White Paper
Effects of the Deduplication/Compression Function in Virtual Platforms
ETERNUS AF series and ETERNUS DX S4/S3 series

Table of Contents

Preface 3

1 Virtual Platform Considerations 4
- 1.1 Efficiency Brought on by Virtualization 4
- 1.2 Improving the Rate of Storage System Consolidation 4

2 Drive Usage and Space Reduction with Deduplication/Compression in Virtual Platforms 5
- 2.1 Drive Usage on the Storage Systems in Physical Environments and Virtual Environments 5
- 2.2 Effects of Deduplication/Compression in Virtual Platforms 6

3 Verifying the Effects of Space Reduction Provided by the Deduplication/Compression Function 7
- 3.1 Verifying the Effects of Space Reduction 7
  - 3.1.1 Reducing the Drive Usage by Compression 8
  - 3.1.2 Reducing the Drive Usage by Deduplication 9
  - 3.1.3 Effects of Deduplication on Continuous Operations 10
- 3.2 Measuring the Effects of Space Reduction 11
  - 3.2.1 Measuring the Effects of Space Reduction by Compression 12
  - 3.2.2 Measuring the Effects of Space Reduction by Deduplication 13
  - 3.2.3 Measuring the Effects of Deduplication/Compression on Continuous Operations 14
- 3.3 Deduplication/Compression and Performance 15
  - 3.3.1 Comparison of Startup Times for Windows Server 2016 15
  - 3.3.2 Time Comparison for Storage vMotion 16

4 Summary 17

List of Figures and Tables

Figure - 1 Overview of deduplication effectiveness 4
Figure - 2 Expanded drive space in a virtual environment 5
Figure - 3 Deduplication between different platforms 6
Figure - 4 Overview of the storage area used for virtualization and deduplication/compression 7
Figure - 5 Overview of the effects of drive usage reduction provided by compression 8
Figure - 6 Overview of the effects of drive usage reduction provided by deduplication 9
Figure - 7 Overview of the effects of drive usage reduction provided by deduplicating the files in the OS of the guests 10
Figure - 8 Virtual disk space and storage system space 12
Figure - 9 Amount of physical space used after multiple Windows Server machines are deployed 13
Figure - 10 Sum of the increased physical space after applying the Service Pack 14
Figure - 11 Comparison of startup times when the Deduplication/Compression function is enabled and disabled 15
Figure - 12 Time comparison for Storage vMotion when the Deduplication/Compression function is enabled and disabled 16

Table - 1 Comparison of physical drive areas 8
Table - 2 Drive spaces allocated to Windows Server 2016 11
Table - 3 Physical drive space and reduction rate after a Windows Server machine is deployed 12
Table - 4 Physical drive space and reduction rate after multiple Windows Server machines are deployed 13
Table - 5 Common file updates on the guest OS by applying the Service Pack 14

Preface

Most systems have already been virtualized and the benefits, such as effective use of resources and spaces, have been provided. However, there is still room for improvement.
This document describes the effects of the Deduplication/Compression function provided by a storage system in virtual platforms by providing the test results of a space reduction in a configuration using VMware vSphere and Microsoft Windows virtual machines.
The verification results described in this document are current as of January 2017.

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

■Target Readers

This document targets readers who have the following knowledge.

- General concepts of storage systems and virtual platforms

Some parts of this document include information on VMware vSphere and file systems (block management).

■Target Models

This document targets the following storage system models.

- FUJITSU Storage ETERNUS AF250 S2/AF650 S2 and ETERNUS AF250/AF650
- FUJITSU Storage ETERNUS DX200 S4, ETERNUS DX500 S4/DX600 S4, ETERNUS DX200 S3, and ETERNUS DX500 S3/DX600 S3 *1

*1 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.

■Abbreviations

The following abbreviations are used in this document.

- ETERNUS AF series ...... FUJITSU Storage ETERNUS AF250 S2/AF650 S2 and ETERNUS AF250/AF650 (All-Flash Arrays)
- ETERNUS DX series ...... FUJITSU Storage ETERNUS DX200 S4, ETERNUS DX500 S4/DX600 S4, ETERNUS DX200 S3, and ETERNUS DX500 S3/DX600 S3
- ETERNUS AF/DX............. FUJITSU Storage ETERNUS AF series and ETERNUS DX S4/S3 series
- TPP .... Thin Provisioning Pool
- TPV .... Thin Provisioning Volume
- HDD .... Hard Disk Drive
- SSD .... Solid State Drive
- VM .... Virtual Machine
- vSphere .... VMware vSphere
- vCenter Server .... VMware vCenter Server
- ESXi host .... VMware vSphere Hypervisor (VMware ESXi)
- VMFS .... Virtual Machine File System (VMware virtual machine file system)
- .vmdk .... Virtual Machine Disk (VMware virtual disk file)
- .vmx .... Virtual Machine Configuration File (VMware virtual machine configuration file)
- .vswp .... Virtual Machine Swap File (VMware virtual machine swap file)

1 Virtual Platform Considerations

1.1 Efficiency Brought on by Virtualization

Virtualization has become generalized because virtual environments are easily acquirable thanks to cost-free hypervisors, open source software to configure cloud environments, and widespread public cloud services. Most systems have been virtualized except for those with license and hardware restrictions, and those that require extremely high performance and extremely large capacities.

Because one of the advantages of virtualization that has helped popularized it is easy implementation, the number of virtual machines (VM) keeps increasing in situations where development/test environments and clones are created, as well as in situations where existing systems are migrated and new systems are configured.
Because of that, there is an obvious need for further efficiency.

1.2 Improving the Rate of Storage System Consolidation

In virtual environments, there are many similar VM environments in which clones are created and the same OS is installed.

In virtual desktop environments, Linked Clone of VMware Horizon View *2 and Data Deduplication of Microsoft Windows Server *3 are also provided. These features reduce the storage space by removing duplicated data areas. However, these features are implemented in specific middleware or within a file system, and have a limited applicable range.

The Deduplication/Compression function of the ETERNUS AF/DX provides the benefits of deduplication to the overall virtual platform without, for example, limiting the range to the machine, middleware, or the OS. For this reason, deduplication works effectively in environments where many VMs with the same data exist.
In addition, deduplicated data is compressed before being saved to the physical drives. This reduces the physical areas even more.

*2 Feature for integrating and deleting OS areas of similar VMs
*3 Feature for reducing space by detecting duplicated data and not saving it

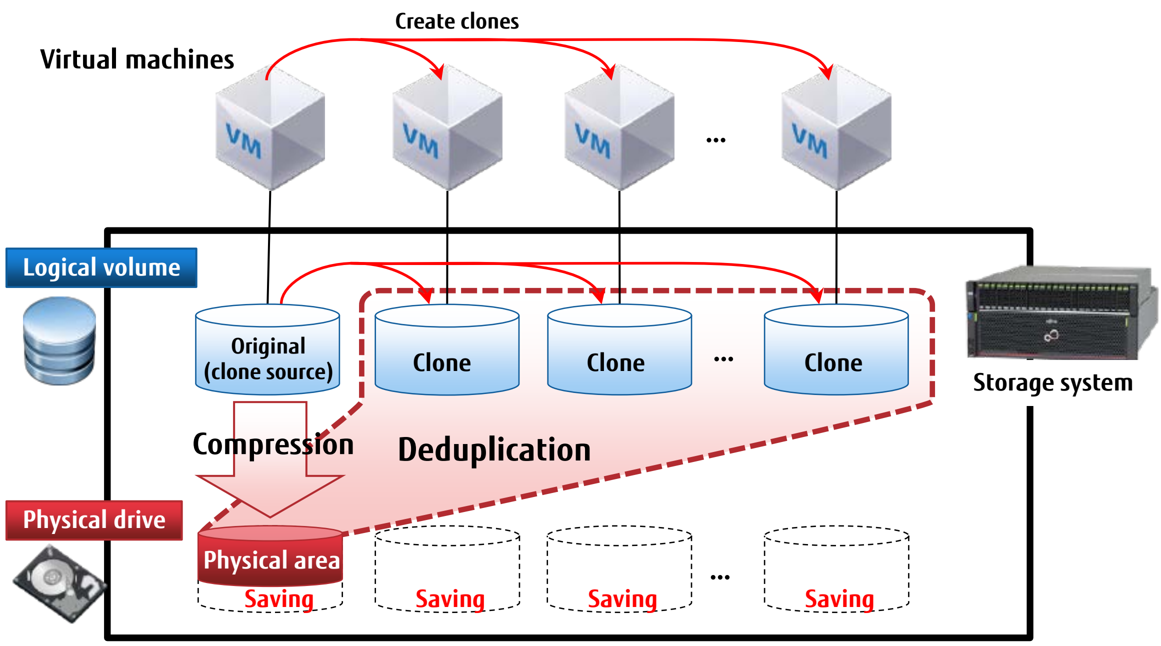


Figure - 1 Overview of deduplication effectiveness

By using the Deduplication/Compression function of the ETERNUS AF/DX, virtual platforms can be further consolidated.
The Deduplication/Compression function is implemented in the storage system and the volumes are available from the servers in the same way as standard volumes. In addition, this function is easy to use without complicated operations during installation.

