NETAPP TECHNICAL REPORT

NetApp Deduplication for FAS Deployment and Implementation Guide

NetApp Technical Marketing
September, 2008 | TR-3505

Fifth Revision

ABSTRACT

This guide introduces NetApp® deduplication for FAS technology, describes in detail how to implement and use it, and provides information on best practices, operational considerations, and troubleshooting.

It should prove useful for both NetApp and channel partner sales and services field personnel who require assistance in understanding details and successfully deploying solutions that include deduplication.

TABLE OF CONTENTS

1 INTRODUCTION AND OVERVIEW OF DEDUPLICATION 4

1.1 HOW DEDUPLICATION FOR FAS WORKS 4

1.2 DEDUPLICATED VOLUMES 5

1.3 DEDUPLICATION METADATA 5

1.4 GENERAL DEDUPLICATION FEATURES 6

2 CONFIGURATION AND OPERATION 7

2.1 CONFIGURATION AND OPERATION 7

2.2 REQUIREMENTS OVERVIEW 7

2.3 INSTALLING AND LICENSING DEDUPLICATION 8

2.4 COMMAND SUMMARY 9

2.5 DEDUPLICATION QUICK START GUIDE 10

2.6 END-TO-END DEDUPLICATION CONFIGURATION EXAMPLE 10

2.7 CONFIGURING DEDUPLICATION SCHEDULES 12

3 SIZING FOR PERFORMANCE AND SPACE EFFICIENCY 14

3.1 DEDUPLICATION GENERAL BEST PRACTICES 14

3.2 DEDUPLICATION PERFORMANCE 14

3.3 DEDUPLICATION STORAGE SAVINGS 16

3.4 SPACE SAVINGS ESTIMATION TOOL (SSET) 18

3.5 DEDUPLICATION LIMITATIONS 18

4 DEDUPLICATION WITH OTHER NETAPP FEATURES 21

4.1 DEDUPLICATION AND SNAPSHOT COPIES 21

4.2 DEDUPLICATION AND SNAPRESTORE 21

4.3 DEDUPLICATION AND THE VOL COPY COMMAND 22

4.4 DEDUPLICATION AND FLEXCLONE VOLUMES 22

4.5 DEDUPLICATION AND ACTIVE-ACTIVE CONFIGURATION 22

4.6 DEDUPLICATION AND SNAPMIRROR REPLICATION 23

4.7 DEDUPLICATION AND SNAPVAULT 27

4.8 DEDUPLICATION AND SNAPVAULT FOR NETBACKUP 28

4.9 DEDUPLICATION AND MULTISTORE (VFILER) 29

4.10 DEDUPLICATION AND LUNS 29

5 DEDUPLICATION AND VMWARE 33

6 TROUBLESHOOTING 37

6.1 LICENSING 37

6.2 VOLUME SIZES 37

6.3 LOGS AND ERROR MESSAGES 37

6.4 NOT SEEING SPACE SAVINGS 37
6.5 UNDEDUPLICATING A FLEXIBLE VOLUME 38
6.6 ADDITIONAL REPORTING WITH SIS STAT -1 38
6.7 DEDUPLICATION AND SYSTEM REBOOTS, HALTS, AND SO ON 39
7 APPENDIX – ADDITIONAL READING AND REFERENCES 40

1 INTRODUCTION AND OVERVIEW OF DEDUPLICATION

This section provides an overview of how deduplication for FAS works.

1.1 HOW DEDUPLICATION FOR FAS WORKS

NetApp deduplication for FAS provides block-level deduplication within the entire flexible volume on NetApp storage systems. Figure 1 shows how this works at the highest level.

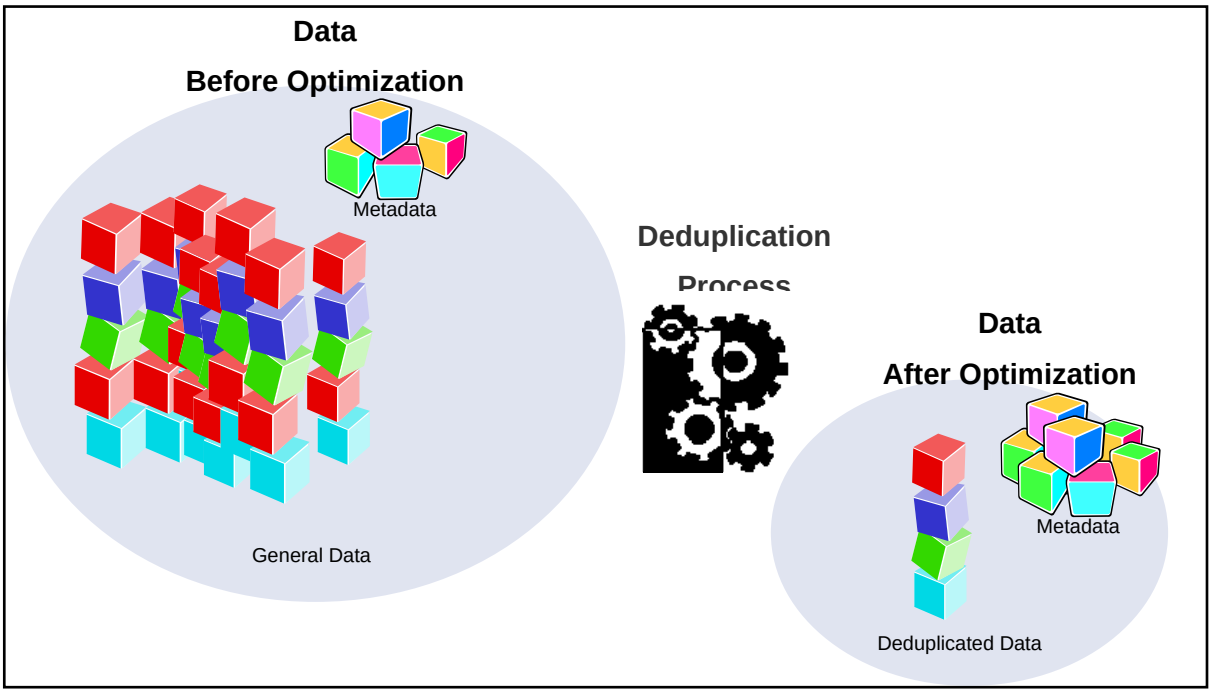

Figure 1) How NetApp deduplication for FAS works.

Essentially, deduplication stores only unique blocks in the flexible volume and creates a small amount of additional metadata in the process. Notable features of deduplication include:

It works with a high degree of granularity; that is at the 4KB block level.

It operates on the active file system of the flexible volume. Any block referenced by a Snapshot™ copy is not made "available" until the Snapshot copy is deleted.

It's a background process that can be configured to run automatically, or it can be scheduled, or run manually through the command line interface (CLI).

It's application transparent, and therefore it can be used for deduplication of data originating from any application using the NetApp system.

It's enabled and managed via a simple CLI.

It can be enabled on and can deduplicate blocks on flexible volumes with new and existing data.

In summary, this is how deduplication works. Newly saved data on the FAS system is stored in 4KB blocks as usual by Data ONTAP®. Each block of data has a digital fingerprint, which is compared to all other fingerprints in the flexible volume. If two fingerprints are found to be the same, a byte-for-byte comparison is done of all bytes in the block and, if there is an exact match between the new block and the existing block on the flexible volume, the duplicate block is discarded and its disk space is reclaimed.

1.2 DEDUPLICATED VOLUMES

Despite the introduction of less expensive ATA disk drives, one of the biggest challenges for storage systems today continues to be the storage cost. There is a desire to reduce storage consumption (and therefore storage cost per MB) by eliminating duplicated data through sharing blocks across files.

The core NetApp technology to accomplish this goal is the deduplicated volume, a flexible volume that contains shared data blocks. Data ONTAP supports shared blocks in order to optimize storage space consumption. Basically, within one volume, there is the ability to have multiple references to the same data block, as shown in Figure 2.

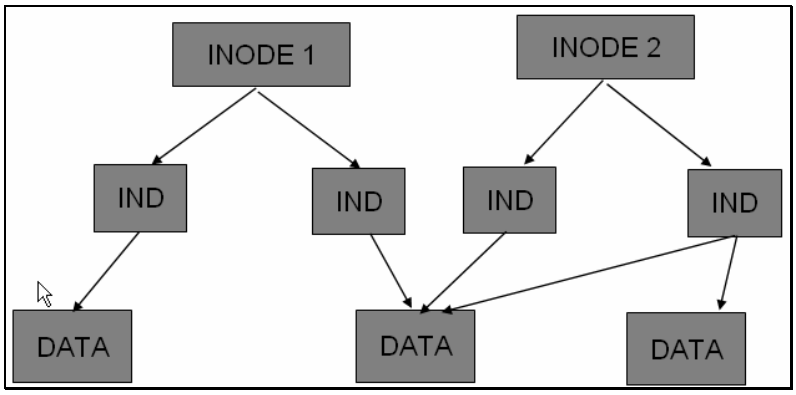

Figure 2) Data structure in a deduplicated volume.

In Figure 2, the number of physical blocks used on the disk is 3 (instead of 5), and the number of blocks saved by deduplication is 2 (5 minus 3). In the remainder of this document, these will be referred to as *used* blocks and *saved* blocks.

Each data block has a block count reference kept in the volume metadata. As additional indirect blocks ("IND" in Figure 2) point to the data, or existing ones stop pointing to it, this value is incremented or decremented accordingly. When no indirect blocks point to a data block, it is released.

The NetApp deduplication technology allows duplicate 4KB blocks anywhere in the flexible volume to be deleted, as described in the following sections.

The maximum sharing for a block is 255. This means, for example, that if there are 500 duplicate blocks, deduplication would reduce that to only 2 blocks. Also note that this ability to share blocks is different from the ability to keep 255 Snapshot copies for a volume.

1.3 DEDUPLICATION METADATA

The core enabling technology of deduplication is fingerprints. These are unique digital "signatures" for every 4KB data block in the flexible volume.

When deduplication runs for the first time on a flexible volume with existing data, it scans the blocks in the flexible volume and creates a fingerprint database, which contains a sorted list of all fingerprints for used blocks in the flexible volume.

After the fingerprint file is created, fingerprints are checked for duplicates and, when found, first a byte-by-byte comparison of the blocks is done to make sure that the blocks are indeed identical, and if they are found to be identical, the block's pointer is updated to the already existing data block and the new (duplicate) data block is released.

Releasing a duplicate data block entails updating the indirect inode pointing to it, incrementing the block reference count for the already existing data block, and freeing the duplicate data block.

In real time, as additional data is written to the deduplicated volume, a fingerprint is created for each new block and written to a change log file. When deduplication is run subsequently, the change log is sorted and its sorted fingerprints are merged with those in the fingerprint file, and then the deduplication processing occurs.

Note that there are really two change log files, so that as deduplication is running and merging the new blocks from one change log file into the fingerprint file, new data that is being written to the flexible volume is causing fingerprints for these new blocks to be written to the second change log file. The roles of the two files are then reversed the next time that deduplication is run. (For those familiar with Data ONTAP usage of NVRAM, this is analogous to when it switches from one half to the other to take a consistency point.)

Note: When deduplication is run for the first time on an empty flexible volume, it still creates the Fingerprint file from the change log.

Here are some additional details about the deduplication metadata:

There is a fingerprint record for every 4KB data block, and the fingerprints for all the data blocks in the volume are stored in the fingerprint database file.

Fingerprints are not deleted from the Fingerprint file automatically when data blocks are freed, but when a threshold of 20% new fingerprints is reached, the stale fingerprints are deleted. This can also be done by a manual operation from the command line.

In Data ONTAP 7.2.X, all the deduplication metadata resides in the flexible volume.

Starting with Data ONTAP 7.3.0, part of the metadata resides in the volume and part of it resides in the aggregate outside the volume. The fingerprint database and the change log files that are used in the deduplication process are located outside of the volume in the aggregate, and are therefore not captured in Snapshot copies. This change enables deduplication to achieve higher space savings. However, some other temporary metadata files created during the deduplication operation are still placed inside the volume. These temporary metadata files are deleted once the deduplication operation is complete. These temporary metadata files can get locked in Snapshot copies, if the Snapshot copies are created during a deduplication operation. The metadata files remain locked until the Snapshot copies are deleted.

