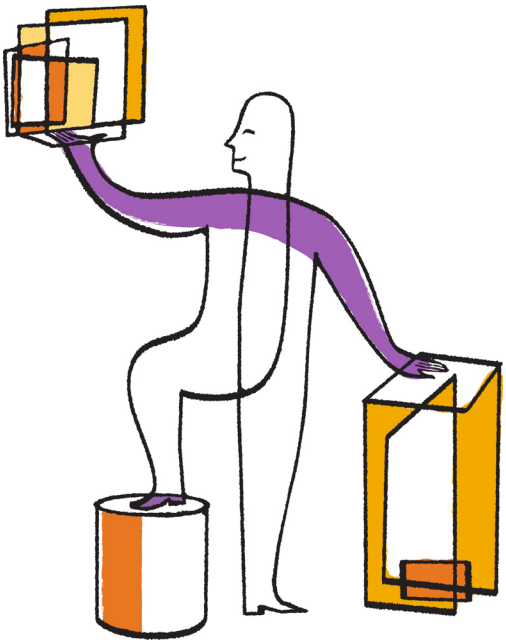
Data ONTAP® 8.2

Storage Management Guide

For 7-Mode



NetApp, Inc.
495 East Java Drive
Sunnyvale, CA 94089
U.S.

Telephone: +1 (408) 822-6000
Fax: +1 (408) 822-4501
Support telephone: +1 (888) 463-8277
Web: www.netapp.com
Feedback: doccomments@netapp.com

Part number: 215-07980_B0
October 2013

Contents

Data ONTAP storage architecture overview

Storage architecture refers to how Data ONTAP provides data storage resources to host or client systems and applications. Data ONTAP distinguishes between the physical layer of data storage resources and the logical layer.

- The physical layer includes drives, array LUNs, virtual disks, RAID groups, plexes, and aggregates.

  Note: A *drive* (or disk) is the basic unit of storage for storage systems that use Data ONTAP to access native disk shelves. An *array LUN* is the basic unit of storage that a storage array provides to a storage system that runs Data ONTAP. A *virtual disk* is the basic unit of storage for a storage system that runs Data ONTAP-v.

- The logical layer includes the file systems—volumes, qtrees, logical unit numbers (LUNs)—and the directories and files that store data.

  Note: LUNs are storage target devices in iSCSI and FC networks.

Aggregates provide storage to volumes. Aggregates can be composed of either drives or array LUNs, but not both. Data ONTAP organizes the drives or array LUNs in an aggregate into one or more RAID groups. RAID groups are then collected into one or two plexes, depending on whether RAID-level mirroring (SyncMirror) is in use. Aggregates can have two formats: 32-bit and 64-bit. An aggregate's format affects its maximum size.

Volumes are data containers. Clients can access the data in volumes through the access protocols supported by Data ONTAP. These protocols include Network File System (NFS), Common Internet File System (CIFS), HyperText Transfer Protocol (HTTP), Web-based Distributed Authoring and Versioning (WebDAV), Fibre Channel (FC), and Internet SCSI (iSCSI).

You can partition volumes and control resource usage using qtrees. You can create LUNs for use in a SAN environment, using the FC or iSCSI access protocols. Volumes, qtrees, and LUNs contain directories and files.

Related concepts

Managing disks using Data ONTAP on page 18
Managing array LUNs using Data ONTAP on page 74
How Data ONTAP uses RAID to protect your data and data availability on page 105
How you use aggregates to provide storage to your volumes on page 126
Using volumes on page 163
About qtrees on page 286

Managing disks using Data ONTAP

Disks provide the basic unit of storage for storage systems running Data ONTAP that use native disk shelves. Understanding how Data ONTAP uses and classifies disks will help you manage your storage more effectively.

How Data ONTAP reports drive types

Data ONTAP associates a type with every drive. Data ONTAP reports some drive types differently than the industry standards; you should understand how Data ONTAP drive types map to industry standards to avoid confusion.

When Data ONTAP documentation refers to a drive type, it is the type used by Data ONTAP unless otherwise specified. *RAID drive types* denote the role a specific drive plays for RAID. RAID drive types are not related to Data ONTAP drive types.

For a specific configuration, the drive types supported depend on the storage system model, the shelf type, and the I/O modules installed in the system. For more information about the types of drives supported by your configuration, see the *Hardware Universe* at *hwu.netapp.com*.

The following tables show how Data ONTAP drive types map to industry standard drive types for the SAS and FC-AL storage connection architectures, storage arrays, and for virtual storage (Data ONTAP-v):

Table 1: SAS storage connection architecture

Data ONTAP drive type	Primary drive characteristic	Industry standard drive type	Description
BSAS	Capacity	SATA	Bridged SAS–SATA disks with added hardware to enable them to be plugged into a SAS shelf.
FSAS	Capacity	NL-SAS	Near Line SAS
MSATA	Capacity	SATA	SATA disk in multi-disk carrier disk shelf
SAS	Performance	SAS	
SATA	Capacity	SATA	Available only as internal disks for FAS20xx systems.
SSD	High-performance	SSD	Solid-state drives

Table 2: FC-AL storage connection architecture

Data ONTAP drive type	Primary drive characteristic	Industry standard drive type	Description
ATA	Capacity	SATA	
FCAL	Performance	FC	

Table 3: Storage arrays

Data ONTAP drive type	Primary drive characteristic	Industry standard drive type	Description
LUN	N/A	LUN	A logical storage device backed by storage arrays and used by Data ONTAP as a drive. These LUNs are referred to as *array LUNs* to distinguish them from the LUNs that Data ONTAP serves to clients.

Table 4: Virtual storage (Data ONTAP-v)

Data ONTAP drive type	Primary drive characteristic	Industry standard drive type	Description
SAS	N/A	VMDK	Virtual drives that are formatted and managed by VMware ESX.

For information about best practices for working with different types of drives, see *Technical Report 3437: Storage Subsystem Resiliency Guide*.

Related concepts

Rules for mixing HDD types in aggregates on page 135

Related information

TR 3437: Storage Subsystem Resiliency Guide

Storage connection architectures and topologies supported by Data ONTAP

Data ONTAP supports two storage connection architectures: serial-attached SCSI (SAS) and Fibre Channel (FC). The FC connection architecture supports three topologies: arbitrated loop, switched, and point-to-point.

- SAS, SATA, BSAS, FSAS, SSD, and MSATA disks use the SAS connection architecture.
- FC and ATA disks use the FC connection architecture with an arbitrated-loop topology (FC-AL).
- Array LUNs use the FC connection architecture, with either the point-to-point or switched topology.

SAS-connected disk shelves are connected to the controller on a daisy chain called a *stack*. FC-connected disk shelves are connected to the controller on a loop. You cannot combine different connection architectures in the same loop or stack.

For the MetroCluster configuration, the FC and SAS connection architectures can be combined in a bridged connection, with FC on the controller side and SAS on the shelf side. The bridged connection can be used in either a direct-attached or switched topology. For more information, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges*.

How disks can be combined for the SAS disk connection type

You can combine SAS disk shelves and SATA disk shelves within the same stack, although this configuration is not recommended.

Each SAS-connected disk shelf can contain only one type of disk (SAS or SATA). The only exception to this rule is if the shelf is being used for a Flash Pool aggregate. In that case, for some SSD sizes and shelf models, you can combine SSDs and HDDs in the same shelf. For more information, see the *Hardware Universe*.

How disks can be combined for the FC-AL disk connection type

You cannot combine disk shelves containing FC disks and disk shelves containing ATA disks in the same loop.

Usable and physical disk capacity by disk size

You cannot use the nominal size of a disk in your aggregate and storage system capacity calculations. You must use either the usable capacity or the physical capacity as calculated by Data ONTAP.

The following table lists the approximate physical and usable capacities for the disk sizes currently supported by Data ONTAP. The numbers shown are in Mebibytes (MiBs). This unit of measure is equivalent to 2 to the 20th power bytes. (MBs, in contrast, are 10 to the sixth power bytes.)

The physical capacities listed in the table are approximations; actual physical disk capacities vary by manufacturer. The technical documentation for your disks contains the exact physical disk capacities.

Disk size as described by manufacturer	**Physical capacity (MiBs, approximate)**	**Usable capacity (MiBs)**
100 GB SSD (X441A-R5)	95,396	95,146
100 GB SSD (X442A-R5)	84,796	84,574
200 GB SSD	190,782	190,532
800 GB SSD	763,097	762,847
300 GB Performance	280,104	272,000
450 GB Performance	420,156	418,000
500 GB Capacity	423,946	423,111
600 GB Performance	560,208	560,000
900 GB Performance	858,483	857,000
1.2 TB Performance	1,144,641	1,142,352
1 TB Capacity	847,884	847,555
2 TB Capacity	1,695,702	1,695,466
3 TB Capacity (X308A, single-drive carrier)	2,543,634	2,538,546
3 TB Capacity (X478A, multi-drive carrier)	2,891,588	2,811,241
4 TB Capacity (X477A, single-drive carrier)	3,815,447	3,807,816
4 TB Capacity (X480A, multi-drive carrier)	3,815,447	3,748,319

Related concepts

Methods of calculating aggregate and system capacity on page 22

Methods of calculating aggregate and system capacity

You use the physical and usable capacity of the disks you employ in your storage systems to ensure that your storage architecture conforms to the overall system capacity limits and the size limits of your aggregates.

To maintain compatibility across different brands of disks, Data ONTAP rounds down (*right-sizes*) the amount of space available for user data. In addition, the numerical base used to calculate capacity (base 2 or base 10) also impacts sizing information. For these reasons, it is important to use the correct size measurement, depending on the task you want to accomplish:

- For calculating overall system capacity, you use the physical capacity of the disk, and count every disk that is owned by the storage system.
- For calculating how many disks you can put into an aggregate before you exceed its maximum size, you use the right-sized, or usable capacity of all data disks in that aggregate.

  Parity and dparity disks are not counted against the maximum aggregate size.

Related references

Usable and physical disk capacity by disk size on page 20

Disk speeds supported by Data ONTAP

For hard disk drives, which use rotating media, speed is measured in revolutions per minute (RPM). Faster disks provide more disk input/output operations per second (IOPS) and faster response time.

It is best to use disks of the same speed in an aggregate.