2 Drive Usage and Space Reduction with Deduplication/Compression in Virtual Platforms

2.1 Drive Usage on the Storage Systems in Physical Environments and Virtual Environments

When the drive space in the physical machines is stored in a Thin Provisioning Pool of a storage system (hereinafter called the physical environment), the physical drive space used is equivalent to the overall data amount that is written to the OS file system.

In a virtual environment, in addition to the drive capacity of the VMs, the file system (VMFS) space where the VMs are deployed and the information that the hypervisor uses to manage the VMs are stored in the physical drive space.
For VMware vSphere, datastore file systems (VMFS), VM configuration (.vmx) files, and swap (.vswp) files are stored in the physical drive space. (In addition, files created by operations such as snapshot and suspend are also stored there.)

When the drive space is virtualized and stored in a Thin Provisioning Pool, a larger physical drive space is used in comparison to when the physical machine's drive space is stored in a Thin Provisioning Pool.

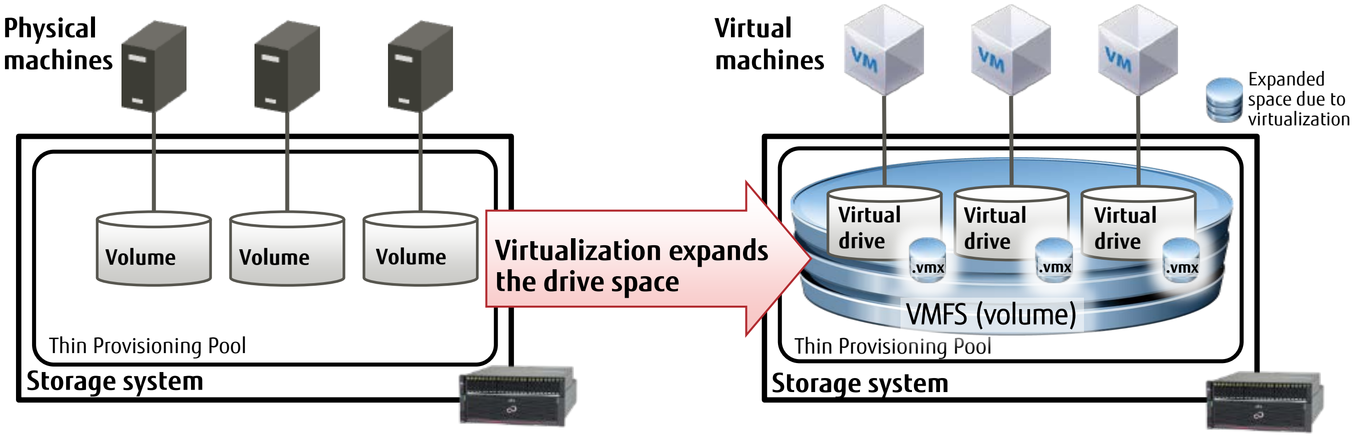


Figure - 2 Expanded drive space in a virtual environment

2.2 Effects of Deduplication/Compression in Virtual Platforms

In order for systems to handle increases in drive usage, the drive space must have free space.
Physical machines have a much larger free space because each system must have its own free space and the free space is increased every time HDDs are added as drive space. This also applies to Standard volumes of a storage system.

The Thin Provisioning function virtualizes drives and allows only the data storage area to be stored in the physical drives. This function also saves physical drive space by allowing the free space to be shared. The efficiency of a space reduction with Thin Provisioning can be estimated from the drive's free space.

Deduplication/compression can further reduce the physical drive space necessary for the Thin Provisioning function.
The Deduplication/Compression function of the ETERNUS AF/DX operates on Thin Provisioning Pools. If duplicate data is stored in a Thin Provisioning Pool with the Deduplication/Compression setting enabled, it is not written in the physical space but the existing physical space is referenced instead. In addition, non-duplicate data is compressed when written in the physical space thereby saving more space.

The Thin Provisioning function interoperates with the OS and hypervisor and saves the physical drive space for unused spaces and drive space where data contains all zeros. However, the Thin Provisioning function of a storage system manages data in large units of MB. Therefore, the physical space can be saved only if several dozen MBs of unused space exist in sequence.

Many recent file systems such as Microsoft Windows' NT File System (NTFS) and Linux's ext4 use 4KB as the default block size. The Deduplication/Compression function of the ETERNUS AF/DX also processes data in units of 4KB. Therefore, in a virtual environment where multiple VMs are installed with the same OS, deduplication works effectively not only on clones (entire copies of VMs) but also on VM files because alignments match even at the file level. In this ways, more space can be saved by finely reducing unused space and duplicates.

The Deduplication/Compression function operates within Thin Provisioning Pools. Therefore, the benefits of deduplication can be obtained by storing the volumes in the same pool even if different environments, such as volumes for physical machines and volumes for VMware vSphere datastores, are mixed together.
Deduplication/compression also applies to the space increased by virtualization as shown in "2.1 Drive Usage on the Storage Systems in Physical Environments and Virtual Environments".

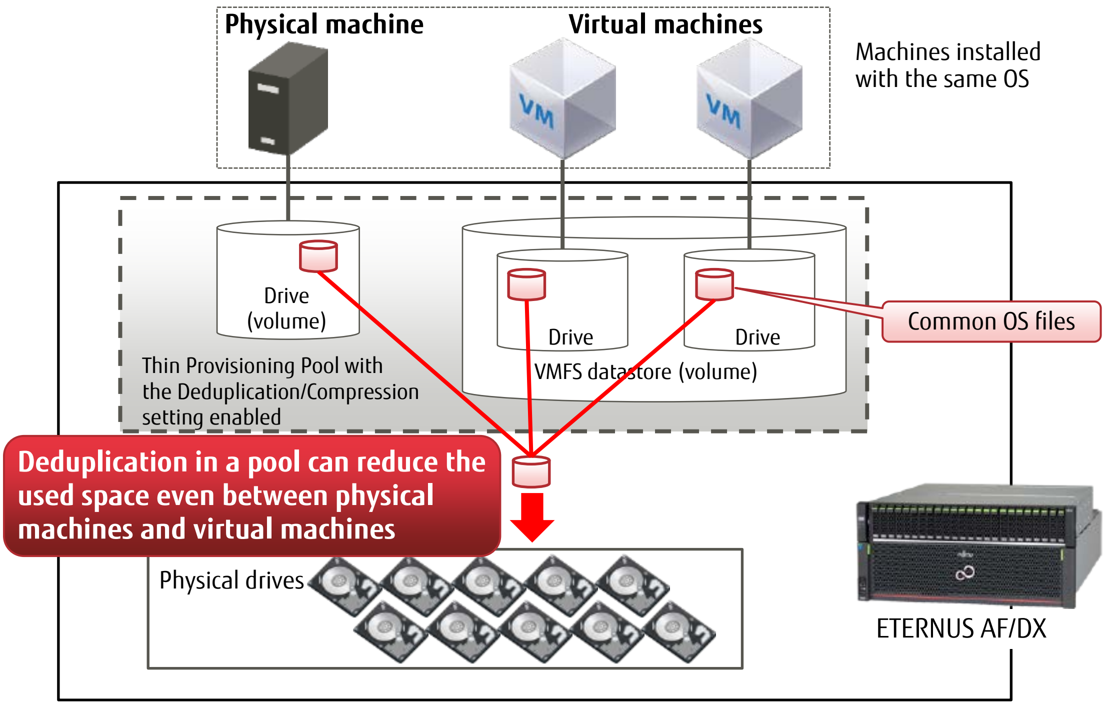


Figure - 3 Deduplication between different platforms

The Deduplication/Compression function performs the process in the storage system to reduce the drive space. This means that the function uses the storage system's resources and requires an extra process between the servers and HDDs/SSDs, which impacts the performance.
In environments where the Deduplication/Compression function is effective, not only is the used drive space reduced, but access to the HDDs/SSDs is required less. However, in environments where the Deduplication/Compression function is not effective, the amount of processes increases.
Because the same OS file exists in each VM, a reduction in the amount of space to be used can be greatly expected in the virtual platform using deduplication/compression.