For the size of the overhead associated with the deduplication metadata files, see section 3.3.2, Deduplication Metadata Overhead.

1.4 GENERAL DEDUPLICATION FEATURES

Deduplication is enabled on a per flexible volume basis.

It can be enabled on any number of flexible volumes in a storage system.

It can be run one of four different ways:

Scheduled on specific days and at specific times

Manually via the command line

Automatically, when 20% new data has been written to the volume

Automatically on the destination volume, when used with SnapVault® or SnapVault for NetBackup™

Only one deduplication process can run on a flexible volume at a time.

Up to eight deduplication processes can run concurrently on the same NetApp storage system.

2 CONFIGURATION AND OPERATION

2.1 CONFIGURATION AND OPERATION

This section discusses what is required to install deduplication, how to configure it, and various aspects of managing it. Although it discusses some basic things, in general it assumes both that the NetApp storage system is already installed and running, and that the reader is familiar with basic NetApp administration.

2.2 REQUIREMENTS OVERVIEW

Table 1 specifies the hardware and software required to run deduplication.

Table 1) Deduplication requirements overview.

Hardware	NearStore® R200 FAS2000 Series FAS3000 Series FAS3100 Series FAS6000 Series IBM N5000 series IBM N7000 series **Note:** Starting with Data ONTAP 7.3, the V-Series systems corresponding to the NetApp FAS systems listed above are also supported.
Data ONTAP	Data ONTAP 7.2.5.1 or later
Software	- `nearstore_option` (for all platforms except R200) license - `a_sis` license
Maximum deduplicated volume size for Data ONTAP 7.2.X (starting with 7.2.5.1) and 7.3.0	FAS6070, FAS6080, N7800: 16TB FAS6030, FAS6040, FAS3170, N7600: 10TB FAS3070, N5600: 6TB NearStore R200: 4TB FAS3040, FAS3140, N5300: 3TB FAS3050, N5500: 2TB FAS2050, FAS3020, N5200: 1TB

	FAS2020: 0.5TB **Note:** A volume should never exceed the deduplicated volume size limit for the entire life of the volume. If a volume ever gets larger than this limit, and is later shrunk to a smaller size, deduplication cannot be enabled on that volume.
Supported protocols	All

2.2.1 WHAT'S SUPPORTED AND WHAT'S NOT

The following NetApp features are supported with deduplication:

Deduplication is supported only on the R200 systems and on FAS systems with the NearStore option license.

Only flexible volumes are supported. Traditional volumes are not supported.

SnapLock® volumes are currently not supported with deduplication.

LUNs are supported with deduplication.

SnapMirror® is supported with deduplication (both qtree SnapMirror and volume SnapMirror).

SnapMirror Sync mode is not supported with deduplication.

SnapVault on the source volume is supported with deduplication.

Starting with Data ONTAP 7.3, SnapVault on the destination volume is supported.

Starting with Data ONTAP 7.3, Open Systems SnapVault is supported.

SnapVault for NetBackup (FSE – File System Export) is supported with deduplication starting with Data ONTAP 7.3.

SnapVault for NetBackup (SOI - Space Optimized Image) is supported with deduplication starting with Data ONTAP 7.2.5.1 (NetApp approval is required for this configuration).

MetroCluster is supported with deduplication (NetApp approval is required for this configuration).

The V-Series product line is supported with deduplication starting with Data ONTAP 7.3.

MultiStore® is supported with deduplication starting with Data ONTAP 7.3.

FlexShare™ is supported with deduplication.

NDMP Dump is supported with deduplication.

2.3 INSTALLING AND LICENSING DEDUPLICATION

Deduplication is included in Data ONTAP and just needs to be licensed. Add the deduplication license by using the following command:

```
license add <a_sis license key>
```

To run deduplication on any of the FAS platforms, you also need to add the NearStore option license:

```
license add <nearstore_option license key>
```

2.3.1 DEDUPLICATION LICENSING IN A CLUSTERED ENVIRONMENT

Deduplication is a licensed option behind the NearStore option license. Both nodes must have the NearStore option licensed. Deduplication must be licensed on both nodes of the cluster as well.

2.4 COMMAND SUMMARY

Table 2 describes all deduplication (related) commands:

Table 2) Deduplication command summary.

sis on <vol>	Enables deduplication on the specified flexible volume.
sis start -s <vol>	Begins the deduplication process on the flexible volume specified and performs a scan of the flexible volume to process existing data. This option is typically used upon initial configuration and deduplication on an existing flexible volume that contains un-deduplicated data. (There's no need to use this option on a volume that has just been created and doesn't contain any data.)
sis start <vol>	Begins the deduplication process on the flexible volume specified.
sis status [-l] <vol>	Returns the current status of deduplication for the specified flexible volume. The `-l` option displays a long list.
df –s <vol>	Returns the value of deduplication space savings in the active file system for the specified flexible volume. Use this command to see how much space has been saved.
sis config [-s sched]\ <vol>	Creates an automated deduplication schedule. When deduplication is first enabled on a flexible volume, a default schedule is configured, running it each day of the week at midnight.
sis stop <vol>	Suspends an active deduplication process on the flexible volume specified.
sis off <vol>	Deactivates deduplication on the flexible volume specified. This means that there will be no more change logging or deduplication operations, but the flexible volume remains a deduplicated volume and the storage savings are kept. If this command is used, and then deduplication is turned back on for this flexible volume, the flexible volume must be rescanned with the `sis start –s` command.
sis check <vol> (This command is available only in	Verifies and updates the fingerprint database for the specified

Diag mode.)	flexible volume; includes purging stale fingerprints.
`sis stat <vol>` (This command is available only in Diag mode.)	Displays the statistics of flexible volumes that have deduplication enabled.
`sis undo <vol>` (This command is available in Advanced and Diag modes.)	Reverts a deduplicated volume to a normal flexible volume.

2.5 DEDUPLICATION QUICK START GUIDE

This section provides a quick run-through of the steps to configure and manage deduplication.

Table 3) Deduplication quick overview.

	New Flexible Volume	**Flexible Volume with Existing Data**
Flexible Volume Configuration	Create flexible volume.	N/A
Enable Deduplication on Flexible Volume	`sis on <vol>`	
Initial Scan	N/A	Scan and deduplicate the existing data. `sis start -s <vol>`
Create, Modify, Delete Schedules (if not doing manually)	Delete or modify the default deduplication schedule that was configured when deduplication was first enabled on the flexible volume, or create the desired schedule. `sis config [-s sched] <vol>`	
Manually Run Deduplication (if not using schedules)	`sis start <vol>`	
Monitor Status of Deduplication (optional)	`sis status <vol>`	
Monitor Space Savings (optional)	`df –s <vol>`	

2.6 END-TO-END DEDUPLICATION CONFIGURATION EXAMPLE

This section steps through the entire process of creating a flexible volume and configuring, running, and monitoring deduplication on it.

Note: The steps are spelled out in detail, so the process appears a lot lengthier than it would be in the real world.

This example creates a place to archive several large data files. The destination NetApp storage system is called r200-rtp01, and it is assumed that deduplication has been licensed on this machine.

1. Create a flexible volume (keeping in mind the maximum allowable volume size for the platform, as specified in the requirements table at the beginning of this section).

```
r200-rtp01*> vol create VolArchive aggr0 200g
Creation of volume 'VolArchive' with size 200g on containing aggregate
'aggr0' has completed.
```

2. Enable deduplication on the flexible volume and verify that it's turned on. The `vol status` command shows the attributes for flexible volumes that have deduplication turned on.

   After you turn deduplication on, Data ONTAP lets you know that if this were an existing flexible volume that already contained data before deduplication was enabled, you would want to run `sis start –s`. In this example it's a brand-new flexible volume so that's not necessary.

```
r200-rtp01*> sis on /vol/VolArchive
Deduplication for "/vol/VolArchive" is enabled.
Already existing data could be processed by running "sis start -s
/vol/VolArchive".

r200-rtp01*> vol status VolArchive
Volume State      Status            Options
VolArchive online raid_dp, flex     nosnap=on
                           sis
Containing aggregate: 'aggr0'
```

3. Another way to verify that deduplication is enabled on the flexible volume is to check the output from running `sis status` on the flexible volume.

```
r200-rtp01*> sis status /vol/VolArchive
Path                   State      Status     Progress
/vol/VolArchive        Enabled    Idle       Idle for 00:00:20
```

4. Turn off the default deduplication schedule.

```
r200-rtp01*> sis config /vol/VolArchive
Path                           Schedule
/vol/VolArchive                sun-sat@0
r200-rtp01*> sis config -s - /vol/VolArchive
r200-rtp01*> sis config /vol/VolArchive
Path                           Schedule
/vol/VolArchive                   -
```

5. NFS-mount the flexible volume to `/testArchives` on a Solaris™ host called `sunv240-rtp01`, and copy lots of files from the users' directories into the new archive directory flexible volume. Here is the result from the host perspective.

```
root@sunv240-rtp01 # pwd
/testPSTs
root@sunv240-rtp01 # df -k .
Filesystem               kbytes    used   avail capacity  Mounted on
r200-rtp01:/vol/VolArchive 167772160 33388384 134383776  20% /testArchives
```

6. Next, examine the flexible volume, run deduplication, and monitor the status. Use the `df –s` command to examine the storage consumed and the space saved. Note that no space savings have been achieved by simply copying data to the flexible volume, even though deduplication is turned on. What has happened is

that all the blocks that have been written to this flexible volume since deduplication was turned on have had their fingerprints written to the change log file.

```
r200-rtp01*> df -s /vol/VolArchive
Filesystem                used      saved       %saved
/vol/VolArchive/          33388384    0           0%
```

7. Run deduplication on the flexible volume. These causes the change log to be processed, fingerprints to be sorted and merged, and duplicate blocks to be found.

```
r200-rtp01*> sis start /vol/VolArchive
The deduplication operation for "/vol/VolArchive" is started.
```

8. Use `sis status` to monitor the progress of deduplication.

```
r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Active     9211 MB Searched

r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Active     11 MB (0%) Done

r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Active     1692 MB (14%) Done

r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Active     10 GB (90%) Done

r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Active     11 GB (99%) Done

r200-rtp01*> sis status /vol/VolArchive
Path                        State      Status     Progress
/vol/VolArchive             Enabled    Idle       for 00:00:07
```

9. When `sis status` indicates that the flexible volume is once again in the Idle state, deduplication has finished running and you can check the space savings it provided in the flexible volume.

```
r200-rtp01*> df -s /vol/VolArchive
Filesystem                used      saved       %saved
/vol/VolArchive/          24072140  9316052       28%
```

That's all there is to it.

2.7 CONFIGURING DEDUPLICATION SCHEDULES

It's best to set up a schedule for deduplication so that you don't have to run it manually each time. This section provides some specifics about configuring schedules with deduplication.

The `sis config` command is used to configure and view deduplication schedules for flexible volumes. The usage syntax is shown below.

```
r200-rtp01*> sis help config
sis config [ [ -s schedule ] <path> | <path> ... ]
Sets up, modifies, and retrieves the schedule of deduplication volumes.
```

Run with no arguments, `sis config` returns the schedules for all flexible volumes that have deduplication enabled. The following example shows the four different formats the reported schedules can have.

```
toaster> sis config
Path                            Schedule
/vol/dvol_1                         -
/vol/dvol_2                     23@sun-fri
/vol/dvol_3                     auto
/vol/dvol_4                     sat@6
```

The meaning of each of these schedule types is as follows:

- On flexible volume `dvol_1`, deduplication is not scheduled to run.
- On flexible volume `dvol_2,` deduplication is scheduled to run every day from Sunday to Friday at 11 p.m.
- On flexible volume `dvol_3`, deduplication is set to auto schedule. This means that deduplication is triggered by the amount of new data written to the flexible volume, specifically when there are 20% new fingerprints in the change log.
- On flexible volume `dvol_4`, deduplication is scheduled to run at 6 a.m. on Saturday.

When the `-s` option is specified, the command sets up or modifies the schedule on the specified flexible volume. The schedule parameter can be specified in one of four ways:

```
[day_list][@hour_list]
[hour_list][@day_list]
-
auto
```

The `day_list` specifies which days of the week deduplication should run. It is a comma-separated list of the first three letters of the day: `sun`, `mon`, `tue`, `wed`, `thu`, `fri`, `sat`. The names are not case sensitive. Day ranges such as `mon-fri` can also be used. The default `day_list` is `sun-sat`.