Data ONTAP supports the following rotational speeds for disks:

- SAS disks (SAS-connected)
  - 10K RPM
  - 15K RPM
- SATA, BSAS, FSAS, and MSATA disks (SAS-connected)
  - 7.2K RPM
- FCAL disks (FC-AL connected)
  - 10K RPM
  - 15K RPM
- ATA disks (FC-AL connected)
  - 5.4K RPM
  - 7.2K RPM

Solid-state disks, or SSDs, are flash memory-based devices and therefore the concept of rotational speed does not apply to them. SSDs provide more IOPS and faster response times than rotating media.

For more information about supported disk speeds, see the *Hardware Universe* at *hwu.netapp.com*.

Related concepts

How you can use disks with mixed speeds in the same aggregate on page 134
How you use aggregates to provide storage to your volumes on page 126

How disk checksum types affect aggregate and spare management

There are two checksum types available for disks used by Data ONTAP: BCS (block) and AZCS (zoned). Understanding how the checksum types differ and how they impact storage management enables you to manage your storage more effectively.

Both checksum types provide the same resiliency capabilities; BCS optimizes data access speed and capacity for disks that use 520 byte sectors. AZCS provides enhanced storage utilization and capacity for disks that use 512 byte sectors (usually SATA disks, which emphasize capacity).

Aggregates have a checksum type, which is determined by the checksum type of the disks that compose the aggregate. The following configuration rules apply to aggregates, disks, and checksums:

- Checksum types cannot be combined within RAID groups.
  This means that you must consider checksum type when you provide hot spare disks.
- When you add storage to an aggregate, if it has a different checksum type than the storage in the RAID group to which it would normally be added, Data ONTAP creates a new RAID group.
- An aggregate can have RAID groups of both checksum types.
  These aggregates have a checksum type of `mixed`.
- For mirrored aggregates, both plexes must have the same checksum type.
- Disks of a different checksum type cannot be used to replace a failed disk.
- You cannot change the checksum type of a disk.

Checksum type by Data ONTAP disk type

You should know the Data ONTAP disk type and checksum type of all of the disks you manage, because these disk characteristics impact where and when the disks can be used.

The following table shows the checksum type by Data ONTAP disk type:

Data ONTAP disk type	Checksum type
SAS or FC-AL	BCS

Data ONTAP disk type	Checksum type
SATA/BSAS/FSAS/ATA	BCS
SSD	BCS
MSATA	AZCS

Drive name formats

Each drive has a name that differentiates it from all other drives. Drive names have different formats, depending on the connection type (FC-AL or SAS) and how the drive is attached.

The following table shows the various formats for drive names, depending on how they are connected to the storage system.

Note: For internal drives, the slot number is zero, and the internal port number depends on the system model.

Drive connection	Drive name	Example
SAS, direct-attached	<slot><port>.<shelfID>.<bay>	The drive in shelf 2, bay 11, connected to onboard port 0a is named 0a.2.11. The drive in shelf 6, bay 3, connected to an HBA in slot 1, port c, is named 1c.6.3.
SAS, direct-attached in multi-disk carrier disk shelf	<slot><port>.<shelfID>.<bay>L<carrierPosition>	Carrier position is 1 or 2.

Drive connection	**Drive name**	**Example**
SAS, direct-attached, for systems running Data ONTAP-v	<slot><port>.<ID>	The third virtual disk connected to the first port is named 0b.3. The second virtual disk connected to the third port is named 0d.2. The range of ports is b through e, and the range of disks is 0 through 15.
SAS, bridge-attached (FibreBridge, used for MetroCluster configurations)	<slot><port>.<loopID>L<LUN>	The drive with LUN 2 behind the bridge connected to port a in slot 3, loop ID 125, is named 3a.125L2.
SAS, bridge-attached through a switch (FibreBridge, used for MetroCluster configurations)	<switch_name>:<switch_port>.<loopID>L<LUN>	The drive with LUN 5 behind the bridge connected to port 2 of switch brcd44, loop ID 126, is named brcd44:2.126L5.
FC-AL, direct-attached	<slot><port>.<loopID>	The drive with loop ID 19 (bay 3 of shelf 1) connected to onboard port 0a is named 0a.19. The drive with loop ID 34 connected to an HBA in slot 8, port c is named 8c.34.
FC-AL, switch-attached	<switch_name>.<switch_port>.<loopID>	The drive with loop ID 51 connected to port 3 of switch SW7 is named SW7.3.51.

For information about determining the LUN for drives using the FibreBridge connection architecture, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges*.

Loop IDs for FC-AL connected disks

For disks connected using Fibre Channel-Arbitrated Loop (FC-AL or FC), the loop ID is an integer between 16 and 126. The loop ID identifies the disk within its loop, and is included in the disk name, which identifies the disk uniquely for the entire system.

The loop ID corresponds to the disk shelf number and the bay in which the disk is installed. The lowest loop ID is always in the far right bay of the first disk shelf. The next higher loop ID is in the next bay to the left, and so on. You can view the device map for your disk shelves with the `fcadmin device_map` command.

For more information about the loop ID map for your disk shelf, see the hardware guide for the disk shelf.

Understanding RAID disk types

Data ONTAP classifies disks as one of four types for RAID: data, hot spare, parity, or dParity. The RAID disk type is determined by how RAID is using a disk; it is different from the Data ONTAP disk type.

Data disk	Holds data stored on behalf of clients within RAID groups (and any data generated about the state of the storage system as a result of a malfunction).
Spare disk	Does not hold usable data, but is available to be added to a RAID group in an aggregate. Any functioning disk that is not assigned to an aggregate but is assigned to a system functions as a hot spare disk.
Parity disk	Stores row parity information that is used for data reconstruction when a single disk drive fails within the RAID group.
dParity disk	Stores diagonal parity information that is used for data reconstruction when two disk drives fail within the RAID group, if RAID-DP is enabled.

How disk sanitization works

Disk sanitization is the process of physically obliterating data by overwriting disks with specified byte patterns or random data so that recovery of the original data becomes impossible. You use the sanitization process to ensure that no one can recover the data on the disks.

Related tasks

Using disk sanitization to remove data from disks on page 46
Destroying data on disks using Storage Encryption on page 101

Disk sanitization process

Understanding the basics of the disk sanitization process helps you understand what to anticipate during the sanitization process and after it is complete.

The disk sanitization process uses three successive default or user-specified byte overwrite patterns for up to seven cycles per operation. The random overwrite pattern is repeated for each cycle.

Depending on the disk capacity, the patterns, and the number of cycles, the process can take several hours. Sanitization runs in the background. You can start, stop, and display the status of the sanitization process.

The sanitization process contains two phases:

1. Formatting phase
   - For capacity HDDs (SATA, BSAS, FSAS, MSATA, or ATA) the formatting phase is skipped.
   - For performance HDDs (SAS or FC), the formatting phase consists of a SCSI format operation.
   - For SSDs, the formatting phase consists of a SCSI sanitize operation.
2. Pattern overwrite phase
   The specified overwrite patterns are repeated for the specified number of cycles.

When the sanitization process is complete, the specified disks are in a sanitized state. They are not returned to spare status automatically. You must return the sanitized disks to the spare pool before the newly sanitized disks are available to be added to another aggregate.

When disk sanitization cannot be performed

Disk sanitization is not supported for all disk types. In addition, there are times when disk sanitization cannot be performed.

You should be aware of the following facts about the disk sanitization process:

- It is not supported on all SSD part numbers.
  For information about which SSD part numbers support disk sanitization, see the *Hardware Universe* at *hwu.netapp.com*.
- It is not supported in takeover mode for systems in an HA pair.
- It cannot be performed on disks that were failed due to readability or writability problems.
- It cannot be performed on disks that belong to an SEC 17a-4-compliant SnapLock volume until the expiration periods on all files have expired--that is, all of the files have reached their retention dates.
- It does not perform its formatting phase on ATA drives.
- If you are using the random pattern, it cannot be performed on more than 100 disks at one time.
- It is not supported on array LUNs.
- If you sanitize both SES disks in the same ESH shelf at the same time, you see errors on the console about access to that shelf, and shelf warnings are not reported for the duration of the sanitization.

However, data access to that shelf is not interrupted.

- You can perform disk sanitization on disks using Storage Encryption.
  However, there are other methods to obliterate data on disks using Storage Encryption that are faster and do not require an operational storage system.

What happens if disk sanitization is interrupted

Disk sanitization is a long-running operation. If disk sanitization is interrupted by user intervention or an unexpected event such as a power outage, Data ONTAP takes action to return the disks that were being sanitized to a known state, but you must also take action before the sanitization process can finish.

If the sanitization process is interrupted by power failure, system panic, or manual intervention, the sanitization process must be repeated from the beginning. The disk is not designated as sanitized.

If the formatting phase of disk sanitization is interrupted, Data ONTAP must recover any disks that were corrupted by the interruption. After a system reboot and once every hour, Data ONTAP checks for any sanitization target disk that did not complete the formatting phase of its sanitization. If any such disks are found, Data ONTAP recovers them. The recovery method depends on the type of the disk. After a disk is recovered, you can rerun the sanitization process on that disk; for HDDs, you can use the `-s` option to specify that the formatting phase is not repeated again.

How selective disk sanitization works

Selective disk sanitization consists of physically obliterating data in specified files or volumes while preserving all other data located on the affected aggregate for continued user access. Because a file can be stored on multiple disks, there are three parts to the process.

To selectively sanitize data contained in an aggregate, you must carry out three general tasks:

1. Delete the files, directories or volumes that contain the data you want to sanitize from the aggregate that contains them.
2. Migrate the data that you want to preserve to a new set of disks in a destination aggregate on the same storage system.
   You can migrate data using the `ndmpcopy` command or qtree SnapMirror.
3. Destroy the original aggregate and sanitize all the disks that were RAID group members in that aggregate.

Related tasks

Removing data from disks using selective disk sanitization on page 49

Tips for creating and backing up aggregates containing data to be sanitized

If you are creating or backing up aggregates to contain data that might need to be sanitized, following some simple guidelines will reduce the time it takes to sanitize your data.

- Make sure your aggregates containing sensitive data are not larger than they need to be.

If they are larger than needed, sanitization requires more time, disk space, and bandwidth.