3 Verifying the Effects of Space Reduction Provided by the Deduplication/Compression Function

This chapter verifies the effects of space reduction provided by the Deduplication/Compression function of the ETERNUS AF/DX. The verification uses an environment where VMware vSphere 6.5 is used for the virtual platform and Microsoft Windows Server 2016 is used as the OS for the VMs.

The specific space reduction results are reflective of the environment tested. Under different environments (for the storage area of the user data or between VMs installed with different OSs), the amount of space reduction will vary to a degree based on the amount of common files across VMs or the compression ratio.

In this test, deduplication/compression is applied to an environment where the OS areas are created by cloning. Therefore, there are comparatively more common files between the VMs.

In this chapter, "Thin Provisioning Pool" refers to a Thin Provisioning Pool that does not use the Deduplication/Compression function and "Deduplication/Compression Pool" refers to a Thin Provisioning Pool that uses it.

3.1 Verifying the Effects of Space Reduction

This test verifies the effects by comparing the physical drive usage when the Deduplication/Compression function is used and when it is not used.
The required space of the physical drive when only the OS is used is calculated from the required capacity of the physical drive in the physical environment. Then, the increased physical drive capacity in the virtual platform is checked.

The effects of space reduction using the Deduplication/Compression function are outlined in "Figure - 3 Deduplication between different platforms".
In the physical environment, a single Windows Server OS uses a volume as the drive area. In the virtual environment, it uses a virtual drive as the drive area. Two Windows Server machines are used in each environment.

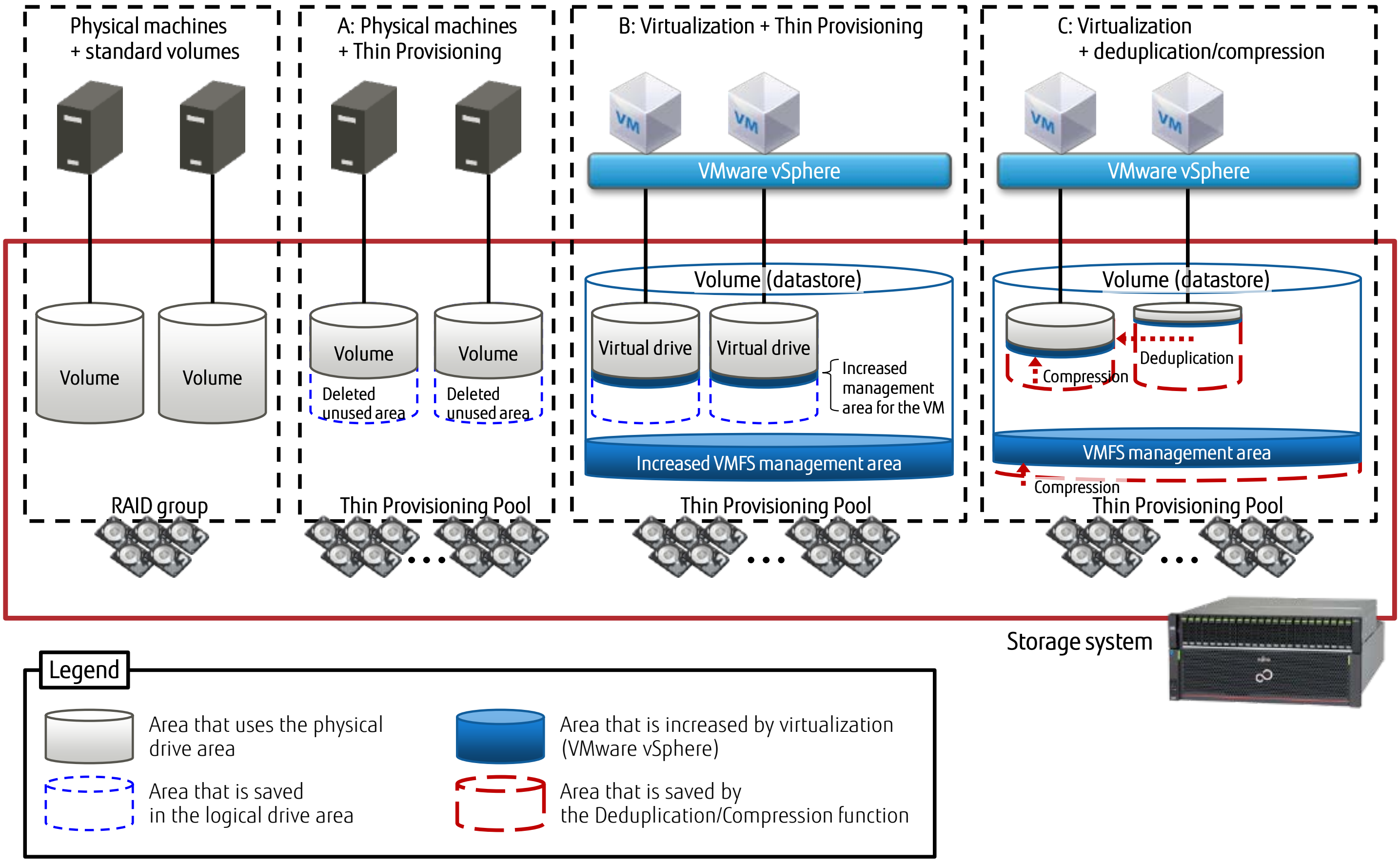


Figure - 4 Overview of the storage area used for virtualization and deduplication/compression

This test verifies the space reduction effects by comparing the physical drive areas in the three cases that use Thin Provisioning Pools.

Verification cases are shown in "Table - 1 Comparison of physical drive areas".

The physical drive space used by a standard volume depends on the size that is allocated to the volume. For this reason, a configuration using standard volumes is not targeted for this test.

Case	Configuration	Description
A	Physical machines + Thin Provisioning	The physical drive areas are used only for the drive areas used by the OS because the Thin Provisioning function is enabled. The drive space used only by the OS is checked.
B	Virtualization + Thin Provisioning	The physical drive areas (such as VMFS and VM management areas) required in the virtual platform are checked. In addition, the physical drive usage is checked before the space is reduced using the Deduplication/Compression function.
C	Virtualization + deduplication/compression	The effects of space reduction provided by the Deduplication/Compression function are checked.

Table - 1 Comparison of physical drive areas

3.1.1 Reducing the Drive Usage by Compression

The Deduplication/Compression function of the ETERNUS AF/DX removes duplicate data and compresses the remaining data to be stored in physical drives.
For this test, a single machine with Windows Server 2016 Standard (Desktop Experience) OS is deployed in a storage pool and the physical drive usage is checked. The physical drive usage in each pool is compared to verify the effects of compression.