The `hour_list` specifies which hours of the day deduplication should run on each scheduled day. The `hour_list` is a comma-separated list of the integers from 0 to 23. Hour ranges such as 8-17 are allowed.

Step values can be used in conjunction with ranges. For example, `0-23/2` means "every 2 hours." The default `hour_list` is 0; that is, midnight on the morning of each scheduled day.

If "`-`" is specified, there is no scheduled deduplication operation on the flexible volume.

The `auto` schedule causes deduplication to run on that flexible volume whenever there are 20% new fingerprints in the change log. This check is done in a background process and occurs every hour.

When deduplication is enabled on a flexible volume for the first time, an initial schedule is assigned to the flexible volume. This initial schedule is `sun-sat@0`, which means "once every day at midnight."

To configure the schedules shown earlier in this section, the following commands would be issued:

```
toaster> sis config -s - /vol/dvol_1
toaster> sis config -s 23@sun-fri /vol/dvol_2
toaster> sis config –s auto /vol/dvol3
toaster> sis config –s sat@6 /vol/dvol_4
```

3 SIZING FOR PERFORMANCE AND SPACE EFFICIENCY

This section discusses the deduplication behavior that you can expect. Information in this section comes from testing, observations, and knowledge of how deduplication functions.

3.1 DEDUPLICATION GENERAL BEST PRACTICES

This section contains deduplication best practices and lessons learned from internal tests and from deployments in the field.

- Deduplication consumes system resources and can alter the data layout on disk. Due to the application's I/O pattern, and the effect of deduplication on the data layout, the read and write I/O performance can vary considerably. The space savings and the performance impact vary significantly depending on the application and the data contents.
- NetApp recommends that the performance impact due to deduplication be carefully considered and measured in a test setup and taken into sizing considerations before deploying deduplication in performance-sensitive solutions. For more information on the impact of deduplication on other applications, contact the specialists at NetApp for their advice and test results of your particular application with deduplication.
- If there is a small amount of new data, run deduplication infrequently, because there's no benefit in running it frequently in such a case, and it consumes CPU resources. How often you run it depends on the rate of change of the data in the flexible volume.
- The more concurrent deduplication processes you're running, the more system resources are consumed.
- Given the previous two items, the best option is to do one of the following:
    - Use the auto mode so that deduplication runs only when significant additional data has been written to each particular flexible volume (this tends to naturally spread out when deduplication runs).
    - Stagger the deduplication schedule for the flexible volumes so that it runs on alternative days.
    - Run deduplication manually.
- If Snapshot copies are required, run deduplication before creating the Snapshot copy to minimize the amount of data before the data gets locked in to the copies. (Make sure that deduplication has completed before creating the copy.) If a Snapshot copy is created on a flexible volume before deduplication has a chance to run and complete on that flexible volume, this could result in lower space savings.
- If Snapshot copies are to be used, the Snapshot reserve should be greater than 0. (An exception to this might be in an FCP/iSCSI LUN scenario, where it is often set to zero for thin provisioning reasons.)
- For deduplication to run properly, you need to leave some free space for the deduplication metadata. For information about how much extra space to leave in the volume and in the aggregate, see section 3.3.2, "Deduplication Metadata Overhead."

3.2 DEDUPLICATION PERFORMANCE

This section discusses the performance aspects of deduplication.

Since deduplication is a part of Data ONTAP, it is tightly integrated with the WAFL® file structure. Because of this, deduplication is performed with high efficiency. It is able to leverage the internal characteristics of Data ONTAP to create and compare digital fingerprints, redirect data pointers, and free up redundant data areas.

However, the following factors can affect the performance of the deduplication process and the I/O performance of deduplicated volumes.

- The application and the type of data set being used

- The data access pattern (for example, sequential Vs. random access, the size and pattern of the I/O)
- The amount of duplicate data, the amount of total data, and the average file size
- The nature of the data layout in the volume
- The amount of changed data between deduplication runs
- The number of concurrent deduplication sessions
- Hardware platform—the amount of CPU/memory in the system
- Amount of load on the system
- Disk types ATA/FC, and the RPM of the disk
- Number of disk spindles in the aggregate

Because of these factors, NetApp recommends that the performance impact due to deduplication be carefully considered and measured in a test setup and taken into sizing considerations before deploying deduplication in performance-sensitive solutions.

3.2.1 The Performance of the Deduplication Operation

The performance of the deduplication operation itself varies widely depending on the factors listed above, and this determines how long it takes this background process to finish running.

On a FAS6080 with no other load on the system, we have seen deduplication performances of up to 120 MBytes/sec (running a single deduplication session). If multiple deduplication streams are running, this total bandwidth gets divided evenly into the number of streams.

To get an idea of how long it takes for a deduplication process to complete, let's say that the deduplication process is running on a flexible volume at 25MB/sec. If 1TB of new data has been added to the volume since the last deduplication update, this deduplication operation takes about 10 to 12 hours and to complete. (There are no configurable parameters that can tune the deduplication process; that is, the priority of this background process in Data ONTAP is fixed.)

3.2.2 Impact on the System During the Deduplication Process

The deduplication operation runs as a low priority background process on the system. However, it can still affect the performance of user I/O and other applications running on the system.

The number of deduplication processes that are running and the phase that each process is running in can cause performance impacts to other applications running on the system (up to eight deduplication processes can actively run at any time on a system). The following are some observations made when running deduplication on a FAS3050 system:

- With eight deduplication processes running, and no other processes running, deduplication uses 15% of the CPU in its least invasive phase, and nearly all the available CPU in its most invasive phase.
- When one deduplication process is running, there is a 0% to 15% performance degradation on other applications.
- With eight deduplication processes running, there may be as much as a 15% to 50% performance penalty on other applications running on the system.

3.2.3 The I/O Performance of Deduplicated Volumes

Write Performance to a Deduplicated Volume

The impact of deduplication on the write performance of a system is a function of the hardware platform that is being used, as well as the amount of load that is placed on the system.

If the load on a system is low—that is, for systems in which the CPU utilization is around 50% or lower—there is a negligible difference in performance when writing data to a deduplicated volume, and there is no noticeable impact on other applications running on the system. On heavily used systems, however, where the system is nearly saturated with the amount of load on it, the impact on write performance can be expected to be around

15% for most NetApp systems. The performance impact is more noticeable on higher-end systems than on lower-end systems. On the FAS6080 system, this performance impact can be as much as 35%. Note that these numbers are for FC drives; if ATA drives are used in a system, the performance impact would be greater.

Read Performance from a Deduplicated Volume

When data is read from a deduplication-enabled volume, the impact on the read performance varies depending on the difference between the deduplicated block layout compared to the original block layout. There is minimal impact on random reads.

Since deduplication alters the data layout on the disk, it can affect the performance of sequential read applications such as Dump Source, QSM/SV Source, SV Restore, and other sequential read-heavy applications. This impact is more noticeable with data sets that are zero-padded, or data sets that contain blocks with repeating patterns (such as applications that preinitialize data blocks to a value of zero). Significant performance degradation has been measured in sequential reads with these types of data patterns. The performance impact is also more significant on sequential reads from SATA drives as compared to FC drives. Therefore, if an application depends on sequential read performance, the impact of deduplication on read performance should be carefully considered before implementation.

3.2.4 The Performance Acceleration Module (PAM)

- The PAM card is available with Data ONTAP 7.3 and later.
- In environments where there are shared blocks that are read repeatedly, the PAM card can help reduce the number of disk reads, thus improving the read performance.
- The amount of performance improvement with the PAM card depends the duplication rate, the access rate, the active data set size, and the data layout.

3.3 DEDUPLICATION STORAGE SAVINGS

This section discusses storage savings that deduplication can be expected to deliver.

Comprehensive testing of various data sets has been performed to determine typical space savings in different environments. These results were obtained in three ways:

1. Running deduplication on various production data sets within NetApp.
2. NetApp systems deployed in the real world running deduplication.
3. NetApp and end users running a simulation tool on various data sets. See section 3.4, “Space Savings Estimation Tool,” for information about how to use this tool.

Table 4 summarizes the test results.

Table 4) Deduplication storage savings for various environments.

Data Types	Typical Space Savings	Range
Backup Data	90%	85-95%
VMware VMs	70%	50-90%
Geoseismic	55%	40-70%
Database Backups	55%	40-70%
Home Directories	35%	20-50%
CIFS Shares	35%	20-50%
E-mail PSTs	30%	20-40%
Mixed Enterprise Data	30%	20-40%
Document Archives	25%	20-30%
Engineering Archives	25%	20-30%

Note that nonrepeating archival data such as image files and encrypted data is generally not considered a good candidate for deduplication.

The results reported in Table 4 are considered realistic and typically achievable, but still conservative.

It is important to note that the space savings in the table are from deduplicating a data set one time, with the following exception. In cases where the data is being backed up or archived over and over again, the realized storage savings get better and better, achieving 20:1 (95%) in many instances.

3.3.1 Deduplication and Space Savings on Existing Data

A major benefit of deduplication is that it can be used to deduplicate existing data in the flexible volumes. It is realistic to assume that there will be Snapshot copies—perhaps many—of this existing data. Here's what happens when you run deduplication in this case.

When you first run deduplication on this flexible volume, the storage savings will probably be rather small or even nonexistent.

- Previous Snapshot copies expire, and as they do some small savings are realized, but they too are likely to be low.
- During this period of old Snapshot copies expiring, it is fair to assume that new data is being created on the flexible volume and that Snapshot copies being created.
- The storage savings may continue to stay low.
- When the last Snapshot copy that was created before deduplication was run is deleted, the storage savings should increase noticeably.

The question thus becomes when to run deduplication again. The best case is before creation of each and every Snapshot copy; this provides the most storage savings benefit. However, depending on the flexible volume size and possible performance impact on the system, this may not always be advisable.

3.3.2 Deduplication Metadata Overhead

This section discusses storage overhead that deduplication introduces. While deduplication can provide substantial storage savings in many environments, there is a small amount of storage overhead associated with it. This should be considered when sizing the flexible volume.

The total storage used by the deduplication metadata files is approximately 1% to 3% of the total data in the volume (that is, the size of the data before it is deduplicated). So for 1TB of total data, the metadata overhead would be approximately 10GB to 30GB. The breakdown of the overhead associated with the deduplication metadata is as follows:

- There is a fingerprint record for every 4KB data block, and the fingerprint records for all of the data blocks in the volume are stored in the fingerprint database file. There is an overhead of less than 1% associated with this database file.
- The size of the deduplication change log files depends on the rate of change of the data and on how frequently deduplication is run. This accounts for less than 1% overhead in the volume.
- Finally, when deduplication is running, it creates some temporary files that could account for up to 1% of the size of the volume. These temporary metadata files are deleted when the deduplication process has finished running.

In Data ONTAP 7.2.X, all of the above deduplication metadata files reside in the volume, and this metadata is therefore captured and locked in the Snapshot copies of the volume as well.

Starting with Data ONTAP 7.3, part of the metadata still resides in the volume and part of it resides in the aggregate outside of the volume. The fingerprint database and the change log files are located outside of the volume in the aggregate and are therefore not captured in Snapshot copies. This change enables deduplication to achieve higher space savings. However, the other temporary metadata files created during the deduplication

operation are still placed inside the volume. These temporary metadata files are deleted when the deduplication operation completes. However, if Snapshot copies are created during a deduplication operation, these temporary metadata files can get locked in Snapshot copies, and they remain there until the Snapshot copies are deleted.

The guideline for the amount of extra space that should be left in the aggregate or volume for the deduplication metadata overhead is as follows:

- If you're running Data ONTAP 7.2.X, leave about 3% extra space inside the volume on which you plan to run deduplication.
- If you're running Data ONTAP 7.3, leave about 1% extra space inside the volume on which you plan to run deduplication, and around 2% extra space outside the volume in the aggregate.

3.4 SPACE SAVINGS ESTIMATION TOOL (SSET)

The actual amount of data space reduction depends on the type of data. For this reason, the SSET tool should be used to analyze the actual data set and determine the effectiveness of deduplication on that particular data set.