- When you back up aggregates containing sensitive data, avoid backing them up to aggregates that also contain large amounts of nonsensitive data.
  This will reduce the resources required to move nonsensitive data before sanitizing sensitive data.

How Data ONTAP monitors disk performance and health

Data ONTAP continually monitors disks to assess their performance and health. When Data ONTAP encounters certain errors or behaviors from a disk, it takes the disk offline temporarily or takes the disk out of service to run further tests.

What happens when Data ONTAP takes disks offline

Data ONTAP temporarily stops I/O activity to a disk and takes a disk offline when Data ONTAP is updating disk firmware in background mode or when disks become non-responsive. While the disk is offline, Data ONTAP performs a quick check on it to reduce the likelihood of forced disk failures.

A disk can be taken offline only if its containing RAID group is in a normal state and the plex or aggregate is not offline.

While the disk is offline, Data ONTAP reads from other disks within the RAID group while writes are logged. When the offline disk is ready to come back online, Data ONTAP resynchronizes the RAID group and brings the disk online. This process generally takes a few minutes and incurs a negligible performance impact.

How Data ONTAP reduces disk failures using Rapid RAID Recovery

When Data ONTAP determines that a disk has exceeded its error thresholds, Data ONTAP can perform Rapid RAID Recovery by removing the disk from its RAID group for testing and, if necessary, failing the disk. Spotting disk errors quickly helps prevent multiple disk failures and allows problem disks to be replaced.

By performing the Rapid RAID Recovery process on a suspect disk, Data ONTAP avoids three problems that occur during sudden disk failure and the subsequent RAID reconstruction process:

- Rebuild time
- Performance degradation
- Potential data loss due to additional disk failure during reconstruction

During Rapid RAID Recovery, Data ONTAP performs the following tasks:

1. Places the suspect disk in pre-fail mode.
2. Selects a hot spare replacement disk.

   Note: If no appropriate hot spare is available, the suspect disk remains in pre-fail mode and data continues to be served. However, a suspect disk performs less efficiently. Impact on performance ranges from negligible to worse than degraded mode. For this reason, make sure hot spares are always available.

3. Copies the suspect disk's contents to the spare disk on the storage system before an actual failure occurs.
4. After the copy is complete, attempts to put the suspect disk into the maintenance center, or else fails the disk.

   Note: Tasks 2 through 4 can occur only when the RAID group is in normal (not degraded) mode.

   If the suspect disk fails on its own before copying to a hot spare disk is complete, Data ONTAP starts the normal RAID reconstruction process.

   A message is sent to the log file when the Rapid RAID Recovery process is started and when it is complete. The messages are tagged "raid.rg.diskcopy.start:notice" and "raid.rg.diskcopy.done:notice".

Related concepts

About degraded mode on page 115
When Data ONTAP can put a disk into the maintenance center on page 31
How Data ONTAP works with hot spare disks on page 113

How the maintenance center works

When a disk is in the maintenance center, it is subjected to a number of tests. If the disk passes all of the tests, it is redesignated as a spare. Otherwise, Data ONTAP fails the disk.

The maintenance center is controlled by the `disk.maint_center.enable` option. It is on by default.

Data ONTAP puts disks into the maintenance center only if there are two or more spares available for that disk.

You can control the number of times a disk is allowed to go to the maintenance center using the `disk.maint_center.allowed_entries` option. The default value for this option is `1`, which means that if the disk is ever sent back to the maintenance center, it is automatically failed.

You can also put a disk into the maintenance center manually by using the `disk maint start` command. If the target disk is in use, it does not enter the maintenance center until its contents have been copied to another disk (unless you include the -i option).

Data ONTAP informs you of these activities by sending messages to the following destinations:

- The console
- A log file at `/etc/maintenance.log`

  Note: When Data ONTAP puts a drive into the maintenance center, and that drive is housed in a disk shelf that supports automatic power cycling, power to that drive might be turned off for a short period of time. If the drive returns to a ready state after the power cycle, the maintenance center tests the drive. Otherwise, the maintenance center fails the drive immediately.

You can see the power-cycle status for ESH4 disk shelves by using the `environment shelf_power_status` command.

For information about best practices for working with the maintenance center, see *Technical Report 3437: Storage Best Practices and Resiliency Guide*.

Related information

TR 3437: Storage Best Practices and Resiliency Guide

When Data ONTAP can put a disk into the maintenance center

When Data ONTAP detects certain disk errors, it tries to put the disk into the maintenance center for testing. Certain requirements must be met for the disk to be put into the maintenance center.

If a disk experiences more errors than are allowed for that disk type, Data ONTAP takes one of the following actions:

- If the `disk.maint_center.spares_check` option is set to `on` (the default) and two or more spares are available, Data ONTAP takes the disk out of service and assigns it to the maintenance center for data management operations and further testing.
- If the `disk.maint_center.spares_check` option is set to `on` and fewer than two spares are available, Data ONTAP does not assign the disk to the maintenance center. It simply fails the disk and designates the disk as a broken disk.
- If the `disk.maint_center.spares_check` option is set to `off`, Data ONTAP assigns the disk to the maintenance center without checking the number of available spares.

Note: The `disk.maint_center.spares_check` option has no effect on putting disks into the maintenance center from the command-line interface.

Data ONTAP does not put SSDs into the maintenance center.

How Data ONTAP uses continuous media scrubbing to prevent media errors

The purpose of the continuous media scrub is to detect and correct media errors in order to minimize the chance of storage system disruption due to a media error while a storage system is in degraded or reconstruction mode.

By default, Data ONTAP runs continuous background media scrubbing for media errors on all storage system disks. If a media error is found, Data ONTAP uses RAID to reconstruct the data and repairs the error.

Media scrubbing is a continuous background process. Therefore, you might observe disk LEDs blinking on an apparently idle storage system. You might also observe some CPU activity even when no user workload is present.

How continuous media scrubbing impacts system performance

Because continuous media scrubbing searches only for media errors, its impact on system performance is negligible. In addition, the media scrub attempts to exploit idle disk bandwidth and free CPU cycles to make faster progress. However, any client workload results in aggressive throttling of the media scrub resource.

If needed, you can further decrease the CPU resources consumed by a continuous media scrub under a heavy client workload by increasing the maximum time allowed for a media scrub cycle to complete. You can do this by using the `raid.media_scrub.rate` option.

Why continuous media scrubbing should not replace scheduled RAID-level disk scrubs

Because the continuous media scrub process scrubs only media errors, you should continue to run the storage system's scheduled complete RAID-level scrub operation. The RAID-level scrub finds and corrects parity and checksum errors as well as media errors.

Related concepts

How you schedule automatic RAID-level scrubs on page 117

Increasing storage availability by using ACP

ACP, or Alternate Control Path, is a protocol that enables Data ONTAP to manage and control a SAS disk shelf storage subsystem. It uses a separate network (alternate path) from the data path, so management communication is not dependent on the data path being intact and available.

You do not need to actively manage the SAS disk shelf storage subsystem. Data ONTAP automatically monitors and manages the subsystem without operator intervention. However, you must provide the required physical connectivity and configuration parameters to enable the ACP functionality.

Note: You can install SAS disk shelves without configuring ACP. However, for maximum storage availability and stability, you should always have ACP configured and enabled.

After you enable ACP, you can use the `storage show acp` and `acpadmin list_all` commands to display information about your ACP subsystem.

Because ACP communication is on a separate network, it does not affect data access in any way.

Enabling ACP

ACP can increase your storage availability when you use SAS disk shelves. If your storage system model has a dedicated port for ACP, then ACP is enabled by default and you do not need to explicitly enable ACP.

Before you begin

- The ACP subnet must be cabled on an isolated network, with no switches or hubs.
  For more information, see the *Installation and Service Guide* for your disk shelf.
- You must have identified a port that is not in use by any other subsystem.
- If you are configuring ACP for disk shelves attached to an HA pair, you must have recorded the domain name and network mask to ensure that they are the same for both nodes.

About this task

The ACP subnet is a private Ethernet network that enables the ACP processor in the SAS module to communicate both with Data ONTAP and the SAS IOMs in the disk shelves.

The ACP subnet is separate from the I/O data path that connects the disk shelves to the HBA on the storage controller. When you configure ACP on one of the system's network interfaces, you must supply a private domain name that conforms to the standard for private internet addresses (RFC1918). You can use the system default domain or another network name (that is, an IP address ending in 0) that conforms to the standard.

Steps

1. If your system does not have a dedicated port for ACP (e0P), ensure that the port you are assigning to ACP is not in use by any other subsystem by reviewing your `/etc/rc` file and entering the following command:

   `ifconfig` *`interface_name`*

   The interface you use for ACP should not be part of an Interface Group, and it should have no VLANs or IP addresses configured on it.

2. At the Data ONTAP command line, enter the following command:

   `acpadmin configure`

   If you have not previously configured the networking information for ACP, you are prompted for that information. When you select a domain name and network mask for the ACP interface, Data ONTAP automatically assigns IP addresses for the ACP interface on the storage controller and both I/O modules on each disk shelf on the ACP subnet.

3. If you configured ACP to use a different port, reboot the node.

   The previous port becomes available for use by another subsystem.

4. Verify your ACP connectivity by entering the following command:

`storage show acp`

The ACP Connectivity Status should show “Full Connectivity”.

Example

For example, if you select e0P as the interface for ACP traffic, 192.168.0.0 as the ACP domain, and 255.255.252.0 as the network mask for the ACP subnet, the `storage show acp` command output looks similar to the following:

```
my-sys-1> storage show acp

 Alternate Control Path:            Enabled
 Ethernet Interface:                e0p
 ACP Status:                        Active
 ACP IP address:                    192.168.2.61
 ACP domain:                        192.168.0.0
 ACP netmask:                       255.255.252.0
 ACP Connectivity Status:           Full Connectivity
 ACP Partner Connectivity Status:   NA

Shelf Module      Reset Cnt    IP address      FW Version    Module Type   Status
-------------------------------------------------------------------------------
7a.001.A          002          192.168.0.145    01.05         IOM6          active
7a.001.B          003          192.168.0.146    01.05         IOM6          active
7c.002.A          000          192.168.0.206    01.05         IOM6          active
7c.002.B          001          192.168.0.204    01.05         IOM6          active
```

How you use SSDs to increase storage performance

Solid-state disks (SSDs) are flash memory-based storage devices that provide better overall performance than hard disk drives (HDDs), which are mechanical devices using rotating media. You should understand how Data ONTAP manages SSDs and the capability differences between SSDs and HDDs.