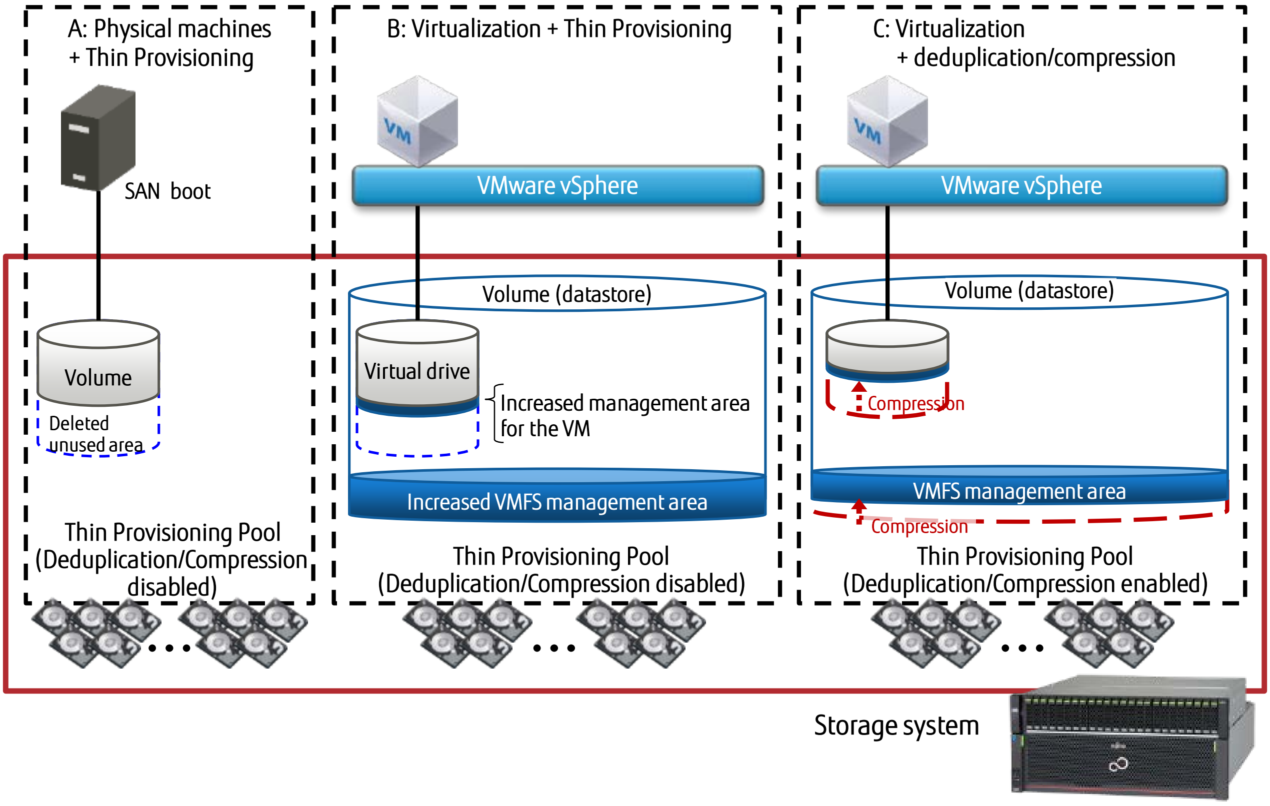


Figure - 5 Overview of the effects of drive usage reduction provided by compression

The test is performed using the following steps.

Configuring "A: Physical machines + Thin Provisioning"

- The physical machine is connected to the TPV and a SAN boot environment is configured.
- Windows Server 2016 is installed and the physical drive usage of the Thin Provisioning Pool is checked.

Configuring "B: Virtualization + Thin Provisioning" and " C: Virtualization + deduplication/compression" (virtual environments)

- A VM installed with Windows Server 2016 is prepared for a work datastore.
  (A template with Windows Server 2016 is prepared for creating a VM with the template.)
- The VM is turned off and has no snapshots.
- The created VM is migrated to the datastore for verification by using the Storage vMotion function of VMware vSphere.
- After Storage vMotion, the physical drive usage is checked in the environment where a single VM is deployed in the Thin Provisioning Pool.

3.1.2 Reducing the Drive Usage by Deduplication

For this test, VM clones are deployed in the environment described in "3.1.1 Reducing the Drive Usage by Compression" and the effects of space reduction provided by deduplication are checked.
Because VM clones are created by copying the virtual disk, the virtual disk contents of the clones are identical and high deduplication effects can be expected.
Deduplication effects for the virtual disk (.vmdk) in the VMware vSphere datastore (VMFS) are also checked in this test.

The main action performed by the VM clones is to copy the virtual disk files (.vmdk) in the datastore.
The VMware vSphere datastore is a file system (VMFS) with an extremely large block size of 1MB and exclusively stores large VM files.

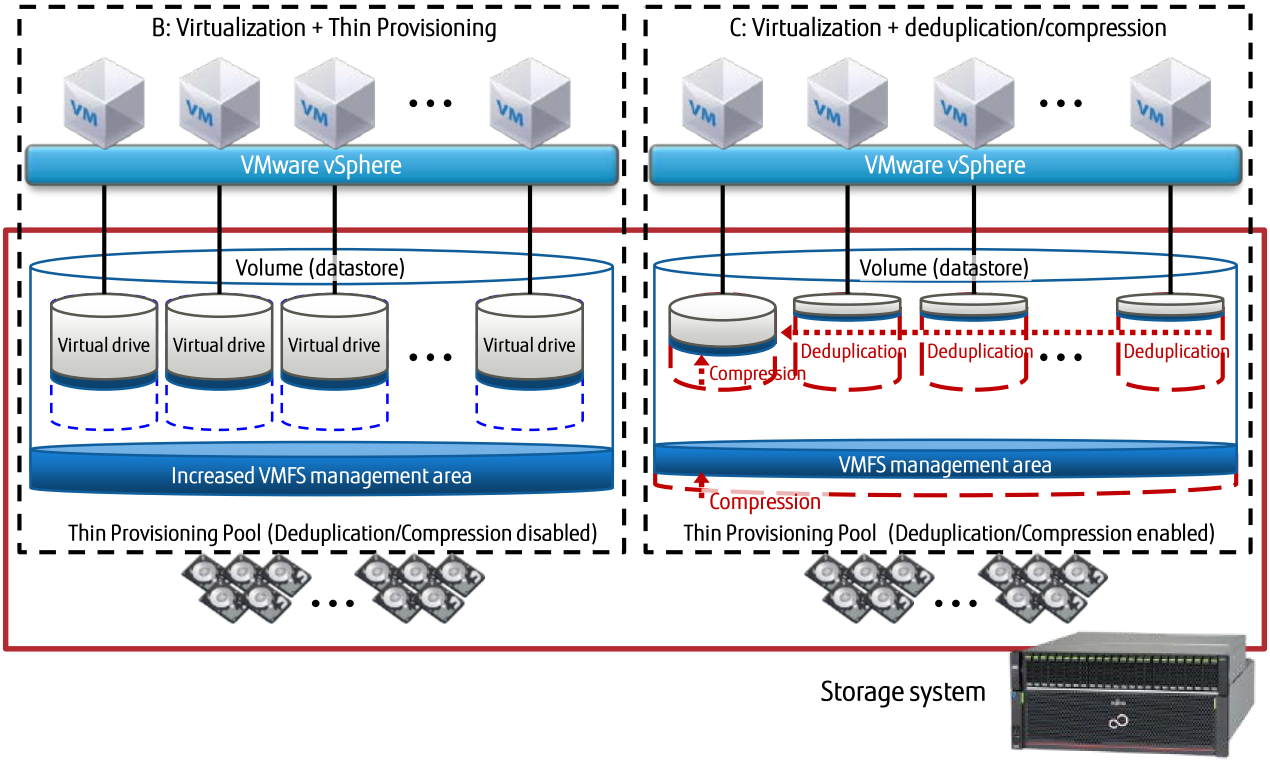


Figure - 6 Overview of the effects of drive usage reduction provided by deduplication

The test is performed using the following steps.

- VMs are created in the work datastore using the same template as the VM in "3.1.1 Reducing the Drive Usage by Compression".
  (Creation of VM clones)
- The created VM's machine name and network address in the Windows Server are changed.
- The created VM is migrated to the datastore for verification using the Storage vMotion function.
- With the above operation, the multiple VMs are deployed in the datastore and the physical space that is used in each pool is checked.

The space reduction effects are checked by comparing the physical spaces used in the Thin Provisioning Pool and in the Deduplication/Compression Pool.

3.1.3 Effects of Deduplication on Continuous Operations

For this test, the deduplication effects are checked when common files are updated with the Windows update programs for multiple VMs after they are deployed.