When executed, the SSET crawls through all the files in the specified path and estimates the space savings that will be achieved by deduplication. Although actual deduplication space savings may deviate from what the estimation tool predicts, use and testing so far indicate that in general, the actual results are within +/-5% of the space savings that the tool predicts.

3.4.1 Overview of SSET

The SSET is available to NetApp system engineers and performs nonintrusive testing of the data set to determine the effectiveness of deduplication.

This tool is intended for use only by NetApp personnel to analyze data at current or prospective NetApp users. By installing this software, the user agrees to keep this tool and any results from this tool confidential between them and NetApp.

The deduplication Space Savings Estimator Tool is available for Linux® and Windows® systems, which have the data available locally or via CIFS/NFS. See the SSET `readme` file for complete usage information.

3.4.2 Limitations of the SSET

The SSET runs on either a Linux system or a Windows system.

It is limited to evaluating 2TB of data or less. If the path given contains more than 2TB, once the tool has processed the first 2 TB of data, the tool indicates that the maximum size has been reached and displays the results of the data that it has processed until that time (the rest of the data is ignored).

The tool is designed to examine data that is available either locally or via NFS/CIFS only.

For more information about SSET, read the SSET documentation. The SSET tool can be downloaded from the NetApp internal and PartnerCenter Web sites.

3.5 DEDUPLICATION LIMITATIONS

This section discusses what's supported and what's not supported, and the do's and don'ts. Some of this information may be covered elsewhere in this report as well.

3.5.1 General Caveats

Deduplication metadata (fingerprint file and change logs) is not deduplicated.

Other metadata (such as directory metadata) is not deduplicated either. Therefore, for heavily replicated directory environments with a large number of small files (for example, Web space), the amount of space savings that can be achieved may be low.

Backup of the deduplicated volume via NDMP is supported, but there is no space optimization when the data is written to tape because it's a logical operation. (This could actually be considered an advantage, since in this case the tape does not contain a proprietary format.)

When deduplication is used in an environment where quotas are used, the quotas cannot be oversubscribed on a volume. For example, a user with a quota limit of 1TB can't store more than 1TB of data in a deduplicated volume even if this data fits into less than 1TB of physical space on the storage system. Storage administrators can use the saved space as desired.

Only data in the active file system is deduplicated. Data pointed to by Snapshot copies that were created before deduplication is run are not released until the Snapshot copy is deleted.

3.5.2 Maximum Flexible Volume Size

The maximum flexible volume size limitation for deduplication varies based on the platform (this number depends primarily on the amount of system memory). When this limit is reached, writes to the volume fail just as they would with any other volume after it is full.

This could be important to consider if the flexible volumes are ever moved to a different platform with a smaller maximum flexible volume size.

Table 5 shows the maximum usable flexible volume size limits (including any `snap reserve` space) for the different NetApp storage system platforms. (Note that if a volume ever gets larger than this limit, and is later shrunk to a smaller size, deduplication cannot be enabled on that volume.)

Table 5) Maximum deduplicated volume sizes for Data ONTAP 7.2.X (starting with 7.2.5.1) and Data ONTAP 7.3.0.

FAS2020	FAS3020 N5200 FAS2050	FAS3050 N5500	FAS3040 FAS3140 N5300	R200	FAS3070 N5600	FAS6030 FAS6040 N7600 FAS3170	FAS6070 FAS6080 N7800
0.5TB	1TB	2TB	3TB	4TB	6TB	10TB	16TB

The maximum shared data limit per volume for deduplication is 16TB, regardless of the platform type. Once this limit is reached, there is no more deduplication of data in the volume, but writes to the volume continue to work successfully until the volume gets completely full.

Table 6 shows the maximum total data limit per deduplicated volume for each platform. This is the maximum amount of data that can be stored in a deduplicated volume. This limit is equal to the maximum volume size plus the maximum shared data limit. (For example, in an R200 system that can have a deduplicated volume of up to 4TB in size, 20TB of data can be stored; that is 4TB + 16TB = 20 TB.

Table 6) Maximum total data limit in a deduplicated volume for Data ONTAP 7.2.X (starting with 7.2.5.1) and Data ONTAP 7.3.0.

FAS2020	FAS3020 N5200 FAS2050	FAS3050 N5500	FAS3040 FAS3140 N5300	R200	FAS3070 N5600	FAS6030 FAS6040 N7600 FAS3170	FAS6070 FAS6080 N7800
16.5TB	17TB	18TB	19TB	20TB	22TB	26TB	32TB

3.5.3 Number of Deduplication Processes

A maximum of 8 deduplication processes can be run at the same time on one FAS system.

- If another flexible volume is ***scheduled*** to have deduplication run while 8 deduplication processes are already running, deduplication for this additional flexible volume is queued. For example, suppose that a user sets a default schedule (`sun-sat@0`) for 10 deduplicated volumes. Eight will run at midnight and the remaining 2 will be queued.
    - As soon as one of the 8 current deduplication processes completes, one of the queued ones starts; when another deduplication process completes, the second queued one starts.
    - Next time deduplication is scheduled to run on these same 10 flexible volumes, a round-robin paradigm is used so that the same volumes aren't always the first ones to run.
- With Data ONTAP 7.2.X, for *manually* triggered deduplication runs, if 8 deduplication processes are already running when a command is issued to start another one, the request fails and the operation is not queued. However, starting with Data ONTAP 7.3, the manually triggered deduplication runs are also queued if 8 deduplication operations are already running (including the `sis start –s` command).

4 DEDUPLICATION WITH OTHER NETAPP FEATURES

For the versions of Data ONTAP that are required to run deduplication with the NetApp features described in this section, read the section on Deduplication Limitations.

4.1 DEDUPLICATION AND SNAPSHOT COPIES

Deduplication only deduplicates data in the active file system, and that data could be locked in Snapshot copies taken before deduplication, causing reduced storage savings.

There are two types of data that can be locked in Snapshot copies:

- Data can be locked in a Snapshot copy if the copy is created before deduplication is run. This effect can be mitigated by always running deduplication before a Snapshot copy is created. (This sounds easy, but it could become complicated, considering the performance implications of deduplication and the best practice of running it infrequently.)
- Deduplication metadata could get locked in a Snapshot copy when the copy is created. In Data ONTAP 7.2.X, all the deduplication metadata resides in the volume. Starting with Data ONTAP 7.3.0, part of the metadata resides in the volume and part of it resides in the aggregate outside the volume. The fingerprint database and the change log files that are used in the deduplication process are located outside of the volume in the aggregate, and are therefore not captured in Snapshot copies. This change enables deduplication to achieve higher space savings. However, some other temporary metadata files created during the deduplication operation are still placed inside the volume. These temporary metadata files are deleted when the deduplication operation completes. (For the size of these temporary metadata files, see section 3.3.2, "Deduplication Metadata Overhead.") These temporary metadata files can get locked in Snapshot copies, if the copies are created during a deduplication operation. The metadata files remain locked until the Snapshot copies are deleted.

For deduplication to provide the most benefit when used in conjunction with Snapshot copies, the following best practices should be considered:

- Run deduplication before creating new Snapshot copies.
- Remove unnecessary Snapshot copies maintained in deduplicated volumes.
- If possible, reduce the retention time of Snapshot copies maintained in deduplicated volumes.
- Schedule deduplication only after significant new data has been written to the volume.
- Configure appropriate reserve space for the Snapshot copies.
- If the space used by Snapshot copies grows to more than 100%, it will cause `df -s` to report incorrect results, because some space from the active file system is being taken away by Snapshot and therefore actual savings from deduplication aren't reported.
- If snap reserve is 0, you should turn off the Snapshot auto-create schedule (this is the case in most LUN deployments).

4.2 DEDUPLICATION AND SNAPRESTORE

The SnapRestore® functionality is supported with deduplication, and it works in the same way with deduplication as it does without deduplication. If you're running Data ONTAP 7.3, please note the following.

Starting with Data ONTAP 7.3, the deduplication metadata files (the fingerprint database and the change log files) do not get restored when SnapRestore is executed, because they are located outside the volume in the aggregate. In this case, after the SnapRestore operation, there is not a fingerprint database file in the active file system for the data. This data, however, retains the original space savings. After SnapRestore, if deduplication is enabled on the volume, any new data written to the volume continues to be deduplicated. However, the deduplication process obtains space savings in the new data only, and does not deduplicate between the new data and the restored data. To run deduplication for all the data in the volume (and thus obtain higher space

savings), use the `sis start -s` command. This command builds the fingerprint database for all the data in the volume. Depending on the size of the logical data in the volume, this process can take a long time to complete.

Before using the `sis start -s` command, make sure that both the volume and the aggregate containing the volume have sufficient free space to accommodate the addition of the deduplication metadata. For information about how much extra space to leave for the deduplication metadata, see section 3.3.2, "Deduplication Metadata Overhead."

4.3 DEDUPLICATION AND THE VOL COPY COMMAND

When deduplicated data is copied by using the volume copy command, the copy of the data at the destination location inherits all the deduplication attributes and storage savings of the original data.

Starting with Data ONTAP 7.3, some of the deduplication metadata files do not get copied by the `vol copy` command, because they are located outside of the volume in the aggregate. In this case, there is no fingerprint database file in the destination volume for the data. However, the data retains the space savings. The deduplication process also continues for any new data written to the destination volume, and creates the fingerprint database for the new data. The deduplication process obtains space savings in the new data only, and does not deduplicate between the new data and the old data. To run deduplication for all the data in the cloned volume (and thus obtain higher space savings), use the `sis start -s` command. Depending on the size of the logical data in the volume, this process can take a long time to complete.

4.4 DEDUPLICATION AND FLEXCLONE VOLUMES

When a FlexClone® volume (cloned volume) is created:

- The FlexClone volume of a deduplicated volume is a deduplicated volume.
- The cloned volume inherits the deduplication configuration of the parent volume, such as the deduplication schedule.
- Starting with Data ONTAP 7.3, the deduplication metadata files (the fingerprint database and the change log files) do not get cloned, because they are located outside the volume in the aggregate. In this case, there is no fingerprint database file in the cloned volume for the data that came from the parent. However, the data in the cloned volume inherits the space savings of the original data. The deduplication process also continues for any new data written to the clone, and creates the fingerprint database for the new data. However, the deduplication process obtains space savings in the new data only, and does not deduplicate between the new data and the old data. To run deduplication for all the data in the cloned volume (and thus obtain higher space savings), use the `sis start -s` command. Depending on the size of the logical data in the volume, this process can take a long time to complete.

4.4.1 Volume Splitting

When a cloned volume is split from the parent volume, all of the data in the clone that was a part of the parent volume (that is, not including the data that was written to the clone volume after the clone was created) gets undeduplicated after the volume split operation. If deduplication is running on the clone volume, this data gets deduplicated again in subsequent deduplication operations on the volume.

4.5 DEDUPLICATION AND ACTIVE-ACTIVE CONFIGURATION

Although deduplication is not yet fully integrated with cluster failover, NetApp cluster services are supported with deduplication in the following manner upon failover to the partner node.

- Writes to the flexible volume have fingerprints written to the change log.
- In Data ONTAP 7.2.X, no `sis` administration operations or deduplication function. However, starting with Data ONTAP 7.3, the following commands are supported for partner volumes in takeover mode: `sis status, sis stat, sis on, sis off`.
- Also, starting with Data ONTAP version 7.3, for SV-NBU, block sharing is supported for partner volumes in takeover mode.

- Upon failback, normal deduplication operations continue and the updated change log is processed.

Since deduplication is a licensed option behind the NearStore option license, NetApp recommends having both nodes in an active-/active configuration licensed with the NearStore option and with deduplication.

Deduplication does not add any overhead in an active-active configuration other than additional disk I/O.

4.6 DEDUPLICATION AND SNAPMIRROR REPLICATION

Although there are substantial benefits to be achieved with deduplication alone, a complete storage solution typically involves the need to additionally mirror the data to another location for disaster recovery purposes.

Replication of the deduplicated volume is supported by using SnapMirror in two ways—volume SnapMirror and –qtree SnapMirror, as discussed in the next two subsections. Keep in mind that deduplication is supported only on NetApp storage systems that are running the NearStore option. So any flexible volume shown in the following figures with deduplication running, even if it's a SnapMirror "primary," is on a NearStore option licensed system.

