87GB	**1,539MB** in F:\test	32%	**1,047MB**

Case when storing the same files in two directories (F:\test and F:\test - Copy) of the partition (with duplicates)

Table 14 Results when the same files are stored in two directories (F:\test and F:\test - Copy) of the partition
(The whole F:\test directory is copied to create the F:\test -Copy directory)

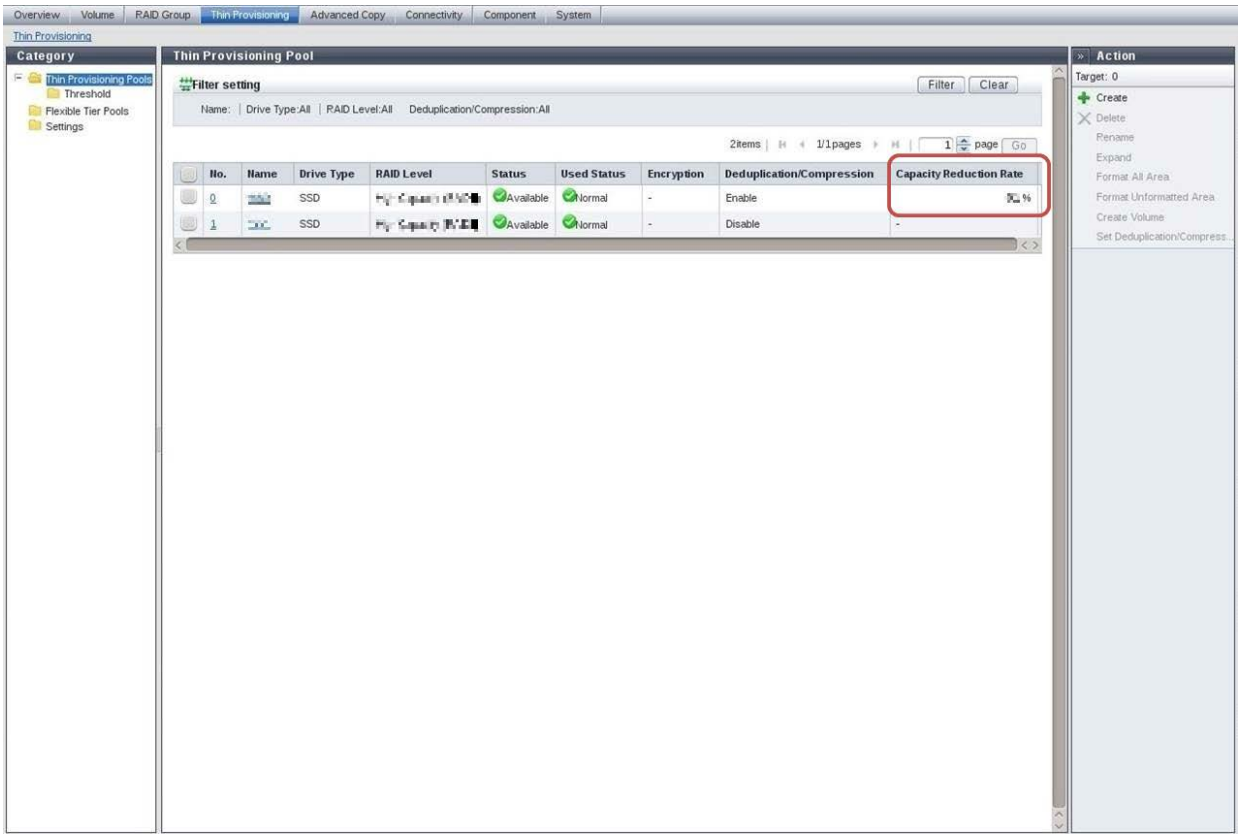
Size of the TPP with the Deduplication/Compression setting enabled	Deduplication/Compression Volume size	Partition size	Destination location and stored file size	Space reduction rate	[Stored file size x (1 - space reduction rate/100)] (*1)
512GB	256GB	255.87GB	1,539MB in F:\test, 1,539MB in F:\test - Copy Total **3,078MB**	65%	**1,077MB**

(*1) This formula can be used to calculate the approximate value after a space reduction.

2.2.6. Display Example of the Space Reduction Rate

The following ETERNUS Web GUI screenshot shows an example of the space reduction rate.
When [Deduplication/Compression] is [Enable], the value is displayed under [Capacity Reduction Rate].
The space reduction rate is an instantaneous value of the TPP with the Deduplication/Compression setting enabled and excludes the data of unfinished writing processes in the server such as when the cache of the server OS is not flashed.

Figure 14 Display example of the space reduction rate
(The following screen is only for reference and is not relevant when the configuration and values are being verified.)



2.2.7. Summary

As the results show, the function is beneficial for managing documents and image files in locations such as file servers and VDIs that store the same files.
In addition, using this function reduces the storage capacity without user intervention.
Furthermore, by reducing the storage space with the function, the size and the number of data writes to the physical area can be reduced.
This improves the overall durability of hardware such as hard drives and flash memory, which leads to improvements in reliability.

3. Conclusion

The Deduplication/Compression function is effective in the following environments.

- A backup environment that manages a large number of generations
- An environment that manages large capacity backup data as a disaster recovery measure or for complying with laws and regulations
- An environment that stores a large number of files that have a high space reduction rate

Reducing storage space and providing efficient use, the Deduplication/Compression function plays an important role in reducing initial investment costs and implementation costs as well as storage and installation costs.
Refer to this document to identify the advantages of using this function and then apply it to a suitable environment.
The Deduplication/Compression function contributes considerably to system cost reduction.

Contact
FUJITSU Limited
Website: http://www.fujitsu.com/eternus/