Depending on your storage system model, you can use SSDs in two ways:

- You can create Flash Pool aggregates--aggregates composed mostly of HDDs, but with some SSDs that function as a high-performance cache for your working data set.
- You can create aggregates composed entirely of SSDs, where the SSDs function as the persistent storage for all data in the aggregate.

You manage Flash Pool aggregates and aggregates composed entirely of SSDs the same way you manage aggregates composed entirely of HDDs. However, there are some differences in the way you manage SSDs from the way you manage disks. In addition, some Data ONTAP capabilities are not available on SSDs and Flash Pool aggregates.

SSDs are not supported on all storage system models. For information about which models support SSDs, see the *Hardware Universe* at *hwu.netapp.com*.

Related concepts

How Flash Pool aggregates work on page 130
Restrictions for using aggregates composed of SSDs on page 133

How Data ONTAP manages SSD wear life

Solid-state disks (SSDs) have a different end-of-life behavior than rotating media (hard disk drives, or HDDs). Data ONTAP monitors and manages SSDs to maximize storage performance and availability.

In the absence of a mechanical failure, rotating media can serve data almost indefinitely. This is not true for SSDs, which can accept only a finite (though very large) number of write operations. SSDs provide a set of internal spare capacity, called *spare blocks*, that can be used to replace blocks that have reached their write operation limit. After all of the spare blocks have been used, the next block that reaches its limit causes the disk to fail.

Because a drive failure is an undesirable occurrence, Data ONTAP replaces SSDs before they reach their limit. When a predetermined percentage of the spare blocks have been used (approximately 90%), Data ONTAP performs the following actions:

1. Sends an AutoSupport message.
2. If a spare SSD is available, starts a disk copy to that spare.
3. If no spare is available, starts a periodic check for a spare so that the disk copy can be started when a spare becomes available.
4. When the disk copy finishes, fails the disk.

Note: You do not need to replace SSDs before they are failed by Data ONTAP. However, when you use SSDs in your storage system (as for all disk types), it is important to ensure that you have sufficient hot spares available at all times.

Capability differences between SSDs and HDDs

Usually, you manage SSDs the same as HDDs, including firmware updates, scrubs, and zeroing. However, some Data ONTAP capabilities do not make sense for SSDs, and SSDs are not supported on all hardware models.

SSDs cannot be combined with HDDs within the same RAID group. When you replace an SSD in an aggregate, you must replace it with another SSD. Similarly, when you physically replace an SSD within a shelf, you must replace it with another SSD.

The following capabilities of Data ONTAP are not available for SSDs:

- Disk sanitization is not supported for all SSD part numbers.
  For information about which SSD part numbers support sanitization, see the *Hardware Universe*.
- The maintenance center
- FlexShare

SSDs are not supported on all storage system models. For information about which models support SSDs, see the *Hardware Universe* at *hwu.netapp.com*.

Guidelines and requirements for using multi-disk carrier disk shelves

Data ONTAP automatically handles most of the extra steps required to manage disks in multi-disk carriers. However, there are some extra management and configuration requirements that you must understand before incorporating multi-disk carrier disk shelves in your storage architecture.

When using storage from multi-disk carrier disk shelves such as the DS4486, you must familiarize yourself with the guidelines and requirements governing the following topics:

- The process that Data ONTAP uses to avoid impacting any RAID groups when a multi-disk carrier needs to be removed
- When it is safe to remove a multi-disk carrier after a disk failure
- The minimum required number of spares for multi-disk carrier disks
- Multi-disk carrier disk shelf configuration
- Aggregate configuration requirements when using multi-disk carrier disk shelves
- Guidelines and best practices for using disks from a multi-disk carrier disk shelf in an aggregate

How Data ONTAP avoids RAID impact when a multi-disk carrier must be removed

Data ONTAP takes extra steps to ensure that both disks in a carrier can be replaced without impacting any RAID group. Understanding this process helps you know what to expect when a disk from a multi-disk carrier disk shelf fails.

A multi-disk carrier disk shelf, such as the DS4486, has double the disk density of other SAS disk shelves. It accomplishes this by housing two disks per disk carrier. When two disks share the same disk carrier, they must be removed and inserted together. This means that when one of the disks in a carrier needs to be replaced, the other disk in the carrier must also be replaced, even if it was not experiencing any issues.

Removing two data or parity disks from an aggregate at the same time is undesirable, because it could leave two RAID groups degraded, or one RAID group double-degraded. To avoid this situation, Data ONTAP initiates a disk evacuation operation for the carrier mate of the failed disk, as well as the usual reconstruction to replace the failed disk. The disk evacuation operation copies the contents of the carrier mate to a disk in a different carrier so the data on that disk remains available when you remove the carrier. During the evacuation operation, the status for the disk being evacuated shows as `evacuating`.

In addition, Data ONTAP tries to create an optimal layout that avoids having two carrier mates in the same RAID group. Depending on how the other disks are laid out, achieving the optimal layout can require as many as three consecutive disk evacuation operations. Depending on the size of the disks

and the storage system load, each disk evacuation operation could take several hours, so the entire swapping process could take an entire day or more.

If insufficient spares are available to support the swapping operation, Data ONTAP issues a warning and waits to perform the swap until you provide enough spares.

How to determine when it is safe to remove a multi-disk carrier

Removing a multi-disk carrier before it is safe to do so can result in one or more RAID groups becoming degraded, or possibly even a storage disruption. Data ONTAP provides several indications of when it is safe to remove a multi-disk carrier.

When a multi-disk carrier needs to be replaced, the following events must have occurred before you can remove the carrier safely:

- An AutoSupport message must have been logged indicating that the carrier is ready to be removed.
- An EMS message must have been logged indicating that the carrier is ready to be removed.
- Both disks in the carrier must be displayed in the list of broken disks.

  You can see the list of broken disks by using the `aggr status -f` command.

  The disk that was evacuated to allow the carrier to be removed shows the outage reason of `evacuated`.
- The amber LED on the carrier must be lit continuously.
- The green LED on the carrier must show no activity.

Attention: You cannot reuse the carrier mate of a failed disk. When you remove a multi-disk carrier that contains a failed disk, you must replace and return the entire carrier.

Spare requirements for multi-disk carrier disks

Maintaining the proper number of spares for disks in multi-disk carriers is critical for optimizing storage redundancy and minimizing the amount of time Data ONTAP must spend copying disks to achieve an optimal disk layout.

You must maintain a minimum of two hot spares for multi-disk carrier disks at all times. To support the use of the Maintenance Center, and to avoid issues caused by multiple concurrent disk failures, you should maintain at least four hot spares for steady state operation, and replace failed disks promptly.

If two disks fail at the same time with only two available hot spares, Data ONTAP might not be able to swap the contents of both the failed disk and its carrier mate to the spare disks. This scenario is called a *stalemate*. If this happens, you are notified through EMS messages and AutoSupport messages. When the replacement carriers become available, you must follow the instructions provided by the EMS messages or contact technical support to recover from the stalemate.

Shelf configuration requirements for multi-disk carrier disk shelves

You can combine multi-disk carrier disk shelves with single-disk carrier disk shelves (standard disk shelves) on the same storage system. However, you cannot combine the two disk shelf types in the same stack.

Aggregate requirements for disks from multi-disk carrier disk shelves

Aggregates composed of disks from multi-disk carrier disk shelves must conform to some configuration requirements.

The following configuration requirements apply to aggregates composed of disks from multi-disk carrier disk shelves:

- The RAID type must be RAID-DP.
- The format must be 64-bit.
- All HDDs in the aggregate must be the same Data ONTAP disk type.

  The aggregate can be a Flash Pool aggregate.
- If the aggregate is mirrored, both plexes must have the same Data ONTAP disk type (or types, in the case of a Flash Pool aggregate).
- The aggregate cannot be a traditional volume.

Related concepts

How Flash Pool aggregates work on page 130

Considerations for using disks from a multi-disk carrier disk shelf in an aggregate

Observing the requirements and best practices for using disks from a multi-disk carrier disk shelf in an aggregate enables you to maximize storage redundancy and minimize the impact of disk failures.

Disks in multi-disk carriers always have the Data ONTAP disk type of MSATA. MSATA disks cannot be mixed with HDDs from a single-carrier disk shelf in the same aggregate.

The following disk layout requirements apply when you are creating or increasing the size of an aggregate composed of MSATA disks:

- Data ONTAP prevents you from putting two disks in the same carrier into the same RAID group.
- Do not put two disks in the same carrier into different pools, even if the shelf is supplying disks to both pools.
- Do not assign disks in the same carrier to different nodes.
- For the best layout, do not name specific disks; allow Data ONTAP to select the disks to be used or added.

  If the operation cannot result in an optimal layout due to the placement of the disks and available spares, Data ONTAP automatically swaps disk contents until an optimal layout is achieved. If there are not enough available spares to support the swaps, Data ONTAP issues a warning and

waits to perform the disk swaps until you provide the necessary number of hot spares. If you name disks and an optimal layout cannot be achieved, you must explicitly force the operation; otherwise the operation fails.

Aggregate creation example

To create an aggregate using MSATA disks, you can specify the disk type and size but leave the disk selection and layout to Data ONTAP by using a command like this:

```
aggr create n1_aggr1 -T MSATA 14
```

Adding disks to a storage system

You add disks to a storage system to increase the number of hot spares, to add space to an aggregate, or to replace disks.

Before you begin

You must have confirmed that your storage system supports the type of disk you want to add. For information about supported disk drives, see the *Hardware Universe* at *hwu.netapp.com*.

About this task

You use this procedure to add physical disks to your storage system. If you are administering a storage system that uses virtual disks, for example, a system based on Data ONTAP-v technology, see the installation and administration guide that came with your Data ONTAP-v system for information about adding virtual disks.

Steps

1. Check the NetApp Support Site for newer disk and shelf firmware and Disk Qualification Package files. If your system does not have the latest versions, update them before installing the new disk.
2. Install one or more disks according to the hardware guide for your disk shelf or the hardware and service guide for your storage system.