NetApp recommends *not* using deduplication with *Sync* SnapMirror. Although technically it will work, the integration and scheduling of deduplication with Sync SnapMirror are complicated to implement in the type of rigorous real-world scenarios that demand synchronous replication.

For a complete discussion of SnapMirror, refer to TR-3390, "SnapMirror Deployment and Implementation Guide."

4.6.1 Replicating with Volume SnapMirror

A deduplicated volume can be replicated to a secondary storage system (destination) by using Volume SnapMirror, as shown in Figure 3.

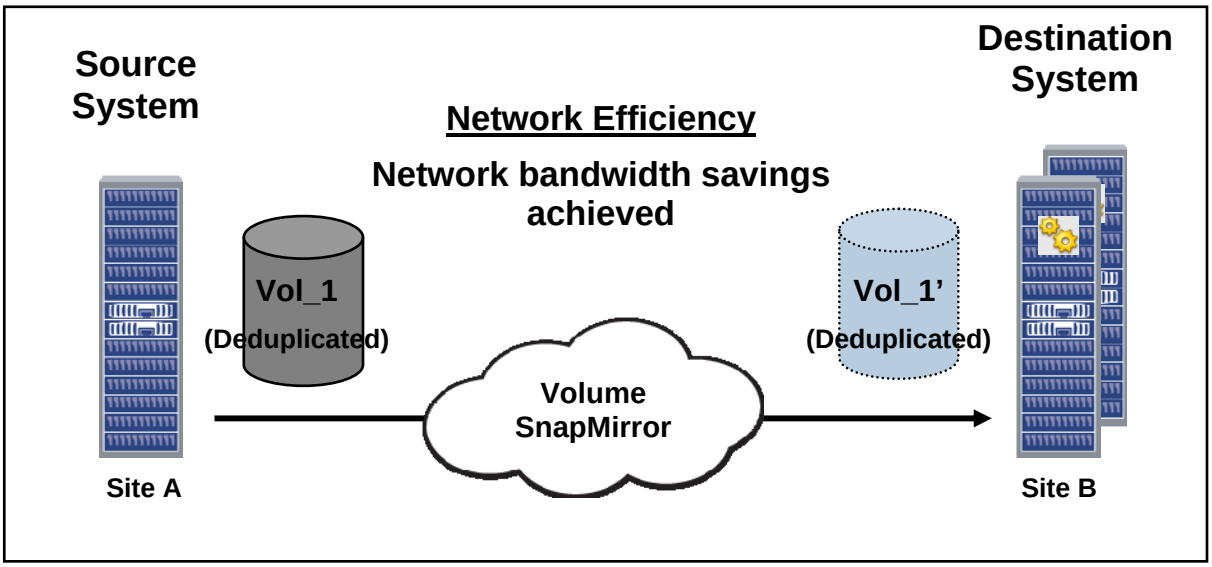

Figure 3) Volume SnapMirror replication of a deduplicated volume for disaster recovery.

To run deduplication with volume SnapMirror:

- Deduplication must be licensed at the primary location (source). However, the NearStore option must be licensed on both the source and destination (or an R200 must be used in one or both locations).
- Deduplication does not need to be licensed at the destination. However, if the primary site is down and the secondary location becomes the new primary, deduplication needs to be licensed for continued deduplication to occur. Therefore the best practice is to have deduplication licensed at both locations.
- Deduplication can be enabled, run, and managed only from the primary location. However, the flexible volume at the secondary location inherits all the deduplication attributes and storage savings via SnapMirror.

- Shared blocks are transferred only once, so deduplication reduces network bandwidth usage too.
- The volume SnapMirror update schedule is not tied to the deduplication schedule.
- The maximum volume size limit is imposed based on the lower maximum volume size limit of the source and destination volumes.

When configuring volume SnapMirror and deduplication, it is important to consider the deduplication schedule and the volume SnapMirror schedule. As a best practice, start volume SnapMirror transfers of a deduplicated volume after deduplication has completed (that is, not in the middle of the deduplication process). This is to avoid sending undeduplicated data and additional temporary metadata files over the network. If the temporary metadata files in the source volume are locked in Snapshot copies, they also consume extra space in the source and destination volumes.

Volume SnapMirror performance degradation can increase with deduplicated volumes. This extra overhead needs to be accounted for when sizing the storage solution. For more information, see the section on "Deduplication Performance."

The Impact of Moving Deduplication Metadata Files Outside the Volume

Starting with Data ONTAP 7.3, most of the deduplication metadata resides in the aggregate outside the volume. Therefore it does not get captured in Snapshot copies, and volume SnapMirror does not replicate this data. This provides additional network bandwidth savings. However, some temporary metadata files are still kept inside the volume and are deleted when the deduplication operation completes. If Snapshot copies are created during the deduplication operation, these temporary metadata files are locked in Snapshot copies, so a volume SnapMirror update that is initiated during a deduplication process transfers these temporary metadata files over the network. To prevent this extra data from being replicated, schedule the Volume SnapMirror updates to take place after the deduplication operation has finished running on the source volume.

In case of a disaster at the primary location, you may need to break the volume SnapMirror relationship and have the volume SnapMirror destination start serving data. In this case, there is no fingerprint database file at the destination for the existing data on the destination volume. However, the existing data retains the space savings from the deduplication operations performed earlier on the original volume SnapMirror source. Also, the deduplication process continues for new data being written to the volume and creates the fingerprint database for this new data. The deduplication process obtains space savings in the new data only, and doesn't deduplicate between the new data and the old data. To run deduplication for all the data in the volume (and thus obtain higher space savings), use the `sis start -s` command. This command builds the fingerprint database for all the data in the volume. Depending on the size of the logical data in the volume, this process may take a long time to complete.

Important: Before using the `sis start -s` command, make sure that both the volume and the aggregate containing the volume have sufficient free space to accommodate the addition of the deduplication metadata. For information about how much extra space to leave for the deduplication metadata, see section Deduplication Metadata Overhead.

4.6.2 Replicating with Qtree SnapMirror

When using qtree SnapMirror with deduplication, remember the following points:

- Deduplication can be enabled on the source system, the destination system, or both.
- Both the deduplication license and the SnapMirror license must be installed on the system where deduplication is required.
- Unlike volume SnapMirror, no network bandwidth savings are obtained with qtree SnapMirror, because the source system sends undeduplicated data to the destination system, even if deduplication is enabled on the source system.
- The deduplication schedule is not tied to qtree SnapMirror updates on either the source or the destination. However, a deduplication schedule can be set up independently of the qtree SnapMirror schedule. For example, on the destination, the deduplication process does not automatically start at the completion of qtree SnapMirror transfers.

As a best practice, NetApp recommends performing qtree SnapMirror updates after the deduplication process on the source volume has finished running. If a qtree SnapMirror update occurs while the deduplication process is running on the source volume, then in addition to the transfer of the changed data blocks, some unchanged

data blocks might also get transferred to the destination. If deduplication is not running on the destination volume, then the redundant data that is transferred occupies extra storage space on the destination volume.

NetApp also recommends that if deduplication is used on the source volume, then it should also be used on the destination volume. However, you don't have to use deduplication on the source volume if you are planning to use deduplication only on the destination volume.

As far as the qtree SnapMirror base Snapshot copy is concerned, there are typically only a couple of Snapshot copies on the destination storage system. If Snapshot copies are not retained long term, they are constantly rotated out and the deduplicated blocks are freed as the Snapshot copies roll off.

If users want to keep Snapshot copies long term (as a replacement for SnapVault, or for other reasons such as the ability to have writeable, reverse, or resync copies in the event of a disaster), it is possible that deduplicated data can be locked in Snapshot copies for longer periods of time, which reduces the deduplication storage savings. This situation can arise when users create Snapshot copies manually or by using `snap sched`.

The best practice when using qtree SnapMirror with deduplication is to let qtree SnapMirror use the minimum number of Snapshot copies it requires (essentially, keep the latest version).

Qtree SnapMirror Replication with Deduplication Enabled on the Source Only

A source deduplicated flexible volume can be replicated to a nondeduplicated volume on the destination by using qtree SnapMirror, as shown in Figure 4.

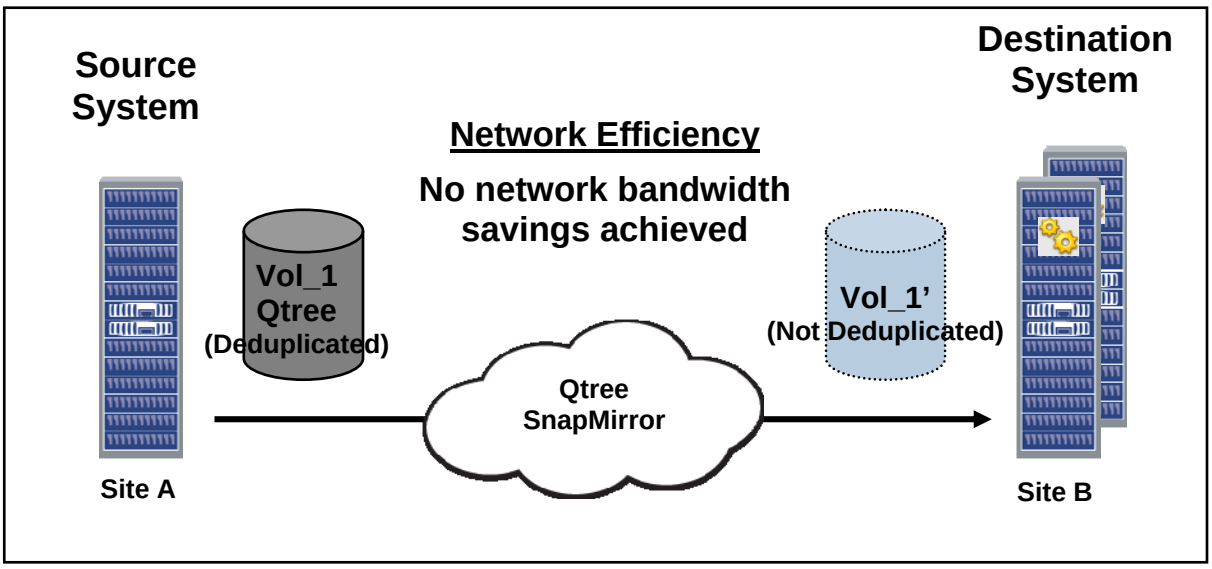

Figure 4) Qtree SnapMirror replication from a deduplicated source volume to a nondeduplicated destination volume.

Keep the following points in mind:

- Deduplication is licensed only on the source system.
- Deduplication is enabled, run, and managed on a flexible volume at the source.
- Deduplication doesn't yield any network bandwidth savings because qtree SnapMirror works at the logical layer and it sends undeduplicated data over the network.
- The deduplication schedule is not integrated with the qtree SnapMirror update, and vice versa; it must be configured independently. The completion of a deduplication process doesn't automatically start a qtree SnapMirror transfer, and qtree SnapMirror updates don't trigger the deduplication operation.
- Deduplication storage savings are achieved only on the source system.

Qtree SnapMirror Replication with Deduplication Enabled on the Destination Only

A nondeduplicated flexible volume on the source can be replicated to a deduplicated volume on the destination by using qtree SnapMirror, as shown in Figure 5.