For recent OSs, patches are provided on regular basis for security and other reasons. This causes common files in multiple VMs to be updated or added. This test verifies that deduplication is effective for the files in Windows Server 2016 (NTFS) and for the VMs in VMware vSphere datastore (VMFS).
The test verifies whether the deduplication function reduces the physical drive usage when common files are updated with, for example, Windows Update for multiple VMs.

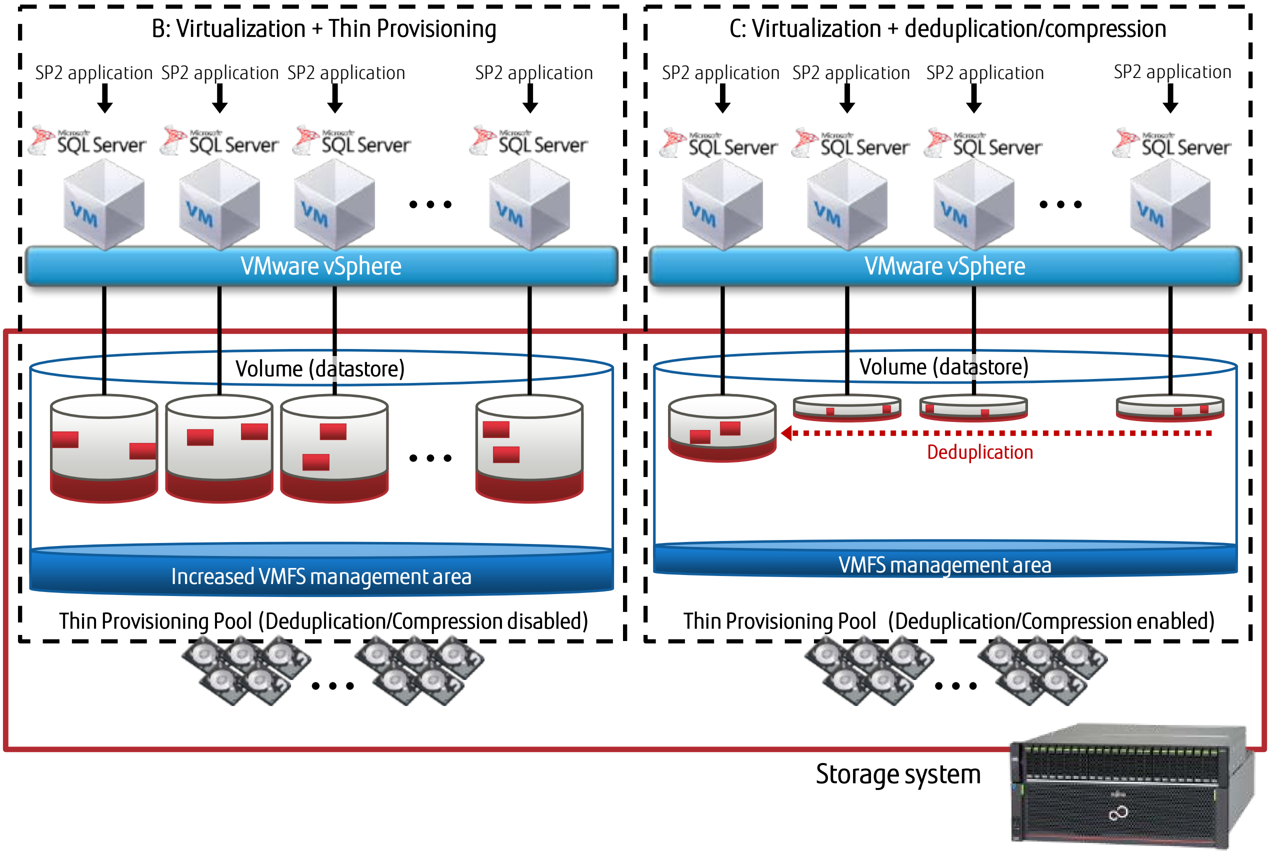


Figure - 7 Overview of the effects of drive usage reduction provided by deduplicating the files in the OS of the guests

The test is performed using the following steps.

- VMs installed with Microsoft SQL Server 2014 (without SP) are used.
- To check the changes of the physical space due to updates, multiple VMs installed with SQL Server are deployed in the datastore for verification before a Service Pack is applied.
- An environment is configured to assume actual operations such as updating and adding different files on each guest OS. (Files with different sizes are deployed to each VM guest OS so that the file system on the guest OS is different for each VM. The file size is in a range of 1MB to 10MB and is not divisible by 4KB [*4].)
- The Service Pack is applied to Microsoft SQL Server 2014.
- With the above operation, the physical space used in each pool is checked after the files are updated by applying the Service Pack.

The space reduction effects are checked by comparing the physical spaces used in the Thin Provisioning Pool and in the Deduplication/Compression Pool after the Service Pack is applied.

*4 The ETERNUS AF/DX performs the Deduplication/Compression function on a 4KB basis. Therefore, this file size should not be divisible 4KB.

3.2 Measuring the Effects of Space Reduction

This section explains the verification environment.

Storage system used to verify the Deduplication/Compression function

Storage system : FUJITSU Storage ETERNUS AF650 (All-Flash Arrays)

Other device and software

VM host : FUJITSU Server PRIMERGY
Virtual platform : VMware vSphere 6.5
VM guest OS : Microsoft Windows Server 2016 Standard Edition (Desktop Experience)
Applied Service Pack : Microsoft SQL Server 2014 and Microsoft SQL Server 2014 Service Pack 2

ETERNUS AF650's volumes, VMware vSphere's datastores, and virtual disks that are assigned to VMs are explained below.

A 100GB disk space is allocated to each machine in which Windows Server 2016 is installed.
The physical machine is assigned a 100GB Thin Provisioning Volume and supports SAN boot.
In the virtual environment, "Thick Provisioning (Lazy Zeroed)" *5 is used by VMware vSphere to format virtual disks in order to verify the effects of a space reduction provided by the storage system.

Thick Provisioning (Lazy Zeroed) secures (thick provisions) an area that is allocated (100GB in this test) to a virtual disk (.vmdk) in the datastore when virtual disks are created. (A 100GB virtual disk file (.vmdk) is created in the datastore.) Without an initialization of the virtual disk area, the required space is deleted and used when accessed for the first time from the VM.
For Thick Provisioning (Lazy Zeroed), the Thin Provisioning function of the storage system works effectively because the drive area is accessed only when the area is used from the VM's guest OS.

In this chapter, the physical space used in each Thin Provisioning Pool is measured after Windows Server 2016 Standard (Desktop Experience) is installed in the 100GB drive space allocated to each machine.

A 100GB SAN boot volume for the physical machine and a 20TB volume for the datastore in the virtual environment are allocated from volumes in the ETERNUS AF650.

The following table shows the drive space that is used for verification.

<table>
<tr><th rowspan="2">Case</th><th rowspan="2">Configuration</th><th colspan="2">Storage volume</th><th rowspan="2">VMware vSphere environment</th><th rowspan="2">Drive allocated to the OS</th></tr>
<tr><th>Volume format</th><th>Volume size</th></tr>
<tr><td>A</td><td>Physical machines
+ Thin Provisioning</td><td rowspan="2">Thin Provisioning
Deduplication/Compression function disabled</td><td>100GB</td><td>–
(SAN boot)</td><td>100GB</td></tr>
<tr><td>B</td><td>Virtualization
+ Thin Provisioning</td><td rowspan="2">20TB</td><td rowspan="2">The entire volume (20TB) is assigned to the datastore formatted with VMFS6</td><td rowspan="2">100GB virtual disk (Thick Provisioning)</td></tr>
<tr><td>C</td><td>Virtualization
+ deduplication/compression</td><td>Thin Provisioning
Deduplication/Compression function enabled</td></tr>
</table>

Table - 2 Drive spaces allocated to Windows Server 2016

*5 Thick Provisioning (Lazy Zeroed)
Creates virtual disks in the default thick format. The required space is allocated when the virtual disks are created. The data remaining in the physical devices is not deleted when the virtual disks are being created but is deleted on demand when the data is written in the VMs for the first time.

The drive spaces used by the VMs are shown in "Figure - 8 Virtual disk space and storage system space".