   The new disks are not recognized until they are assigned to a system and pool. You can assign the new disks manually, or you can wait for Data ONTAP to automatically assign the new disks if your system follows the rules for disk autoassignment.
3. After the new disks have all been recognized, verify their addition and their ownership information by entering the following command:

   `disk show -v`

You should see the new disks, owned by the correct system and in the correct pool, listed as hot spare disks.

4. You can zero the newly added disks now, if needed, by entering the following command:

   `disk zero spares`

   Note: Disks that have been used previously in a Data ONTAP aggregate must be zeroed before they can be added to another aggregate. Zeroing the disks now can prevent delays in case you need to quickly increase the size of an aggregate. The disk zeroing command runs in the background and can take hours to complete, depending on the size of the non-zeroed disks in the system.

Result

The new disks are ready to be added to an aggregate, used to replace an existing disk, or placed onto the list of hot spares.

Related concepts

Guidelines for assigning ownership for disks on page 63
How automatic ownership assignment works for disks on page 59
Managing disks using Data ONTAP on page 18

Related information

Disk Qualification Package Instructions: support.netapp.com/NOW/download/tools/diskqual/
Disk Drive & Firmware Matrix: support.netapp.com/NOW/download/tools/diskfw/

When you need to update the Disk Qualification Package

The Disk Qualification Package (DQP) adds full support for newly qualified disk drives. Each time you update disk firmware or add new disk types or sizes to the storage system, you also need to update the DQP.

You can obtain the DQP from the NetApp Support Site. You need to download and install the DQP in the following situations:

- Whenever you add a new disk type or size to the node
  For example, if you already have 1-TB disks and add 2-TB disks, you need to check for the latest DQP update.
- Whenever you update the disk firmware
- Whenever newer disk firmware or DQP files are available

Related information

Disk Qualification Package Instructions: support.netapp.com/NOW/download/tools/diskqual/
Disk Drive & Firmware Matrix: support.netapp.com/NOW/download/tools/diskfw/

Replacing disks that are currently being used in an aggregate

You can use the `disk replace` command to replace disks that are part of an aggregate without disrupting data service. You do this to swap out mismatched disks from a RAID group. Keeping your RAID groups homogeneous helps optimize storage system performance.

Before you begin

You should already have an appropriate hot spare disk of the correct type, size, speed, and checksum type installed in your storage system. This spare disk must be assigned to the same system and pool as the disk it will replace. For multi-disk carrier disks, you should have at least two hot spare disks available, to enable Data ONTAP to provide an optimal disk layout.

About this task

If you need to replace a disk—for example a mismatched data disk in a RAID group—you can replace the disk. This operation uses Rapid RAID Recovery to copy data from the specified old disk in a RAID group to the specified spare disk in the storage system. At the end of the process, the spare disk replaces the old disk as the new data disk, and the old disk becomes a spare disk in the storage system.

Note: If you replace a smaller disk with a larger disk, the capacity of the larger disk is downsized to match that of the smaller disk; the usable capacity of the aggregate is not increased.

Step

1. Enter the following command:

   `disk replace start [-m]` *`old_disk_name new_spare_disk_name`*

   If you need to use a disk that does not match the speed or pool of the other disks in the aggregate, you can use the `-m` option.

   If you need to stop the disk replace operation, you can use the `disk replace stop` command. If you halt a disk replace operation, the target spare disk needs to be zeroed before it can be used in another aggregate.

Result

The old disk is converted to a spare disk, and the new disk is now used in the aggregate.

Related concepts

How Data ONTAP works with hot spare disks on page 113
Managing disks using Data ONTAP on page 18

Guidelines for assigning ownership for disks on page 63
How automatic ownership assignment works for disks on page 59

Related tasks

Adding disks to a storage system on page 39
Assigning ownership for unowned disks and array LUNs on page 64

Replacing a self-encrypting disk

Replacing a self-encrypting disk (SED) is similar to replacing a regular disk, except that there are some extra steps you must take to reenable Storage Encryption after you replace the disk.

Before you begin

You should know the key used by the SEDs on your storage system so that you can configure the replacement SED to use the same key.

Steps

1. Ensure that reconstruction has started by entering the following command:

   `aggr status -r`

   The status of the disk should display as "Reconstructing".
2. Remove the failed disk and replace it with a new SED, following the instructions in the hardware guide for your disk shelf model.
3. Assign ownership of the newly replaced SED by entering the following command:

   `disk assign disk_name`
4. Confirm that the new disk has been properly assigned by entering the following command:

   `disk encrypt show`

   You should see the newly added disk in the output.
5. Encrypt the disk by entering the following command:

   `disk encrypt rekey key_id disk_name`
6. Finalize the replacement process by entering the following command:

   `disk encrypt lock disk_name`

   The newly replaced SED is ready for use, and Storage Encryption is enabled and working on this system.

Converting a data disk to a hot spare

Data disks can be converted to hot spares by destroying the aggregate that contains them.

Before you begin

The aggregate to be destroyed cannot contain volumes.

About this task

Converting a data disk to a hot spare does not change the ownership information for that disk. You must remove ownership information from a disk before moving it to another storage system.

Step

1. Destroy the aggregate that contains the disk by entering the following command:

   aggr destroy *aggr_name*

   All disks in use by that aggregate are converted to hot spare disks.

Removing disks from a storage system

How you remove a disk from your storage system depends how the disk is being used. By using the correct procedure, you can prevent unwanted AutoSupport notifications from being generated and ensure that the disk will function correctly if it is reused in another storage system.

You cannot reduce the number of disks in an aggregate by removing data disks. The only way to reduce the number of data disks in an aggregate is to copy the data and transfer it to a new aggregate that has fewer data disks.

Removing a failed disk

A disk that is completely failed is no longer counted by Data ONTAP as a usable disk, and you can immediately disconnect the disk from the disk shelf. However, you should leave a partially failed disk connected long enough for the Rapid RAID Recovery process to complete.

About this task

If you are removing a disk because it has failed or because it is producing excessive error messages, you should not use the disk again in this or any other storage system.

Steps

1. Find the disk ID of the failed disk by entering the following command:

   aggr status -f

If the disk does not appear in the list of failed disks, it might be partially failed, with a Rapid RAID Recovery in process. In this case, you should wait until the disk is present in the list of failed disks (which means that the Rapid RAID Recovery process is complete) before removing the disk.

2. Determine the physical location of the disk you want to remove from the output of the `aggr status -f` command. The location is shown in the columns labeled HA, SHELF, and BAY.
3. Remove the disk from the disk shelf, following the instructions in the hardware guide for your disk shelf model.

Related concepts

How to determine when it is safe to remove a multi-disk carrier on page 37

Removing a hot spare disk

Removing a hot spare disk requires you to remove ownership information from the disk. This prevents the disk from causing problems when it is inserted into another storage system, and notifies Data ONTAP that you are removing the disk to avoid unwanted AutoSupport messages.

About this task

Removing a hot spare disk does not make the contents of that disk inaccessible. If you need absolute assurance that the data contained by this disk is irretrievable, you should sanitize the disk instead of completing this procedure.

Steps

1. Find the disk name of the hot spare disk you want to remove:

   `aggr status -s`

   The names of the hot spare disks appear next to the word `spare`. The locations of the disks are shown to the right of the disk name.
2. Enter the advanced privilege level:

   `priv set advanced`
3. Determine the physical location of the disk you want to remove:

   `led_on` ***`disk_name`***

   The fault LED on the face of the disk is lit.
4. If disk ownership automatic assignment is on, turn it off:

   `options disk.auto_assign off`
5. Repeat the previous step on the node's HA partner if it has one.
6. Remove the software ownership information from the disk:

   `disk remove_ownership` ***`disk_name`***

7. Return to admin privilege level:

   ```
   priv set
   ```

8. Remove the disk from the disk shelf, following the instructions in the hardware guide for your disk shelf model.

9. If you turned off disk ownership automatic assignment previously, turn it on now:

   ```
   options disk.auto_assign on
   ```

Related concepts

How to determine when it is safe to remove a multi-disk carrier on page 37

Removing a data disk

The only time that you should remove a data disk from a storage system is if the disk is not functioning correctly. If you want to remove a data disk so that it can be used in another system, you must convert it to a hot spare disk first.

About this task

You can cause Data ONTAP to fail the disk immediately or allow a disk copy to finish before the disk is failed. If you do not fail the disk immediately, you must wait for the disk copy to finish before physically removing the disk. This operation might take several hours, depending on the size of the disk and the load on the storage system.

Do not immediately fail a disk unless it is causing immediate performance or availability issues for your storage system. Depending on your storage system configuration, additional disk failures could result in data loss.

Steps

1. Determine the name of the disk you want to remove.

   If the disk is reporting errors, you can find the disk name in the log messages that report disk errors. The name is prefixed with the word "Disk".

2. Determine the location of the disk you want to remove by entering the following command:

   ```
   aggr status -r
   ```

   The location of the disk appears to the right of its name, in the columns HA, SHELF, and BAY.

3. Take the appropriate action based on whether you need to fail the disk immediately or not.

If you...	Then...
Can wait for the copy operation to finish (recommended)	Enter the following command to pre-fail the disk: **`disk fail disk_name`** Data ONTAP pre-fails the specified disk and attempts to create a replacement disk by copying the contents of the pre-failed disk to a spare disk. If the copy operation is successful, then Data ONTAP fails the disk and the new replacement disk takes its place. If the copy operation fails, the pre-failed disk fails and the storage system operates in degraded mode until the RAID system reconstructs a replacement disk.
Need to remove the disk immediately	Enter the following command to cause the disk to fail immediately: **`disk fail -i disk_name`** The disk fails and the storage system operates in degraded mode until the RAID system reconstructs a replacement disk.

4. Remove the failed disk from the disk shelf, following the instructions in the hardware guide for your disk shelf model.

Related concepts

About degraded mode on page 115
Managing disks using Data ONTAP on page 18
How to determine when it is safe to remove a multi-disk carrier on page 37

Using disk sanitization to remove data from disks

Disk sanitization enables you to remove data from a disk or set of disks so that the data can never be recovered.

Before you begin

The disks that you want to sanitize must be spare disks; they must be owned but not used in an aggregate.