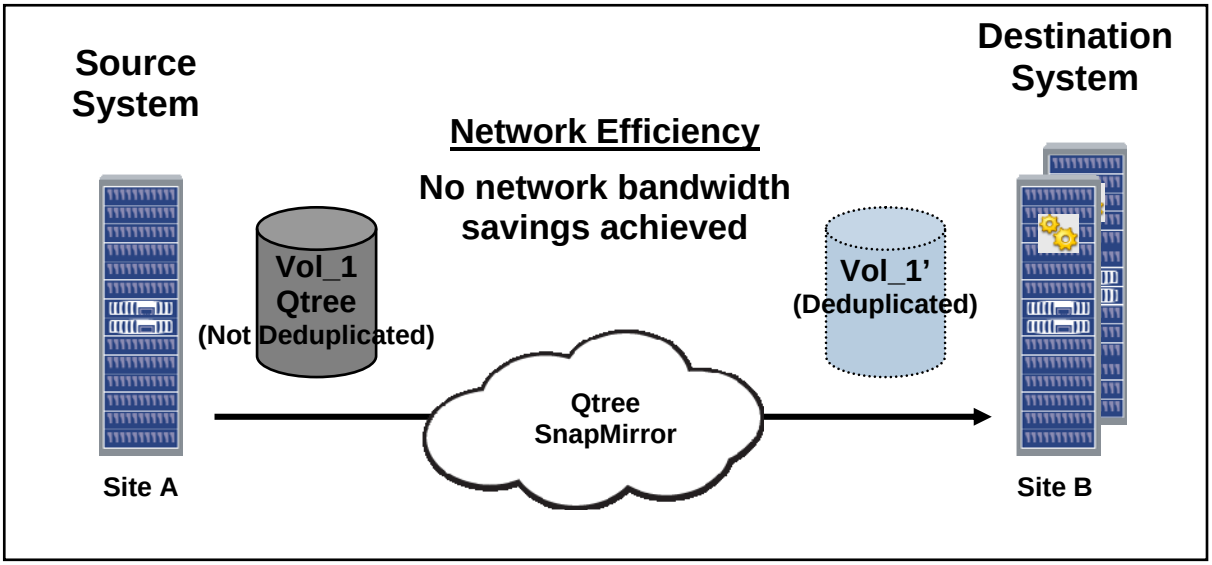

Figure 5) Qtree SnapMirror replication from a nondeduplicated source volume to a deduplicated destination volume.

Keep the following points in mind:

- Deduplication is licensed only on the destination system.
- Deduplication is enabled, run, and managed on a flexible volume at the destination.
- Deduplication doesn't yield any network bandwidth savings.
- The deduplication schedule is not integrated with the qtree SnapMirror update, and vice versa; it must be configured independently. The completion of a qtree SnapMirror update doesn't automatically start a deduplication operation on the destination, and the deduplication operation doesn't trigger a qtree SnapMirror update.
- Deduplication storage savings are achieved only on the destination system.

Qtree SnapMirror with Deduplication Enabled on Both the Source and the Destination

A deduplicated flexible volume on the source can be replicated to a deduplicated volume on the destination by using qtree SnapMirror, as shown in Figure 6.

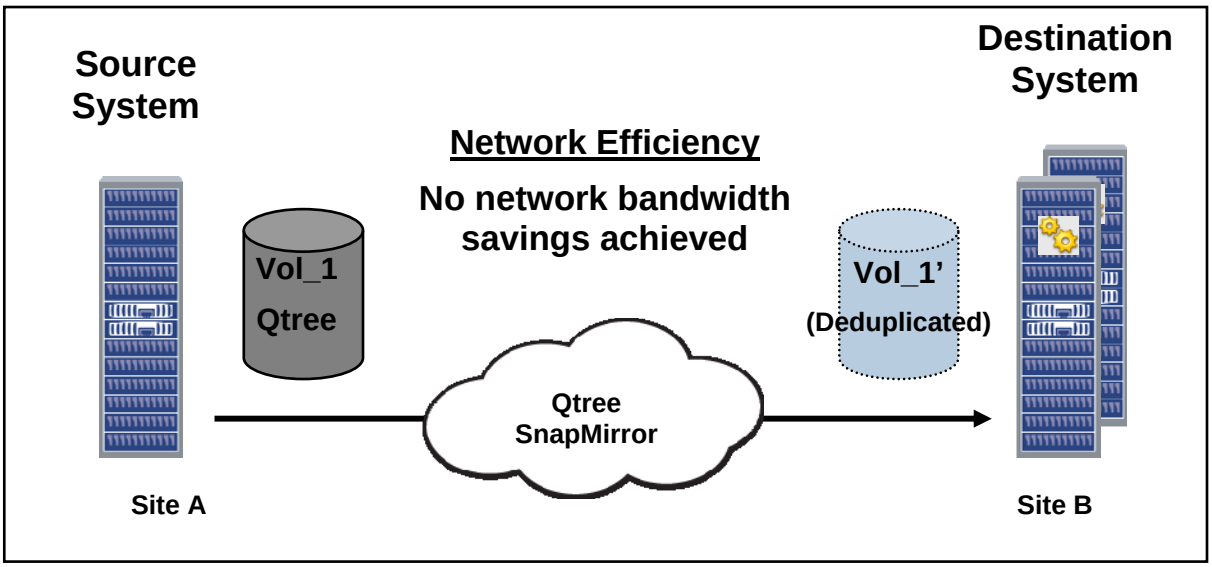

Figure 6) Qtree SnapMirror replication from a deduplicated source volume to a deduplicated destination volume.

Keep the following points in mind:

- Deduplication is licensed on both the source and the destination.
- Deduplication is enabled, run, and managed independently on the source and the destination.
- Deduplication doesn't yield any network bandwidth savings because qtree SnapMirror works at the logical layer, and it sends undeduplicated data over the network.
- Storage savings at the destination are not achieved automatically when qtree SnapMirror updates (unlike volume SnapMirror), because the data that is sent over the network to the destination is not deduplicated. This data must be deduplicated again on the destination after qtree SnapMirror has transferred the data from the source.
- The deduplication schedule is not integrated with the qtree SnapMirror update on either the source or the destination; it must be configured independently.
- Storage savings are achieved on both the source and the destination.

4.7 DEDUPLICATION AND SNAPVAULT

The behavior of deduplication with SnapVault is similar to the behavior of deduplication with qtree SnapMirror, except for the following points. (For information about other aspects of running deduplication with SnapVault, see section 4.2, "Replicating with SnapMirror."

- The deduplication schedule is tied to the SnapVault schedule on the destination system. The deduplication schedule on the source is not tied to the SnapVault update schedule, and it can be configured independently on a volume, just like qtree SnapMirror.
- Every SnapVault update (baseline or incremental) kicks off the deduplication process on the destination after the archival Snapshot is taken.
- The archival Snapshot copy is replaced with a new one after deduplication has finished running on the destination. (The name of this new Snapshot copy is the same as that of the archival copy, but the creation time of this copy is changed).

- The deduplication schedule on the destination cannot be configured manually, and the `sis start` command is not allowed either. However, the `sis start -s` command can be run manually on the destination.
- The SnapVault update is not tied to the deduplication operation; that is, a subsequent incremental update is allowed to run while the deduplication process on the destination volume from the previous backup is still in progress. In this case, the deduplication process continues to run, but the archival Snapshot copy does not get replaced after deduplication has finished running.

4.8 DEDUPLICATION AND SNAPVAULT FOR NETBACKUP

This section provides an overview of how deduplication integrates with SnapVault for NetBackup. The two modes of NetBackup—Space Optimized Image and File System Export—are discussed separately.

4.8.1 SnapVault for NetBackup—Space Optimized Image

SnapVault for NetBackup uses a mode called Space Optimized Image when it backs up supported databases. This provides the ability for the backup image to be aligned on WAFL 4KB block boundaries but not deduplicated real time and not exported. The removal of redundant information is deferred to background processing with deduplication that starts immediately after the backup has completed.

As NetBackup writes the backup data stream to a deduplicated volume on the NetApp Storage System, the following occurs:

- TAR stream is processed on the fly as it is received.
- TAR stream is written to disk (to allow NetBackup-initiated restores).
- Each TAR file header and file data are aligned on a 4KB block boundary.
- Archival Snapshot copy is created of the SnapVault for NetBackup volume.
- The TAR image and point-in-time state of AFS data are preserved.
- Deduplication scan is initiated after every backup, full or incremental.
- Common blocks are removed from independent unpacked file data.
- An archival Snapshot copy of the deduplicated volume is created, which replaces the original SnapVault for NetBackup archival Snapshot copy that was created before the deduplication run.

4.8.2 SnapVault for NetBackup—File System Export

Starting with Data ONTAP 7.3, volume-level deduplication can be enabled for unstructured data backups. This volume level deduplication is in addition to the automatic context-based inline deduplication, which works as described in the following paragraphs.

During the backup of unstructured data, SnapVault for NetBackup performs inline context-based deduplication on the TAR stream. After the initial backup, SnapVault for NetBackup writes only changed blocks to disk during subsequent backups, even though NetBackup sends the whole TAR ball. As these blocks are written to disk, they are aligned on WAFL 4KB block boundaries. This context-based deduplication of unstructured data takes place on the fly, and it is transparent to NetBackup.

Once deduplication is enabled, volume-level deduplication is completely automated, and on the successful competition of a backup, volume-level deduplication starts automatically for that flexible volume.

The archival Snapshot copy is replaced with a new one after deduplication has finished running on the destination. (The name of this new copy is the same as the archival copy, but the creation time of this Snapshot copy is changed).

The SnapVault update is not tied to the deduplication operation; that is, a subsequent incremental update is allowed to run while the deduplication process on the destination volume from the previous backup is still in progress. In this case, the deduplication process continues to run, but the archival Snapshot copy does not get replaced after deduplication has finished running.

4.9 DEDUPLICATION AND MULTISTORE (VFILER)

Starting with Data ONTAP 7.3, deduplication is fully supported with MultiStore.

Note: The deduplication commands are available only at the CLI of vFiler0; however, they allow any volume to be included in the command arguments, regardless of which vFiler™ the volume is associated with.

4.10 DEDUPLICATION AND LUNS

When using NetApp deduplication in a file-based (NFS/CIFS) environment, deduplication is straightforward and automatic; as duplicate blocks are freed, they are marked as available and the NetApp system recognizes these free blocks and makes them available to the volume.

Deduplication in a block-based (FCP/iSCSI) LUN environment is slightly more complicated. This is because of the space guarantees and fractional reservations often used by LUNs. With space guarantees, for instance, a 500GB LUN that is created consumes exactly 500GB of physical disk space. If the data in the LUN is reduced through deduplication, the LUN still reserves the same physical space capacity of 500GB, and the space savings are not apparent to the user.

LUN space guarantees and fractional reserves can be configured so that the use by the NetApp system of the freed blocks changes depending on the configuration. By varying the values of certain parameters, freed blocks can be returned to the LUN overwrite reserve, the volume free pool, or the aggregate free pool, or a combination.

This section describes five common examples of LUN configurations and deduplication behavior, as summarized in Table 7.

Table 7) LUN configuration examples (as described below).

	A (Default)	B	C	D	E
LUN space guarantee value	Yes	Yes	Yes	No	No
Volume fractional reserve value	100	1-99	0	Any	Any
Volume thin provisioned?	No	No	No	No	Yes
After deduplication and thin provisioning (if applicable), free blocks are returned to:	Fractional overwrite reserve	Fractional overwrite reserve + Volume free pool	Volume free pool	Volume free pool	Aggregate free pool

Definitions

- **–Fractional overwrite reserve:** The space that Data ONTAP guarantees will be available for overwriting blocks in a LUN when space guarantee = Yes. Behavior of the fractional reserve space parameter with deduplication is the same as if a Snapshot copy has been taken in the volume and blocks are being overwritten.
- **Volume free pool:** Refers to the free blocks in the parent volume of the LUN. These blocks can be assigned anywhere in the volume as needed.
- **Aggregate free pool:** Refers to the free blocks in the parent aggregate of the LUN. These blocks can be assigned anywhere in the aggregate as needed.

4.10.1 LUN Configuration Examples

Configuration A: The Default LUN Configuration

The default configuration of a NetApp LUN follows. (Best practice for all NetApp LUNs is to turn controller Snapshot off, delete all scheduled Snapshot copies, and set snap reserve to 0).

1) LUN Space Reservation value = On — Default = On
2) Volume Fractional Reserve Value = 100 — Default = 100%
3) Volume Guarantee = volume — Default = volume
4) Snap Reserve = 0% — Default = 20%
5) Autodelete = off — Default = off
6) Autosize = off — Default = off
7) Try_first = volume_grow — Default = volume_grow

Description: When a LUN containing default values is deduplicated, no apparent savings are observed by the storage administrator because the LUN by default was "space reserved" when it was created and fractional reserve was set to 100% in the volume. Any blocks freed through deduplication are allocated to the fractional reserve area. This configuration means that overwrite to the LUN should never fail, even if it is overwritten entirely.