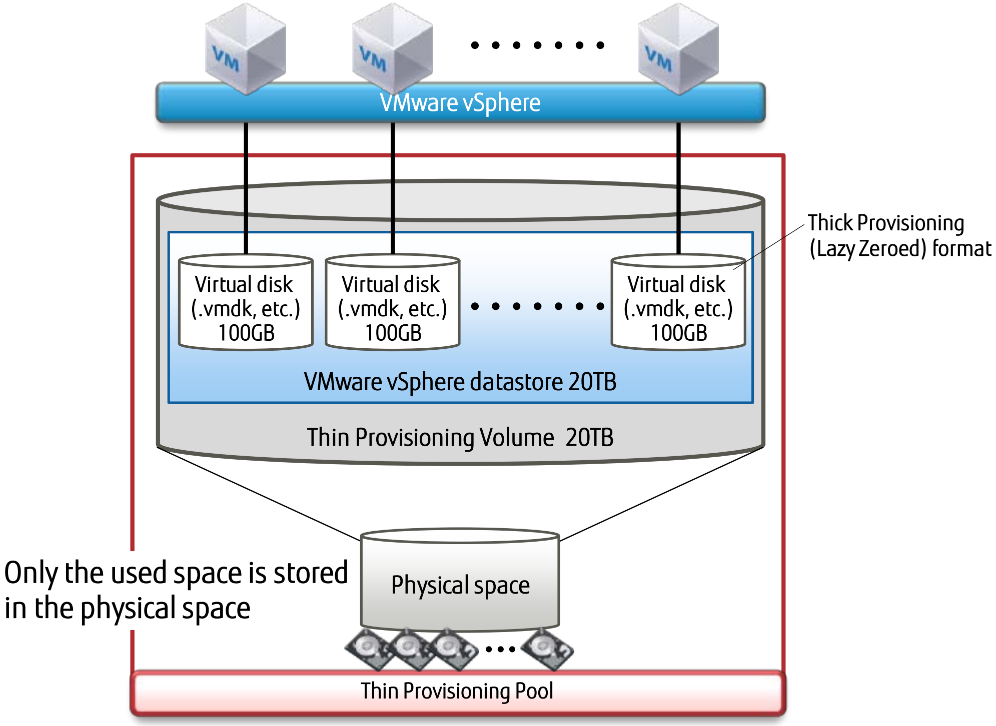


Figure - 8 Virtual disk space and storage system space

3.2.1 Measuring the Effects of Space Reduction by Compression

The following table shows the physical drive space that was used after a machine with Windows Server 2016 was stored in each storage pool.

No.	Measured timing	A: Physical machines + Thin Provisioning	B: Virtualization + Thin Provisioning	C: Virtualization + deduplication/compression	Reduction rate (1 - C ÷ B)
1	After the datastore is created	0.00 GB (after a volume is created)	0.98 GB	0.37 GB	62.5 %
2	After a VM is deployed in the datastore	9.04 GB (after installation)	10.73 GB	7.10 GB	33.8 %

Table - 3 Physical drive space and reduction rate after a Windows Server machine is deployed

The approximate size of the physical space used when Windows Server 2016 is installed is shown in "A: Physical machines + Thin Provisioning". (The physical drive space that is required to install Windows Server 2016 may vary depending on the space used by the Thin Provisioning function and the server hardware configuration.)

The values for No.2 ("After a VM is deployed in the datastore") are compared.
Compared with "A: Physical machines + Thin Provisioning", "B: Virtualization + Thin Provisioning" requires a larger physical space due to the space required for virtualizations such as datastore formatting.

"C: Virtualization + deduplication/compression" requires a smaller physical space than "A: Physical machines + Thin Provisioning". This is mainly due to the effects of compression. The effects of a deduplication are small because only one Windows Server machine is deployed.

When "B: Virtualization + Thin Provisioning" and "C: Virtualization + deduplication/compression" (which are similar environments) are compared, 33.8% of the physical space is reduced due to deduplication/compression.

3.2.2 Measuring the Effects of Space Reduction by Deduplication

The following table shows the physical drive space that was used after multiple Windows Server 2016 VMs created by VM cloning were stored in the Thin Provisioning Pool or the Deduplication/Compression Pool.

No.	Conditions		B: Virtualization + Thin Provisioning			C: Virtualization + deduplication/compression			Physical space reduction rate (1 - C ÷ B)
	No. of VMs to add	Total No. of stored VMs	Used physical space	Increased physical space	Increased space per VM	Used physical space	Increased physical space	Increased space per VM	
1		0 VMs	0.98 GB	–	–	0.37 GB	–	–	62.5 %
2	1 VM	1 VM	10.73 GB	9.75 GB	9.75 GB	7.10 GB	6.73 GB	6.73 GB	33.8 %
3	1 VM	2 VMs	20.45 GB	9.72 GB	9.72 GB	7.53 GB	0.43 GB	0.43 GB	63.2 %
4	1 VM	3 VMs	30.21 GB	9.76 GB	9.76 GB	7.92 GB	0.39 GB	0.39 GB	73.8 %
5	2 VMs	5 VMs	49.79 GB	19.58 GB	9.79 GB	8.74 GB	0.82 GB	0.41 GB	82.4 %
6	5 VMs	10 VMs	98.62 GB	48.83 GB	9.77 GB	10.75 GB	2.01 GB	0.40 GB	89.1 %
7	10 VMs	20 VMs	196.20 GB	97.58 GB	9.76 GB	14.58 GB	3.83 GB	0.38 GB	92.6 %
8	30 VMs	50 VMs	488.11 GB	291.91 GB	9.73 GB	18.46 GB	3.88 GB	0.13 GB	96.2 %

Table - 4 Physical drive space and reduction rate after multiple Windows Server machines are deployed

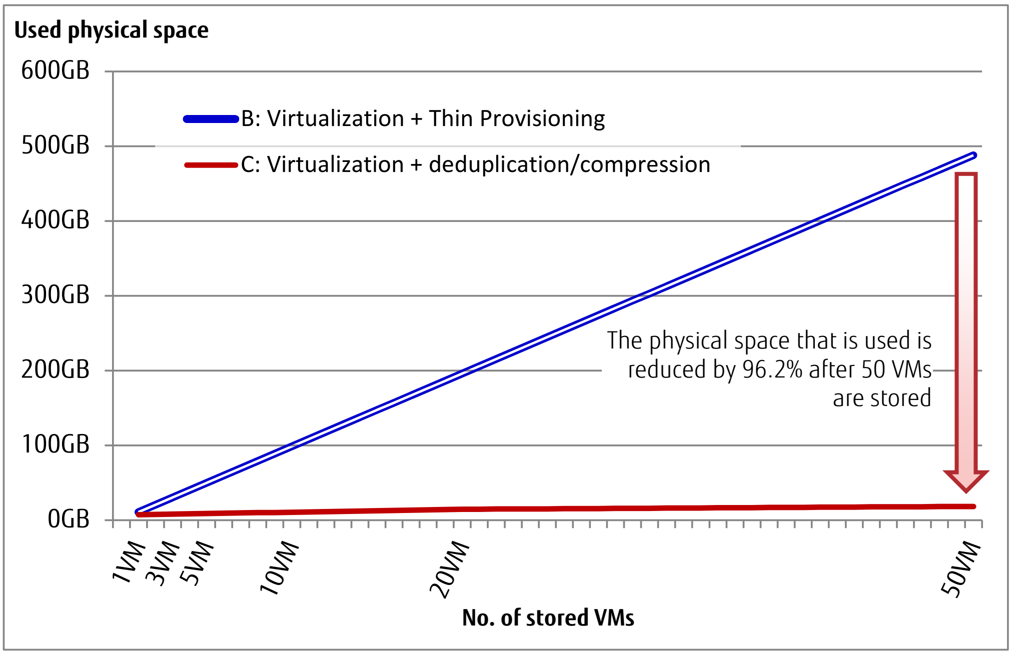


Figure - 9 Amount of physical space used after multiple Windows Server machines are deployed

Conditions No.1 and No.2 (where a VM is stored) are the same as "3.2.1 Measuring the Effects of Space Reduction by Compression". No.3 and later show the physical space that is used when a different number of VM clones are added to each storage pool.