About this task

When disk sanitization is enabled on a storage system, it cannot be disabled again.

If you need to remove data from disks using Storage Encryption, do not use this procedure. Use the procedure for destroying data on disks using Storage Encryption.

Steps

1. Enable disk sanitization by entering the following command:

```
options licensed_feature.disk_sanitization.enable on
```

You are asked to confirm the command, because it is irreversible.

2. Sanitize the specified disks by entering the following command:

   `disk sanitize start [-p pattern1|-r [-p pattern2|-r [-p pattern3|-r]]] [-c cycle_count] disk_list`

   Attention: Do not turn off the storage system, disrupt the storage connectivity, or remove target disks while sanitizing. If sanitizing is interrupted during the formatting phase, the formatting phase must be restarted and allowed to finish before the disks are sanitized and ready to be returned to the spare pool.

   If you need to abort the sanitization process, you can do so by using the `disk sanitize abort` command. If the specified disks are undergoing the formatting phase of sanitization, the abort does not occur until the phase is complete. At that time, Data ONTAP displays a message telling you that the sanitization process was stopped.

   `-p` *`pattern1`* `-p` *`pattern2`* `-p` *`pattern3`* specifies a cycle of one to three user-defined hex byte overwrite patterns that can be applied in succession to the disks being sanitized. The default pattern is three passes, using 0x55 for the first pass, 0xaa for the second pass, and 0x3c for the third pass.

   `-r` replaces a patterned overwrite with a random overwrite for any or all of the passes.

   `-c` *`cycle_count`* specifies the number of times that the specified overwrite patterns are applied. The default value is one cycle. The maximum value is seven cycles.

   `disk_list` specifies a space-separated list of the IDs of the spare disks to be sanitized.

3. If you want to check the status of the disk sanitization process, enter the following command:

   `disk sanitize status [disk_list]`

4. After the sanitization process is complete, return the disks to spare status by entering the following command for each disk:

   `disk sanitize release disk_name`

5. Determine whether all of the disks were returned to spare status by entering the following command:

   `aggr status -s`

If...	Then...
All of the sanitized disks are listed as spares	You are done. The disks are sanitized and in spare status.

If...	Then...

Some of the sanitized disks are not listed as spares	Complete the following steps:
	a. Enter advanced privilege mode:
	`priv set advanced`
	b. Assign the disks to the appropriate storage system by entering the following command for each disk:
	`disk assign disk_name -o system_name`
	c. Return the disks to spare status by entering the following command for each disk:
	`disk unfail -s disk_name`
	d. Return to administrative mode:
	`priv set`

Result

The specified disks are sanitized and designated as hot spares. The serial numbers of the sanitized disks are written to `/etc/log/sanitized_disks`.

Examples

The following command applies the default three disk sanitization overwrite patterns for one cycle (for a total of three overwrites) to the specified disks 8a.6, 8a.7, and 8a.8:

```
disk sanitize start 8a.6 8a.7 8a.8
```

The following command would result in three disk sanitization overwrite patterns for six cycles (for a total of 18 overwrites) to the specified disks:

```
disk sanitize start -c 6 8a.6 8a.7 8a.8
```

Related concepts

How disk sanitization works on page 26
When disk sanitization cannot be performed on page 27
Managing disks using Data ONTAP on page 18

Related tasks

Destroying data on disks using Storage Encryption on page 101

Removing data from disks using selective disk sanitization

The procedure you use to selectively sanitize data depends on whether your data is contained in FlexVol or traditional volumes.

Related concepts

How selective disk sanitization works on page 28
Managing disks using Data ONTAP on page 18

Selectively sanitizing data contained in FlexVol volumes

To selectively sanitize data contained in FlexVol volumes, you need to migrate any data you want to preserve in the *entire aggregate*, because every disk used by that aggregate must be sanitized.

Before you begin

- You must have the `licensed_feature.disk_sanitization.enable` option set to `On`.

  Attention: After disk sanitization is enabled on a storage system, it is permanent, and it prevents certain Data ONTAP commands from being run.
- You need enough free space to duplicate the data you want to preserve, plus extra space for overhead.

  If you have a limited amount of free space, you can decrease the size of the FlexVol volumes after you delete the data you do not want to preserve and before migrating the volume.

Steps

1. Stop any applications that write to the aggregate you plan to sanitize.
2. From a Windows or UNIX client, delete the directories or files whose data you want to selectively sanitize from the active file system.

   Use the appropriate Windows or UNIX command, for example:

   `rm /nixdir/nixfile.doc`
3. Remove NFS and CIFS access to all volumes in the aggregate.
4. From the Data ONTAP command line, enter the following command to delete all volume Snapshot copies of the FlexVol volumes that contained the files and directories you just deleted:

   `snap delete -V -a` *`vol_name`*

   `vol_name` is the FlexVol volume that contains the files or directories that you just deleted.
5. Note the names of the volumes that contain data you want to preserve.
6. Enter the following command for each volume you want to preserve, noting the total size and space used:

   `df -g` *`vol_name`*

7. If you do not have sufficient free space to create an aggregate to contain the migrated volumes at their current size, and the volumes have free space, enter the following command for each volume to decrease its size:

   vol size *vol_name new_size*

   Note: The new size must be larger than the used space in the volume.

8. Create an aggregate to which you will migrate the data you did not delete by entering the following command:

   aggr create *dest_vol disks*

   Example

   aggr create nixdestaggr 8@72G

   This new aggregate provides a migration destination that is free of the data that you want to sanitize.

9. For each FlexVol volume that contains data you want to preserve, enter the following command to create a corresponding FlexVol volume in the new aggregate:

   vol create *dest_vol dest_aggrsize*

   dest_vol is the name of the new FlexVol volume. Use a different name for the new FlexVol volume.

   dest_aggr is the aggregate you just created.

   size must be at least as large as the current size of the FlexVol volume in the aggregate you will sanitize.

   Example

   To create a FlexVol volume to preserve the data in the nixsrcvol volume, which is a little more than 19 GB, you could use the following command:

   vol create nixsrcvol_1 nixdestaggr 20G

   You now have the volumes into which you will copy the data you want to preserve.

10. For each FlexVol volume that contains data you want to preserve, enter the following command to copy the data to the new aggregate:

    ndmpcopy /vol/*src_vol* /vol/*dest_vol*

    src_vol is the FlexVol volume in the aggregate you want to sanitize.

    dest_vol is the new FlexVol volume that you just created that corresponded to the *src_vol* volume.

    Example

    ndmpcopy /vol/nixsrcvol /vol/nixsrcvol_1

For information about the `ndmpcopy` command, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.

All of the data you want to preserve is now contained in the new aggregate.

11. List the disk IDs used by the source aggregate by entering the following command:

`aggr status src_aggr -r`

Example

`aggr status nixsrcaggr -r`

The disks that you will sanitize are listed in the `Device` column of the `aggr status -r` output.

12. Record the disk IDs you listed in the previous step.

13. For each FlexVol volume in the aggregate you are sanitizing, enter the following commands to take the volume offline and destroy it:

`vol offline src_vol`

`vol destroy src_vol`

14. Enter the following commands to take the source aggregate offline and destroy it:

`aggr offline src_aggr`

`aggr destroy src_aggr`

The volumes and aggregate that housed the data you want to sanitize have been destroyed. The disks used in this aggregate are now hot spares.

15. Enter the following command to rename the new aggregate, giving it the name of the aggregate that you just destroyed:

`aggr rename dest_aggr old_src_aggr_name`

Example

`aggr rename nixdestaggr nixsrcaggr`

16. For each FlexVol volume in the new aggregate, enter the following command to rename the FlexVol volume to the name of the original FlexVol volume:

`vol rename dest_vol old_src_vol_name`

Example

`vol rename nixsrcvol_1 nixsrcvol`

17. Reestablish your CIFS or NFS services.

- If the original volume supported CIFS services, restart the CIFS services on the volumes in the destination aggregate after migration is complete.

- If the original volume supported NFS services, enter the following command:

  exportfs -a

Users who were accessing files in the original volume continue to access those files in the renamed destination volume with no remapping of their connections required.

18. Follow the procedure for sanitizing disks on the disks that belonged to the source aggregate.

Related tasks

Using disk sanitization to remove data from disks on page 46

Selectively sanitizing data contained in traditional volumes

To selectively sanitize data contained in traditional volumes, you migrate any data you want to preserve to a new volume, and then sanitize the disks that contained the old volume.

Before you begin

- You must have set the `licensed_feature.disk_sanitization.enable` option to `On`.

  Attention: After disk sanitization is enabled on a storage system, it is permanent, and it prevents certain Data ONTAP commands from being run.

- Your system must have enough free space to duplicate the entire traditional volume you are performing the selective sanitization on, regardless of how much data you are deleting before migrating the data.

Steps

1. Stop any applications that write to the volume you plan to sanitize.
2. From a Windows or UNIX client, delete the directories or files whose data you want to selectively sanitize from the active file system.

   Use the appropriate Windows or UNIX command, such as **rm /nixdir/nixfile.doc**.
3. Remove NFS and CIFS access to the volume you plan to sanitize.
4. Create a traditional volume to which you will migrate the data you did not delete by entering the following command:

   aggr create *dest_vol* -v *disks*

   Note: This traditional volume must have a storage capacity equal to or greater than the volume from which you are migrating. It must have a different name; later, you will rename it to have the same name as the volume you are sanitizing.

Example

aggr create nixdestvol -v 8@72G

This new volume provides a migration destination that is free of the data that you want to sanitize.

5. From the Data ONTAP command line, enter the following command to delete all volume Snapshot copies of the traditional volume that contained the files and directories you just deleted:

   snap delete -V -a *vol_name*

   vol_name is the traditional volume that contained the files or directories that you just deleted.

   Example

   snap delete -V -a nixdestvol

6. Confirm that you have deleted all files or directories that you want to sanitize from the source volume.

7. Copy the data you want to preserve to the destination volume from the volume you want to sanitize by entering the following command:

   ndmpcopy /vol/*src_vol* /vol/*dest_vol*

   src_vol is the volume you want to sanitize.

   dest_vol is the destination volume.

   For information about the ndmpcopy command, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.