Pros and Cons: The advantage of this configuration is that Snapshot copies consume less space when blocks in the active file system are no longer being used. As a result, this volume can hold more Snapshot copies. The disadvantage of this configuration is that free blocks are not returned to either the free volume pool or the free aggregate pool. Moreover, there is no direct space saving in the active file system—in fact, this configuration could consume more space in the volume due to new indirect blocks being written, if no Snapshot copies exist in the volume and the Snapshot schedule is turned off.

Note: If Snapshot copies are turned off for the volume (or no copy exists in the volume) this is **not** a recommended configuration or volume for deduplication.

Configuration B: LUN Configuration for Shared Volume Space Savings

If the user wants to apply the freed blocks to both the fractional overwrite reserve area and the volume free pool, this can be accomplished with the following configuration.

1) LUN Space Reservation value = On
2) Volume Fractional Reserve value = any value from 1 – 99
3) Volume Guarantee = volume
4) Snap Reserve = 0%
5) Autodelete = off
6) Autosize = off
7) Try_first = volume_grow

Description: The only difference between this configuration and configuration A is that the amount of space reserved for overwrite is based on the fractional reserve value set for the volume. As a result, this configuration splits the free blocks between fractional overwrite reserve and volume free space. For instance, if the fractional reserve value is set to 25, 25% of the freed blocks go into fractional overwrite reserve and 75% of the freed blocks are returned to the volume free pool.

Pros and Cons: The advantage of this configuration is that overwrite space reserve does not increase for every block being deduplicated. Freed blocks are split between volume free pool and fractional reserve. The disadvantage of this configuration is that overwrites to the LUN beyond the fractional reserve capacity may fail because freed blocks may have been already allocated. Another disadvantage of this configuration is that freed blocks stay in the parent volume and cannot be provisioned to any other volumes in the aggregate.

Note: If Snapshot copies are turned off for the volume (or if no Snapshot exists in the volume) and percentage of savings due to deduplication is less than the fractional reserve, then this is not a recommended configuration or volume for deduplication.

Configuration C: LUN Configuration for Maximum Volume Space Savings

If the user wants to apply the freed blocks to the volume free pool, this can be accomplished with the following configuration.

1) LUN Space Reservation value = On
2) Volume Fractional Reserve value = 0
3) Volume Guarantee = volume
4) Snap Reserve = 0%
5) Autodelete = off
6) Autosize = off
7) Try_first = volume_grow

Description: The only difference between this configuration and configuration B is that the value of fractional reserve is set to zero. As a result, this configuration "forces" all the freed blocks to the volume free pool and no blocks are set aside for fractional reserve.

Pros and Cons: The advantage of this configuration is that all the freed blocks are returned to the volume free pool. The disadvantage is that the chance of overwrite failure is higher than with configurations A and B because no freed blocks are assigned to the fractional overwrite area.

Configuration D: LUN Configuration for Maximum Volume Space Savings

If the user wants to apply the freed blocks to the volume free pool, this can be accomplished with the following configuration.

1) LUN Space Reservation value = Off
2) Volume Fractional Reserve value = any value from 0-100
3) Volume Guarantee = volume
4) Snap Reserve = 0%
5) Autodelete = off
6) Autosize = off
7) Try_first = volume_grow

Description: The difference between this configuration and configuration C is that the LUN is not space reserved. With LUN space guarantees off, the value for volume fractional reserve is ignored for all LUNs in this volume. From a deduplication perspective, there is no difference between this and the previous configuration, and all freed blocks go to the volume free pool.

Pros and Cons: From a deduplication perspective, this configuration has same advantages and disadvantages as configuration C.

Configuration E: LUN Configuration for Maximum Aggregate Space Savings

In many cases, the user may prefer to reclaim all freed blocks from the volume and return these blocks to the aggregate free pool. This is accomplished with the following configuration.

1) LUN Space Reservation value = Off
2) Volume Fractional Reserve value = any value from 0-100
3) Volume Guarantee = none
4) Snap Reserve = 0%
5) Autodelete = on
6) Autosize = on
7) Try_first = volume_grow

Description: This configuration "forces" the free blocks out of the volume and into the aggregate free pool, where the blocks can be reprovisioned for any other volumes in the aggregate.

Pros and Cons: The advantage of this configuration is that it provides the highest efficiency in aggregate space provisioning. It also uses the thin provisioning features of Data ONTAP; volume autosize and Snapshot autodelete, to help administer the space in the solution.

The disadvantage of this configuration is that it requires the storage administrator to monitor the free space available in the aggregates. With volume autosize and Snapshot autodelete turned on, the volume grows first if space is available in the aggregate; if not, then Snapshots are deleted.

5 DEDUPLICATION AND VMWARE

VMware® environments deduplicate extremely well. However, while working out the VMDK and datastore layouts, keep the following points in mind.

- Operating system VMDKs deduplicate extremely well because the binary files, patches, and drivers are highly redundant between virtual machines (VMs). Maximum savings can be achieved by keeping these in the same volume.
- Application binary VMDKs deduplicate to varying degrees. Duplicate applications deduplicate very well; applications from the same vendor commonly have similar libraries installed and deduplicate somewhat successfully; and applications written by different vendors don't deduplicate at all.
- Application data sets when deduplicated have varying levels of space savings and performance impact based on application and intended use. Careful consideration is needed, just as with nonvirtualized environments, before deciding to keep the application data in a deduplicated volume.
- Transient and temporary data such as VM swap files, pagefiles, and user and system temp directories, do not deduplicate well and potentially add significant performance pressure when deduplicated. Therefore NetApp recommends keeping this data on a separate VMDK and volume that are not deduplicated.

The expectation is that about 30% space savings will be achieved overall. This is a conservative number, and in some cases users have achieved savings of up to 80%. The major factor that affects this percentage is the amount of application data. New installations typically deduplicate extremely well, because they do not contain a significant amount of application data.

Important: In VMware, the need for proper partitioning and alignment of the VMDKs is extremely important (not just for deduplication). To help prevent the negative performance impact of LUN/VMDK misalignment, please read TR 3428, “NetApp/VMware Best Practices Guide,” at http://media.netapp.com/documents/tr-3428.pdf. Also note that the applications in which the performance is heavily affected by deduplication (when these applications are run without VMware) are likely to suffer the same performance impact from deduplication when they are run with VMware.

A deduplication and VMware solution on NFS is easy and straightforward. Combining deduplication and VMware with LUNs requires a bit more work. For more information on this, see section 4.10, “Deduplication and LUNS.”

The following subsections describe the different ways that VMware can be configured. For more information about NetApp storage in a VMware environment, see TR-3428, “Network Appliance and VMware ESX Server 3.0 Storage Best Practices.”

5.1.1 VMFS Data Store on Fibre Channel or iSCSI: Single LUN

This is the default configuration, and it's the way that a large number of VMware installations are done today. Deduplication occurs across the numerous VMDKs.

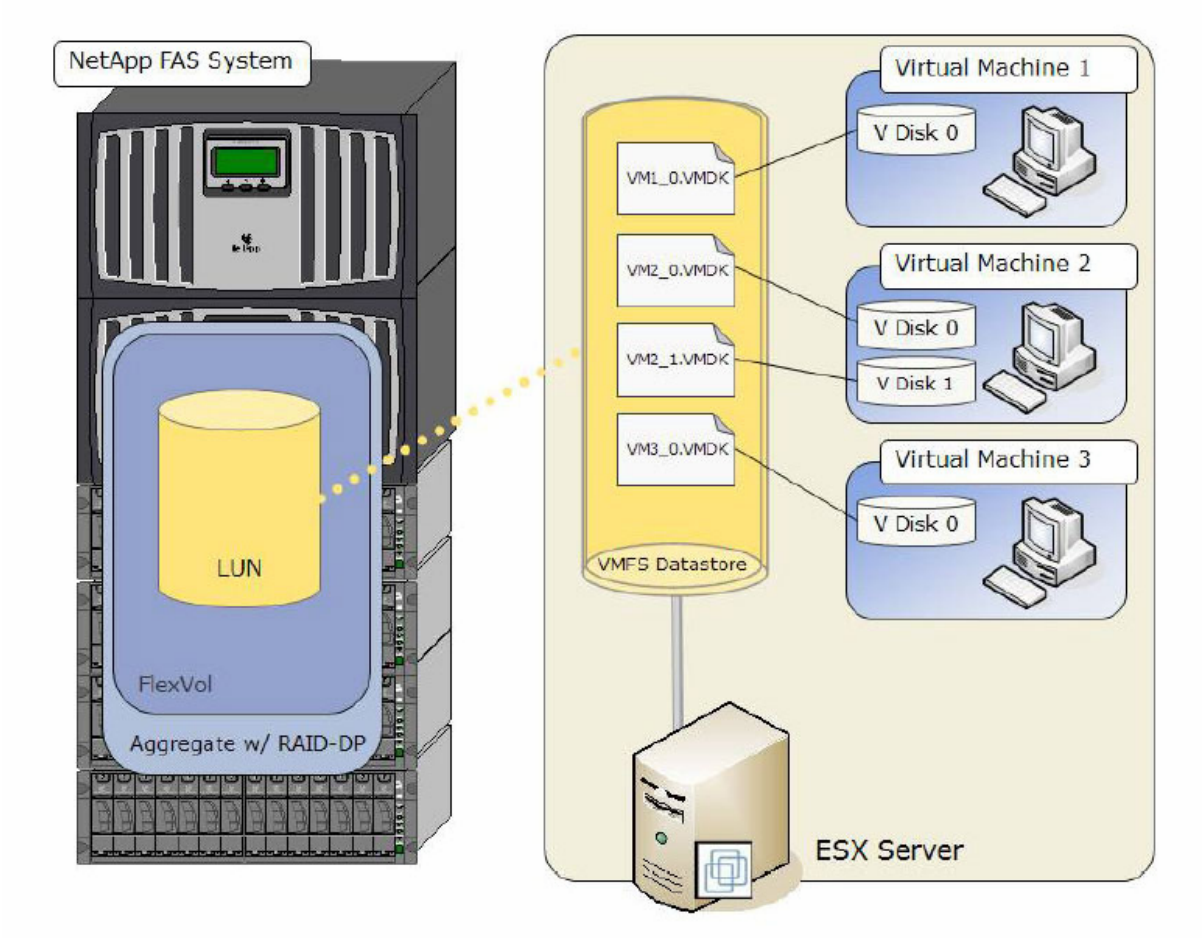

Figure 7) VMFS data store on Fibre Channel or iSCSI—single LUN.

5.1.2 VMware Virtual Disks Over NFS/CIFS

This is a new configuration that became available starting with VMware 3.0. It has a low installed base currently, but it is hot and growing. It is the easiest to configure and allows deduplication to provide the most space savings.