In “B: Virtualization + Thin Provisioning”, the values for “Increased space per VM” show a slight difference even when the number of stored VMs changes from 1 to 50.

"C: Virtualization + deduplication/compression" shows that the values for "Increased space per VM" significantly decrease in No.3 (where 2 VMs are stored) and later. This means that deduplication is functioning effectively.
Deduplication also deletes duplicate data but requires an area to manage them. In addition, the used physical space is slightly increased because of changes made after a VM cloning. These include changes to the computer name of Windows Server 2016, the network address, and additions to the OS log.

This test result shows that the reduction rate reaches more than 90% when 20 VMs are stored.
In a virtual environment with numerous VM clones, the Deduplication/Compression function is highly beneficial.

However, note that this measurement was conducted in a favorable environment where little changes were made to the VM files after cloning. The reduction rate varies depending on updates and additions to the VM files that occur as the result of operations of Windows Server logs, virtual memory files, and software installation.

3.2.3 Measuring the Effects of Deduplication/Compression on Continuous Operations

The following table shows the physical drive space that was used when the files common to the guest OS were updated and added in VMs that were deployed in the Thin Provisioning Pool or the Deduplication/Compression Pool.
This measurement was conducted after 50 VMs that have SQL Server 2014 installed were stored in the datastore and then the files were updated and added by applying Service Pack 2.

No.	Conditions		B: Virtualization + Thin Provisioning				C: Virtualization + deduplication/compression				Reduction rate for "Increased space per VM" (1 - C ÷ B)
	No. of VMs to be applied with SP	Total No. of VMs applied with SP	Used physical space	Increased physical space	Sum of the increased space	Increased space per VM	Used physical space	Increased physical space	Sum of the increased space	Increased space per VM	
1	(50 VMs deployed without SP applied)	0 VMs	861.0 GB	–	–	–	20.1 GB	–	–	–	–
2	1 VM	1 VM	863.5 GB	2.5 GB	2.5 GB	2.5 GB	20.6 GB	0.5 GB	0.5 GB	0.5 GB	80.7%
3	1 VM	2 VMs	866.0 GB	2.5 GB	5.0 GB	2.5 GB	20.7 GB	0.1 GB	0.6 GB	0.1 GB	94.4%
4	1 VM	3 VMs	868.6 GB	2.5 GB	7.6 GB	2.5 GB	20.9 GB	0.1 GB	0.8 GB	0.1 GB	94.1%
5	2 VMs	5 VMs	873.4 GB	4.9 GB	12.4 GB	2.4 GB	21.2 GB	0.3 GB	1.1 GB	0.1 GB	93.9%
6	5 VMs	10 VMs	885.8 GB	12.3 GB	24.8 GB	2.5 GB	21.9 GB	0.8 GB	1.8 GB	0.2 GB	93.7%
7	10 VMs	20 VMs	910.8 GB	25.0 GB	49.8 GB	2.5 GB	23.1 GB	1.1 GB	3.0 GB	0.1 GB	95.4%
8	30 VMs	50 VMs	985.2 GB	74.4 GB	124.2 GB	2.5 GB	26.2 GB	3.1 GB	6.1 GB	0.1 GB	95.8%

Table - 5 Common file updates on the guest OS by applying the Service Pack

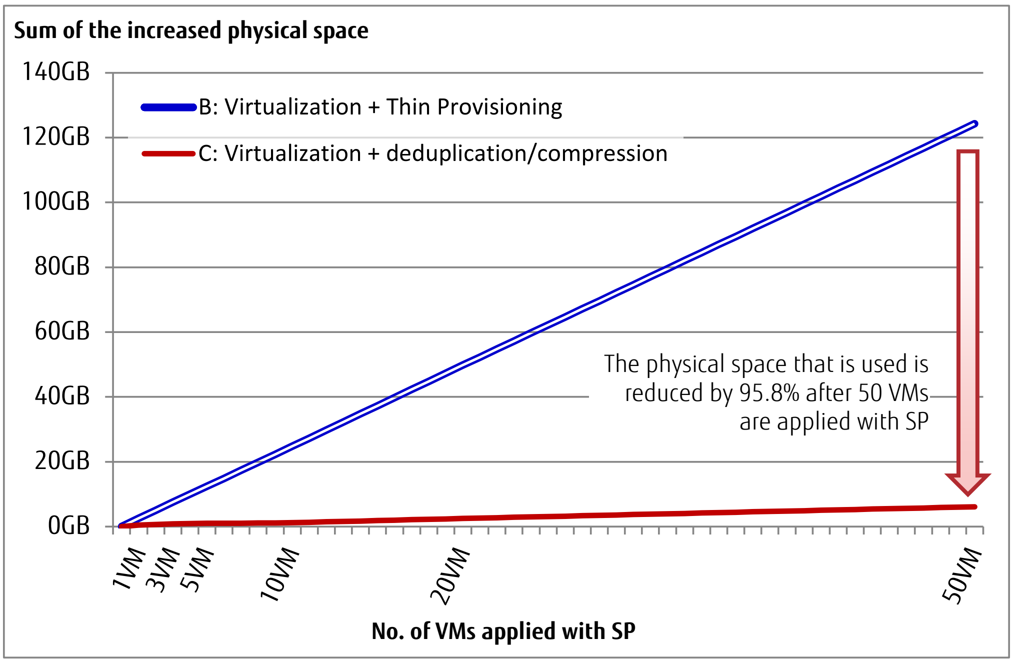


Figure - 10 Sum of the increased physical space after applying the Service Pack

In "B: Virtualization + Thin Provisioning", the values for "Increased space per VM" are almost constant 2.5GB.
On the other hand, in "C: Virtualization + deduplication/compression", the value for "Increased space per VM" is 0.5GB in No.2 ("1 VM with SP applied" which is the first SP application) and approximately 0.1GB in No.3 and later.

The values for "Increased space per VM" in "C: Virtualization + deduplication/compression" are smaller than those of "B: Virtualization + Thin Provisioning".
In addition, similar to "3.2.2 Measuring the Effects of Space Reduction by Deduplication", due to the effects of compression applied to the first 1VM and the effects of deduplication for the second VM and later, the values for "Increased space per VM" are further reduced.

Compared with "3.2.1 Measuring the Effects of Space Reduction by Compression", the result above shows that a higher reduction rate is achieved for the first 1VM. This is due to a high reduction rate in the SQL Server database area that is updated or added after the Service Pack is applied.

The applied Service Pack file size is 876GB after being uncompressed. In No.2 ("1 VM with SP applied") of "C: Virtualization + deduplication/compression", the value for "Increased space per VM" is 0.5GB. Based on the file size of the Service Pack, the reduction rate is approximately 42%.

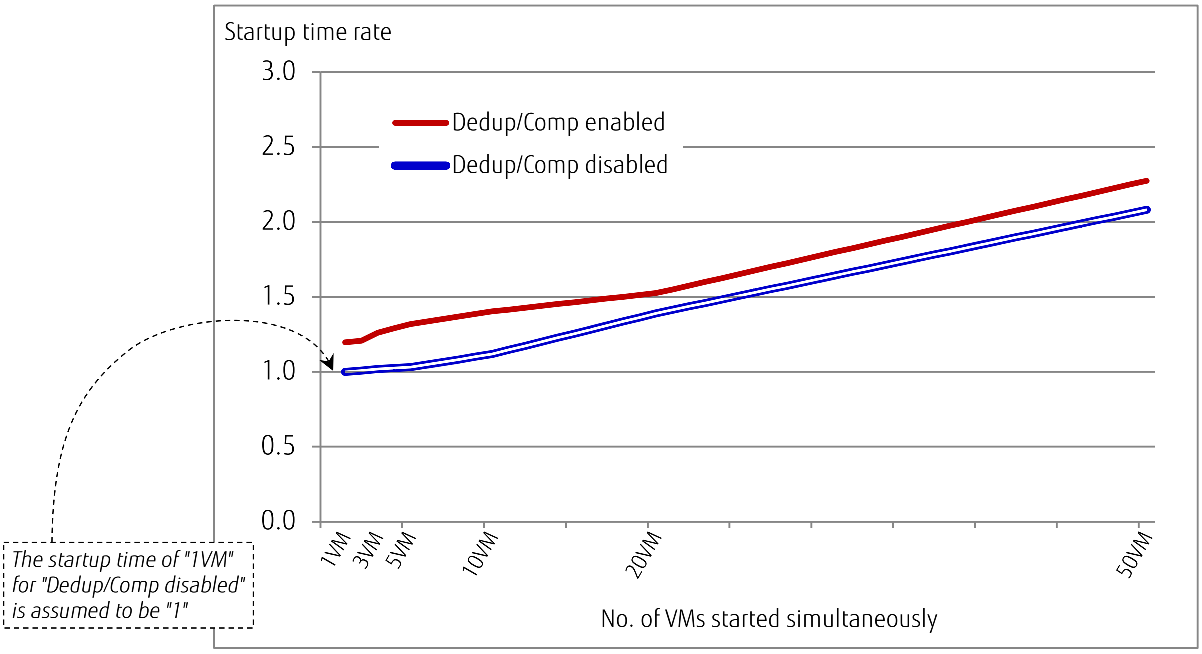
3.3 Deduplication/Compression and Performance