   Example

   ndmpcopy /vol/nixsrcvol /vol/nixdestvol

8. List the disks used in the source volume by entering the following command:

   aggr status *src_vol* -r

   Example

   aggr status nixsrcvol -r

   The disks that you will sanitize are listed in the
   Device
   column of the aggr status -r output.

9. Record the IDs of the disks used in the source volume.

   After that volume is destroyed, you will sanitize these disks.

10. Take the volume you are sanitizing offline and destroy it by entering the following commands:

    aggr offline *src_vol*

    aggr destroy *src_vol*

Example

```
aggr offline nixsrcvol
aggr destroy nixsrcvol
```

11. Rename the new volume, giving it the name of the volume that you just destroyed, by entering the following command:

    ```
    aggr rename dest_vol old_src_vol_name
    ```

    Example

    ```
    aggr rename nixdestvol nixsrcvol
    ```

12. To confirm that the new volume is named correctly, list your volumes by entering the following command:

    ```
    aggr status old_src_vol_name
    ```

13. Reestablish your CIFS or NFS services.

    - If the original volume supported CIFS services, restart the CIFS services on the volumes in the destination aggregate after migration is complete.
    - If the original volume supported NFS services, enter the following command:

      ```
      exportfs -a
      ```

    Users who were accessing files in the original volume will continue to access those files in the renamed destination volume.

14. Follow the procedure for sanitizing disks to sanitize the disks that belonged to the source volume.

Result

After sanitizing, the data that you removed from the source volume no longer exists anywhere on your storage system and cannot be restored.

Related tasks

Using disk sanitization to remove data from disks on page 46

Stopping disk sanitization

You can use the `disk sanitize abort` command to stop an ongoing sanitization process on one or more specified disks.

Step

1. Enter the following command:

   ```
   disk sanitize abort disk_list
   ```

   If the specified disks are undergoing the disk formatting phase of sanitization, the process does not stop until the disk formatting is complete.

Data ONTAP displays the message `Sanitization abort initiated`. After the process stops, Data ONTAP displays another message for each disk to inform you that sanitization is no longer in progress.

How ownership for disks and array LUNs works

Disk and array LUN ownership determines which node owns a disk or array LUN and what pool a disk or array LUN is associated with. Understanding how ownership works enables you to maximize storage redundancy and manage your hot spares effectively.

Data ONTAP stores ownership information directly on the disk or array LUN.

Why you assign ownership of disks and array LUNs

Storage system ownership must be assigned for disks and array LUNs before they become an effective part of your system. You must explicitly assign ownership for array LUNs. Disks can be automatically or manually assigned.

You assign ownership of a disk or array LUN to accomplish the following actions:

- Associate the disk or array LUN with a specific storage system.
  For a stand-alone system, all disks and array LUNs are owned by that system. In an HA pair, the disks and array LUNs could be owned by either system.
- Enable the disk or array LUN to be used and managed by the system that owns it.
  Unowned disks cannot be used as spares and do not receive the automatic firmware updates that owned disks do.
- Associate the disk or array LUN with a specific SyncMirror pool (when SyncMirror is in use).
  If SyncMirror is not in use, all disks and array LUNs are in pool0.

What it means for Data ONTAP to own an array LUN

Data ONTAP cannot use an array LUN presented to it by a storage array until you configure a logical relationship in Data ONTAP that identifies a specific system running Data ONTAP as the *owner* of the array LUN.

A storage array administrator creates array LUNs and makes them available to specified FC initiator ports of storage systems running Data ONTAP. (The process for how to do this varies among storage array vendors.) When you assign an array LUN to a system running Data ONTAP, Data ONTAP writes data to the array LUN to identify that system as the *owner* of the array LUN. Thereafter, Data ONTAP ensures that only the owner can write data to and read data from the array LUN.

From the perspective of Data ONTAP, this logical relationship is referred to as *disk ownership* because Data ONTAP considers an array LUN to be a virtual disk. From the perspective of Data ONTAP, you are assigning disks to a storage system.

An advantage of the disk ownership scheme is that you can make changes through the Data ONTAP software that, on typical hosts, must be done by reconfiguring hardware or LUN access controls. For example, through Data ONTAP you can balance the load of requests among a group of systems

running Data ONTAP by moving data service from one system to another, and the process is transparent to most users. You do not need to reconfigure hardware or the LUN access controls on the storage array to change which system running Data ONTAP is the owner and, therefore, servicing data requests.

Attention: The Data ONTAP software-based scheme provides ownership control only for storage systems running Data ONTAP; it does not prevent a different type of host from overwriting data in an array LUN owned by a system running Data ONTAP. Therefore, if multiple hosts are accessing array LUNs through the same storage array port, be sure to use LUN security on your storage array to prevent the systems from overwriting each other's array LUNs.

Array LUN reconfiguration, such as resizing the array LUN, must be done from the storage array. Before such activities can occur, you must release Data ONTAP ownership of the array LUN.

Why you might assign array LUN ownership after installation

For a V-Series system ordered with disk shelves, you are not required to set up the system to work with array LUNs during initial installation. For a V-Series system using only array LUNs, you need to assign only two array LUNs during initial installation.

If you ordered your V-Series system with disk shelves, you do not need to assign any array LUNs initially because the factory installs the root volume on a disk for you. If you are using only array LUNs, you must configure one array LUN for the root volume and one array LUN as a spare for core dumps during initial installation. In either case, you can assign ownership of additional array LUNs to your system at any time after initial installation.

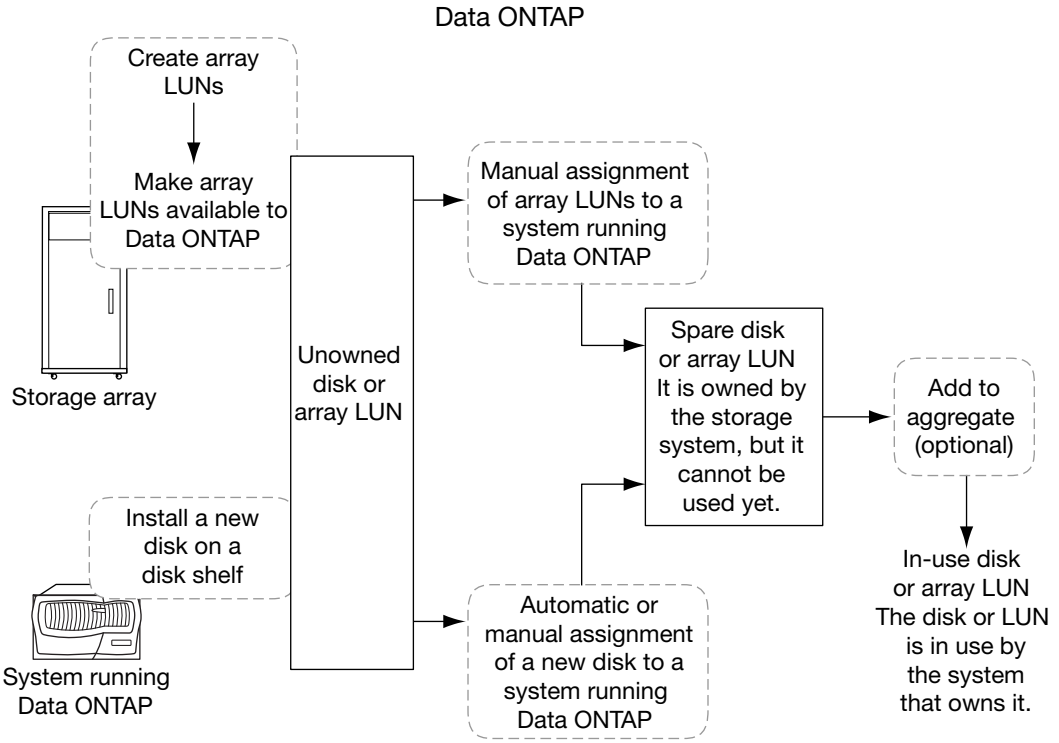
After initial configuration of your system, you might assign ownership of an array LUN in circumstances such as the following:

- You ordered your V-Series system with native disk shelves and you did not set up your system to work with array LUNs initially
- You left some LUNs that the storage array presented to Data ONTAP unowned and you now need to use the storage
- Another system released ownership of a particular array LUN and you want this system to be able to use the LUN
- The storage array administrator had not made the LUNs available to Data ONTAP when you initially configured your system and you now want to use the storage

How disks and array LUNs become available for use

When you add a disk or array LUN to a system running Data ONTAP, the disk or array LUN goes through several stages before it can be used by Data ONTAP to store data or parity information.

The process for making a disk available for use differs slightly from the process for making an array LUN available for use. Both processes are shown in the following diagram.



The process for disks includes the following actions:

1. The administrator physically installs the disk into a disk shelf.
   Data ONTAP can see the disk but the disk is still unowned.
2. If the system is configured to support disk autoassignment, Data ONTAP assigns ownership for the disk; otherwise, the administrator must assign ownership of the disk manually.
   The disk is now a spare disk.
3. The administrator or Data ONTAP adds the disk to an aggregate.
   The disk is now in use by that aggregate. It could contain data or parity information.

The process for array LUNs includes the following actions:

1. The storage array administrator creates the array LUN and makes it available to Data ONTAP.

Data ONTAP can see the array LUN but the array LUN is still unowned.

2. The Data ONTAP administrator assigns ownership for the array LUN to a V-Series system.
   The array LUN is now a spare array LUN.
3. The Data ONTAP administrator adds the array LUN to an aggregate.
   The array LUN is now in use by that aggregate and is used to contain data.

How automatic ownership assignment works for disks

If your configuration follows some basic rules to avoid ambiguity, Data ONTAP can automatically assign ownership and pool membership for disks. Automatic ownership assignment is not available for array LUNs or virtual disks.

If you decide to change the way Data ONTAP has assigned the disks, you can do so at any time.

If you need to temporarily remove disk ownership for a disk while you perform an administrative task, you must disable automatic disk ownership first to prevent Data ONTAP from immediately reassigning ownership for that disk.

What automatic ownership assignment does

When automatic disk ownership assignment runs, Data ONTAP looks for any unassigned disks and assigns them to the same system and pool as all other disks on their loop, stack, or shelf.