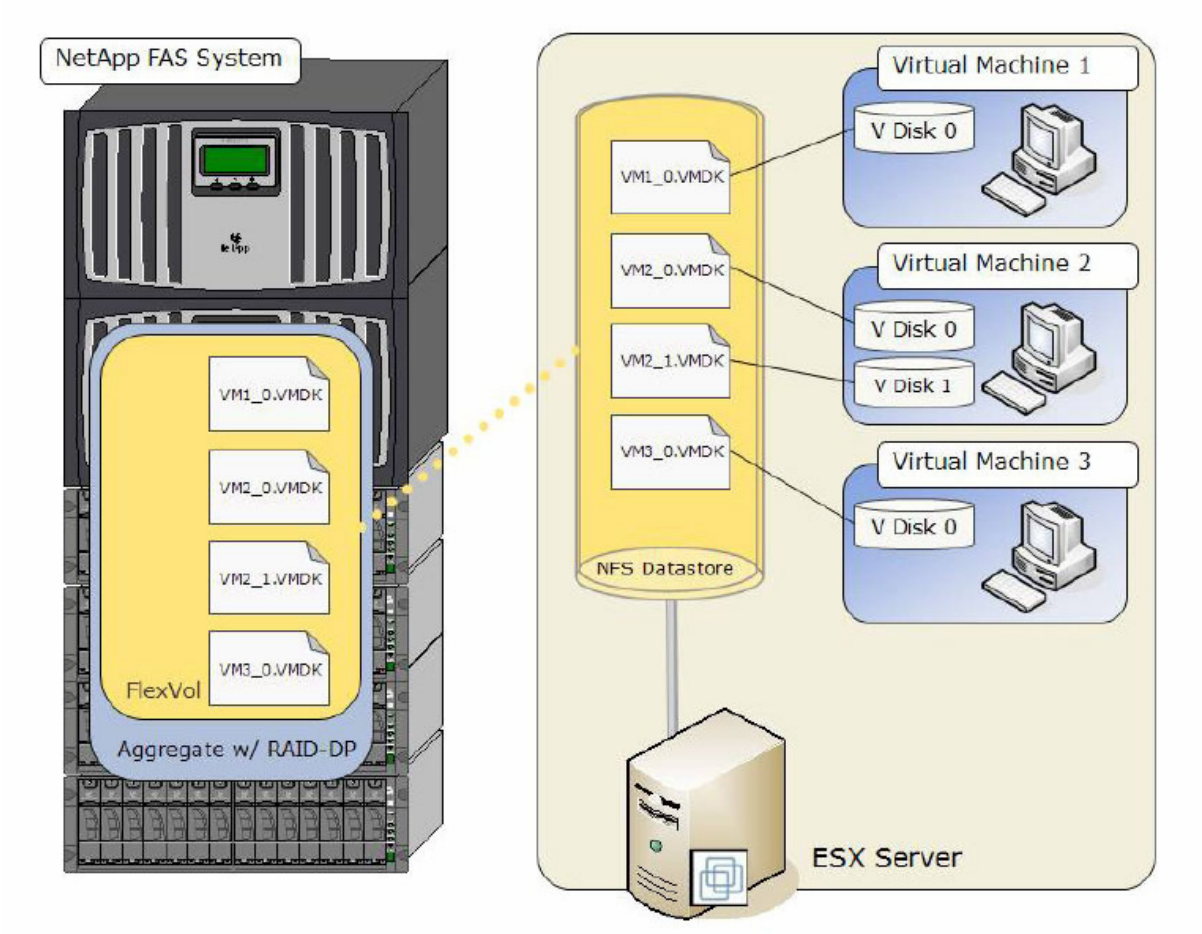

Figure 8) VMware virtual disks over NFS/CIFS.

5.1.3 Deduplication Archive of VMware

Deduplication has proven very useful in VMware archive environments. Figure 9 show an example.

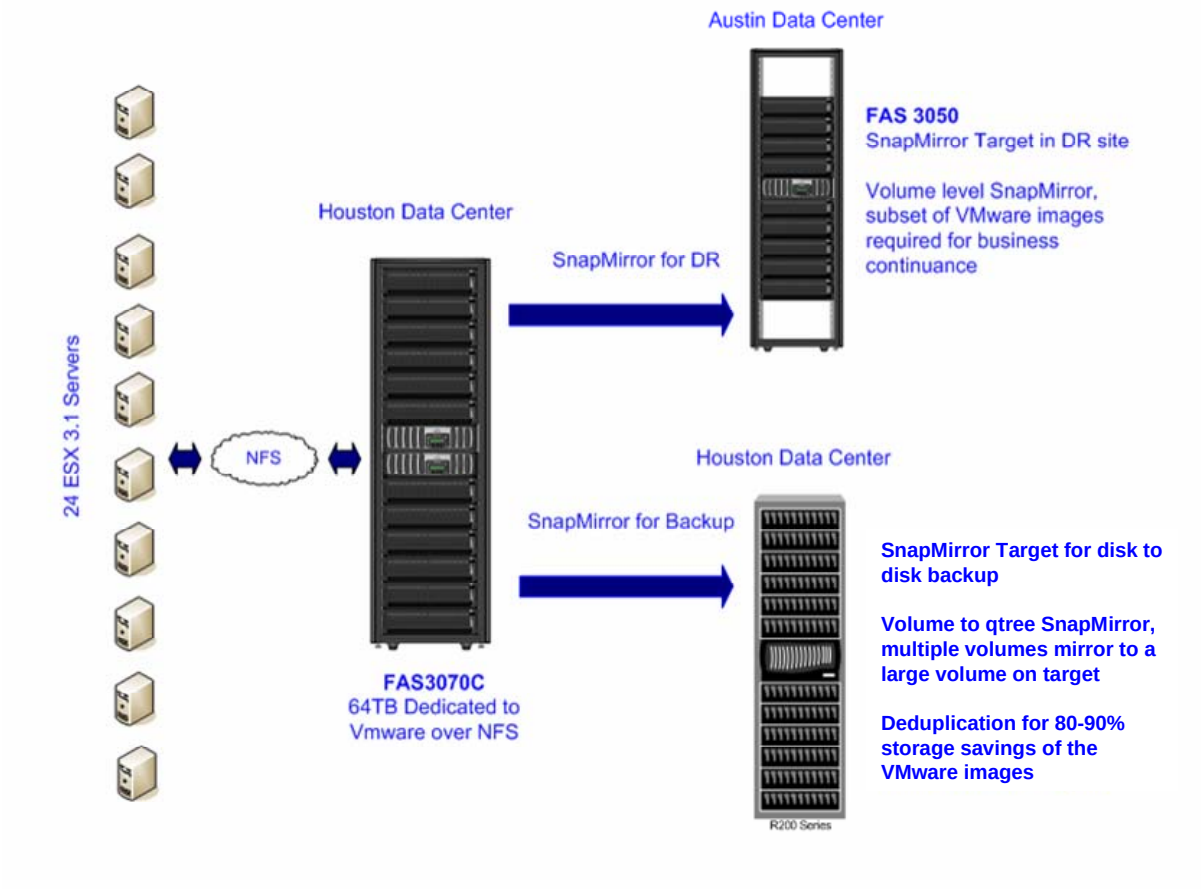

Figure 9) Archive of VMware with deduplication.

Detailed specifications for the example shown in Figure 9:

- In this environment, VMware is done via NFS.
- This environment use approximately 1,800 clone copies of their master VMware image. These images are used to create virtual machines for primary applications and for test and development purposes.
- All 1,800 clone copies (~32TB) are stored on a FAS3070 in the Houston data center.
- The data is mirrored to the remote site in Austin for disaster recovery.
- Once per hour, the FAS3070 images are transferred to an R200 by using SnapMirror.
- Deduplication is run nightly on the R200, and the VMware images are reduced in size by 80-90%.

6 TROUBLESHOOTING

This section covers issues that occasionally come up when configuring and running deduplication.

6.1 LICENSING

Make sure that deduplication is properly licensed and, if the platform is not an R200, make sure that the NearStore option is also properly licensed:

```
fas3070-rtp01*> license
…
a_sis <license key>
nearstore_option <license key>
…
```

If licensing is removed or expired, no additional deduplication can occur and no `sis` commands can run. However, the flexible volume remains a deduplicated volume, the existing storage savings are kept, and all data is usable.

6.2 VOLUME SIZES

From a deduplicated volume size limit perspective, a volume cannot exceed the size limit for the entire life of the volume (that is, if a volume is larger than the maximum size and is then shrunk, you still cannot enable deduplication on that volume). If you need to run deduplication on a volume that was (at some point in time) larger than the maximum supported size, you can do so by creating a new volume and migrating the data from the old volume to the newly created volume.

Here is an example of the message displayed if the volume is, or has been, too large to enable deduplication.

```
london-fs3> sis on /vol/projects
Volume or maxfiles exceeded max allowed for SIS: /vol/projects
```

Note that there must be space available for the `sis on` command to complete successfully. If you're running Data ONTAP 7.2.X, leave about 4% extra space inside the volume on which you plan to run deduplication. If you're running Data ONTAP 7.3, leave 1% extra space inside the volume, and 3% extra space outside the volume in the aggregate. This is done because in Data ONTAP 7.2.X, the metadata files for each volume reside inside the volume, and in Data ONTAP 7.3, some of the metadata files reside inside the volume and some files reside in the aggregate outside the volume.

6.3 LOGS AND ERROR MESSAGES

The location of the deduplication log file is:

```
/etc/log/sis
```

Error messages with explanations:

Registry errors:	Check if vol0 is full (only in Data ONTAP 7.2.X).
Metafile op errors:	Check if the deduplicated volume is full (in Data ONTAP 7.2.X).
Metafile op errors:	Check if the deduplicated aggregate is full (in Data ONTAP 7.2.X).
License errors:	Check if the license is installed.
Change log full error: metafile when finished	Perform a `sis start` operation to empty the change log.

6.4 NOT SEEING SPACE SAVINGS

If you've run deduplication on a flexible volume that you're confident contains data that should deduplicate well, but you are not seeing any space savings, there's a good chance that a bunch of Snapshot copies exist and are locking a lot of data. This tends to happen especially when deduplication is run on existing flexible volumes of data.

Use the `snap list` command to see what Snapshot copies exist and the `snap delete` command to remove them. Alternatively, wait for the Snapshot copies to expire and the space savings to appear (see section 4.1, “Deduplication and Snapshot Copies”).

6.5 UNDEDUPLICATING A FLEXIBLE VOLUME

It is possible, and relatively easy, to undeduplicate a flexible volume that has deduplication enabled by backing out deduplication and turning it back into a regular (nondeduplicated) flexible volume. This can be done while the flexible volume is online, as described below.

Turn deduplication off on the flexible volume.

Note: This command stops fingerprints from being written to the change log as new data is written to the flexible volume. If this command is used, and then deduplication is turned back on for this flexible volume, the flexible volume must be rescanned with the `sis start –s` command.

```
sis off <flexvol>
```

Use the following command[1] to recreate the duplicate blocks in the flexible volume.

```
sis undo <flexvol>
```

When this command completes, it deletes the fingerprint file and the change log files.

Here is an example of undeduplicating a flexible volume.

```
r200-rtp01*> df –s /vol/VolReallyBig2
/vol/VolReallyBig2/   20568276    3768732          15%
r200-rtp01*> sis status /vol/VolReallyBig2
Path                          State      Status     Progress
/vol/VolReallyBig2            Enabled    Idle       Idle for 11:11:13
r200-rtp01*> sis off /vol/VolReallyBig2
SIS for "/vol/VolReallyBig2" is disabled.
r200-rtp01*> sis status /vol/VolReallyBig2
Path                          State      Status     Progress
/vol/VolReallyBig2            Disabled   Idle       Idle for 11:11:34
r200-rtp01*> sis undo /vol/VolReallyBig2
Wed Feb  7 11:13:15 EST [wafl.scan.start:info]: Starting SIS volume scan on
volume VolReallyBig2.
r200-rtp01*> sis status /vol/VolReallyBig2
Path                          State      Status     Progress
/vol/VolReallyBig2            Disabled   Undoing    424 MB Processed
r200-rtp01*> sis status /vol/VolReallyBig2
No status entry found.
r200-rtp01*> df -s /vol/VolReallyBig2
Filesystem               used      saved       %saved
/vol/VolReallyBig2/   24149560          0           0%
```

Note: If sis `undo` starts processing and then there is not enough space to undeduplicate, it stops, sends a message about insufficient space, and leaves the flexible volume deduplicated. Use df –s to understand how much free space you really have, and then delete either data or Snapshot copies to provide the needed free space.

6.6 ADDITIONAL REPORTING WITH SIS STAT -1

For additional status information, you can use `priv set diag` and then use the `sis stat –l` command for long, detailed listings.

[1] The `undo` option of the `sis` command is available only in the diag mode, accessed via the `priv set diag` command.

6.7 DEDUPLICATION AND SYSTEM REBOOTS, HALTS, AND SO ON

If a NetApp storage system crashes, reboots, halts, and so on while deduplication is running, when it reboots, deduplication is in the “Idle” state for that flexible volume. There is no checkpoint or restart for deduplication, so the next scheduled or manual deduplication processing starts all over again, cleaning up all the intermediate sort bin files that were created by the previous deduplication operation. During this operation, the change logs are processed as follows:

- First process changelog1 that was processed before (fingerprint writes for new blocks are directed to changelog2).
- Then process the active changelog2 (fingerprint writes for new blocks are rotated to changelog1).

7 APPENDIX – ADDITIONAL READING AND REFERENCES

NetApp Documents

- TR-3390, SnapMirror Deployment and Implementation Guide
- TR-3428, Network Appliance and VMware ESX Server 3.0 Storage Best Practices
- TR-3465, SnapVault for NetBackup Deployment and Implementation Guide
- TR-3483, Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment
- TR-3488, SnapVault Design and Implementation Guide
- NetApp Data Online Backup and Recovery Guide