This section shows comparisons of the impact on performance when the Deduplication/Compression function is enabled and disabled. The performance is compared by measuring the processing time of VM operations in VMware vSphere when a Thin Provisioning Volume consisting of SSDs is used as the datastore.

Using the operations of the VMs described in "3.2 Measuring the Effects of Space Reduction" as an example, the performance trend depending on the setting of the Deduplication/Compression function is shown below.

3.3.1 Comparison of Startup Times for Windows Server 2016

This test compares the times to complete automatic login operations of all VMs with Windows Server 2016 Standard when every VM in VMware vSphere is powered on simultaneously. Read access to the datastore was mainly performed.
The following graph shows the comparison between startup times when the Deduplication/Compression function is enabled and disabled. The startup time of "1VM" for "Dedup/Comp disabled" is assumed to be "1".



* Example for an SSD RAID5(3+1) configuration

Figure - 11 Comparison of startup times when the Deduplication/Compression function is enabled and disabled

When the Deduplication/Compression function is disabled, simultaneously starting 20 VMs takes approximately 1.4 times longer than 1VM and 50 VMs takes approximately 2.1 times longer.
When the Deduplication/Compression function is enabled, the startup times according to the number of VMs that are started simultaneously tend to be the same as when the function is disabled. Startup takes approximately 1.2 times [*6] longer than when the function is disabled.

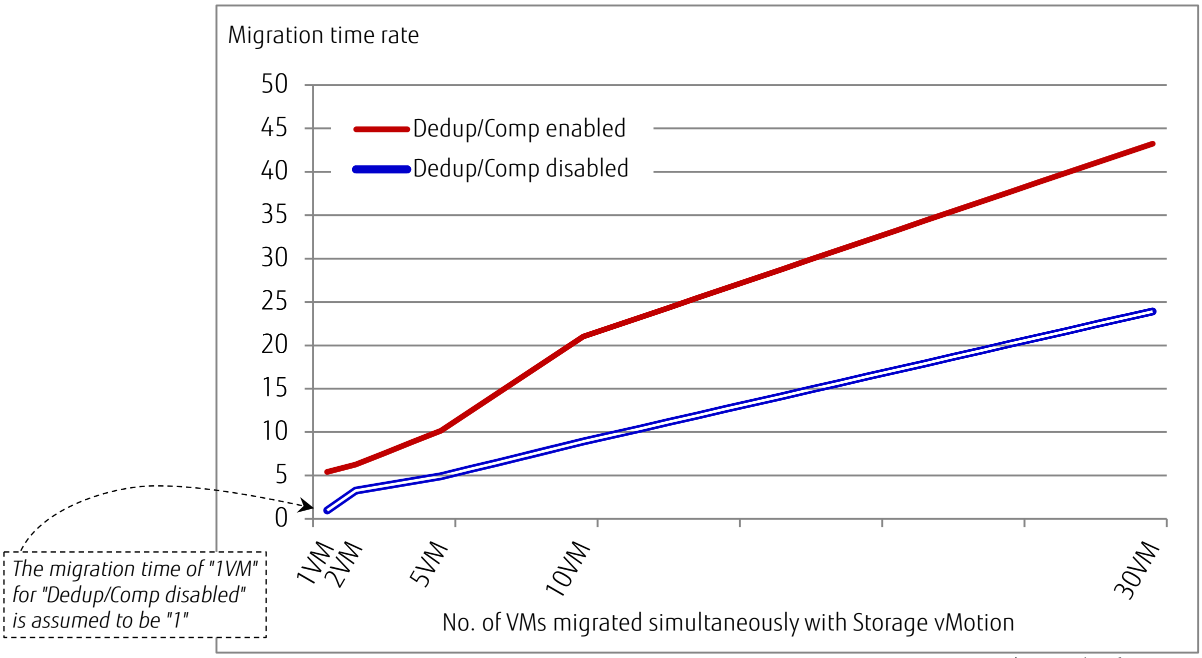
The performance trend does not change even when 50 VMs are started up simultaneously. This indicates that the system has enough power left, although the system performance also depends on the server performance.

*6 The performance may be lower than the values described here depending on conditions such as the amount of data to be stored.

3.3.2 Time Comparison for Storage vMotion

This test compares the time to migrate the VMs in VMware vSphere to the target datastore with Storage vMotion. (Processing times when the VMs are migrated to the datastore using the steps described in "3.1.2 Reducing the Drive Usage by Deduplication".) Read access to the target datastore was mainly performed.

The following graph shows the comparison between migration times when the Deduplication/Compression function is enabled and disabled. The migration time of "1VM" for "Dedup/Comp disabled" is assumed to be "1".



* Example of an SSD RAID5(3+1) configuration

Figure - 12 Time comparison for Storage vMotion when the Deduplication/Compression function is enabled and disabled

When the Deduplication/Compression function is disabled, the migration time increases at an almost constant rate. When 30 VMs are migrated simultaneously (with Storage vMotion), the process takes approximately 24 times longer than that of 1VM. This means that the migration time does not increase as much as the number of VMs.

When the Deduplication/Compression function is enabled and the first VM is stored in the datastore, duplicate data does not exist in the datastore. Deduplication is not performed but processes such as data compression and data writes are executed in the drive. Therefore, the migration time takes approximately 5.4 [*7] times longer than when the Deduplication/Compression function is disabled. (The time is measured by migrating a single VM, by migrating two VMs simultaneously, five VMs simultaneously, and so on to an empty datastore.)

However, when two or more VMs are migrated, the process only takes 2.0 to 2.4 [*7] times longer due to the effects of deduplication. The rate of the processing time has the same tendency as when the Deduplication/Compression function is disabled regardless of the number of VMs to migrate simultaneously.

The processing time takes longer than when the Deduplication/Compression function is disabled but shorter than when HDDs are used. If a RAID5(3+1) configuration consisting of four SSDs with deduplication/compression enabled is compared with a RAID1+0(4+4) configuration consisting of eight HDDs (twice as many drives as the RAID5(3+1) configuration to prioritize performance) with deduplication/compression disabled, the migration time of 30 VMs with Storage vMotion is shorter in the SSD configuration with deduplication/compression enabled or approximately 1/3 the time of the other configuration. The migrations were completed approximately 6 times faster per drive and more than 6 times faster when differences in the RAID configurations are considered.

*7 The performance may be lower than the values described here depending on conditions such as the amount of data to be stored.

4 Summary

The test results prove the outstanding effectiveness of the Deduplication/Compression function not only for VM clones in the virtual platform, but also for file updates that are common between the VM guests which include OS patches required for continuous operation.

The Deduplication/Compression function of the ETERNUS AF/DX drastically reduces the physical drive space to be used in environments that have a large number of duplicate data such as clones in a virtual platform. Furthermore, using SSDs enhances performance.

In addition, the ETERNUS AF/DX enables a volume-based setting for the Deduplication/Compression function.

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.
This flexible setting provides a balance between reducing the space to be used and maintaining performance in a single storage system by, for example, enabling the Deduplication/Compression function in the OS area where the effects are highly expected and by disabling it in the system data area where only a small amount of duplicate data exists.

The ETERNUS AF/DX reduces the drive space to be used in the virtual platform, reduces costs, and enhances performance.

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