Note: If a single loop or stack has disks assigned to multiple systems or pools, Data ONTAP does not perform automatic ownership assignment on that loop or stack. Automatic assignment works only when it is clear which system or pool to assign unowned disks to. For this reason, always follow the disk assignment guidelines for your automatic assignment configuration.

You configure Data ONTAP to automatically assign disks at the stack or shelf level, depending on your system requirements and configuration. By default, autoassignment is at the stack or loop level. Data ONTAP automatically assigns the unowned disks to the system that owns the rest of the disks in that stack or loop.

If you need to have the disks in a single stack owned by more than one system, you can configure Data ONTAP to perform automatic disk assignment at the shelf level. In this case, Data ONTAP automatically assigns the unowned disks to the same owner as the already assigned disks on that shelf.

Related concepts

Guidelines for assigning ownership for disks on page 63
Managing disks using Data ONTAP on page 18

When automatic ownership assignment is invoked

Automatic disk ownership assignment does not happen immediately after disks are introduced into the storage system.

Automatic ownership assignment is invoked at the following times:

- Every five minutes during normal system operation
- Ten minutes after the initial system initialization
  This delay enables the person configuring the system enough time to finish the initial disk assignments so that the results of the automatic ownership assignment are correct.
- Whenever you enable automatic ownership assignment.

How disk ownership works for platforms based on Data ONTAP-v technology

You manage ownership for virtual disks by using the same commands you use for physical disks. However, automatic ownership assignment works differently for virtual disks.

Storage systems based on Data ONTAP-v technology automatically assign ownership for all virtual disks defined during the initial system setup. Automatic assignment is not available for any virtual disks you add after the initial system setup.

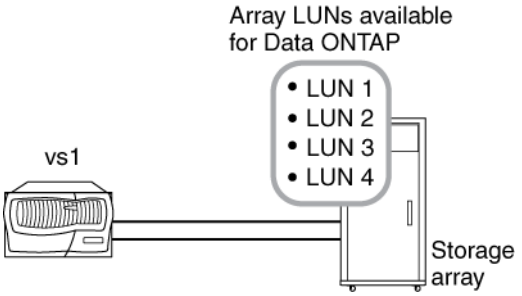
For information about adding virtual disks and managing a storage system based on Data ONTAP-v technology, see the installation and administration guide that came with your Data ONTAP-v system.

Examples showing when Data ONTAP can use array LUNs

After an array LUN has been assigned to a storage system, it can be added to an aggregate and used for storage or it can remain a spare LUN until it is needed for storage.

No storage system owns the LUNs yet

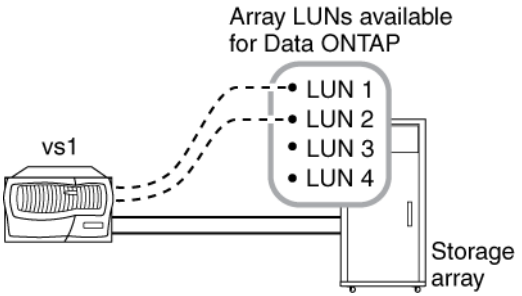
In this example, the storage array administrator made the array LUNs available to Data ONTAP. However, system vs1 has not yet been configured to "own" any of the LUNs. Therefore, it cannot read data from or write data to any array LUNs on the storage array.



Only some array LUNs are owned

In this example, vs1 was configured to own array LUNs 1 and 2, but not array LUNs 3 and 4. LUNs 3 and 4 are still available to Data ONTAP, however, and can be assigned to a storage system later.

Data ONTAP used the smallest of the two array LUNs, LUN 1, for the root volume. System vs1 can read data from and write data to LUN 1, because LUN 1 is in an aggregate. LUN 2 remains a spare LUN because it has not yet been added to an aggregate. System vs1 cannot read data from and write data to LUN 2 while it is a spare.

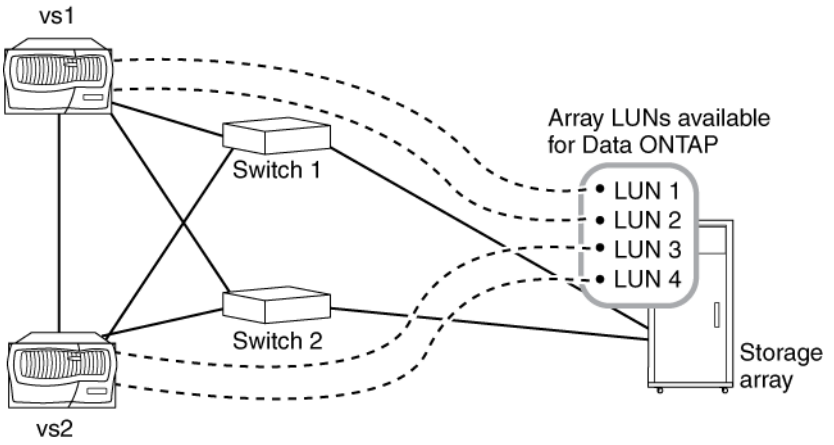


After you perform initial setup of the storage system, you could configure vs1 to also own LUN 3, LUN 4, both, or neither, depending on your storage needs.

Ownership of LUNs in an HA pair

In this example, two storage systems running Data ONTAP are configured in an HA pair. In an HA pair, only one node can be the owner of a particular LUN, but both nodes must be able to see the same LUNs so that the partner can take over if the owning node becomes unavailable.

LUN 1 through LUN 4 were created on the storage array and mapped to the ports on the storage array to which the storage systems are connected. All four LUNs are visible to each node in the HA pair.



Assume that during initial setup vs1 was assigned ownership of LUN 1 and LUN 2. LUN 1 was automatically added to the root volume, so LUN 1 is now "in use" by vs1. LUN 2 remains a spare until it is explicitly added to an aggregate on vs1. Similarly, assume that during initial setup vs2 was assigned ownership of LUN 3 and LUN 4, with LUN 3 assigned to the root volume. LUN 4 remains a spare LUN until it is explicitly added to an aggregate.

The key points of this example are as follows:

- By deploying the storage systems in an HA pair, one system can take over services for its partner if the partner becomes unavailable.
- Only one storage system can own a specific array LUN. However, all array LUNs assigned to a node in an HA pair must be visible to—but not assigned to or owned by—the other node in the HA pair.
- By deploying two switches, if one switch fails, the other switch provides the alternate path to the storage array.
- Both switches must be zoned correctly so that each storage system in the HA pair can see the array LUNs owned by its partner.

Managing ownership for disks and array LUNs

Disk ownership determines which node owns a disk and what pool a disk is associated with. Data ONTAP stores ownership information directly on the disk.

Guidelines for assigning ownership for disks

When you assign ownership for disks, you need to follow certain guidelines to keep automatic ownership assignment working and to maximize fault isolation.

Use these guidelines for configuring automatic disk ownership at the stack or loop level:

- Always assign all disks on the same loop or stack to the same system and pool.
- Always assign all loops or stacks connected to the same adapter to the same pool.
- For HA configurations using SyncMirror, pool0 is typically assigned to the local pool and pool1 is assigned to the remote pool.
- Always assign disks in the same multi-disk carrier to the same system.

Note: On storage system models that only support one loop or stack, you can configure your system to have both pools on a single loop or stack. Automatic assignment at the stack or loop level does not work, but automatic assignment at the shelf level works for this configuration.

Use these guidelines for configuring automatic disk ownership at the shelf level:

- Always assign all disks on the same shelf to the same system and pool.
- On storage systems that support two controllers but only a single stack, if the stack contains more than one shelf, you can use shelf-level assignment.

For more information about configuring SyncMirror with disks or array LUNs, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

For the guidelines for assigning disk ownership in a MetroCluster configuration, see the *Data ONTAP High Availability and MetroCluster Configuration Guide for 7-Mode*.

Assigning ownership for unowned disks and array LUNs

Disks and array LUNs must be owned by a storage system before they can be used in an aggregate. If your system is not configured to assign ownership automatically, or if your system contains array LUNs, you must assign ownership manually.

About this task

This procedure assigns ownership of disks and array LUNs that are currently unowned. If you want to change the ownership of disks or array LUNs that are already owned by a system, you must use the procedure “Modifying assignment of spare disks or array LUNs”.

Steps

1. Use the `disk show -n` command to view all disks and array LUNs that do not have assigned owners.

   Note: You must make array LUNs available to Data ONTAP before they can be assigned to a system.

2. Assign the disks and array LUNs that are labeled `Not Owned` to a storage system:

   `disk assign {`*`disk_list`*` | all | [-T `*`storage_type`*`] -n `*`count`*`|auto} [-c block | zoned] [-o `*`owner_name`*`] [-s `*`sysid`*`] [-f] [-p `*`pool`*`]`

   You can specify the disks and array LUNs to be assigned in the following ways:

   - *`disk_list`* specifies one or more individual disk or array LUN names.
     This is the most explicit way to specify disks and array LUNs. However, you have to manually enter each disk name.
   - You can use *`disk_list`* with the wildcard character (*) to specify a group of disks or array LUN names.
   - `all` specifies all unowned disks and array LUNs.
   - `-n` *`count`* specifies a number of unassigned disks and array LUNs to be assigned
   - `auto` causes Data ONTAP to initiate automatic disk ownership assignment.

     Note: Only disks installed in loops, stacks, or shelves that conform to the automatic ownership assignment guidelines are affected. Automatic ownership assignment does not operate on array LUNs.

     Use the following options to fine-tune automatic ownership assignment for disks:

     - To configure automatic ownership assignment to a stack, set the `disk.auto_assign` option to `on`.
     - To configure automatic ownership assignment to a shelf, the `disk.auto_assign` and `disk.auto_assign_shelf` options to set `on`.
   - You can use the `-shelf` option to manually assign unowned disks to the specified shelf.

You use the following options to further qualify which disks and array LUNs Data ONTAP assigns:

- `-T` specifies the Data ONTAP disk type of the storage to be assigned.

  Note: If you have different disk types or disks and array LUNs on your system, always use the `-T` option to ensure that Data ONTAP uses the disks or array LUNs that you expect. Without this option, Data ONTAP uses the type of disk or array LUN with the most spares.

  This option cannot be used with a list of disk or array LUN names. You must use the `-n` option with the `-T` option.

- `-c` specifies the checksum type for the disk or array LUNs to be assigned.
  For array LUNs, the checksum type can be `block` or `zoned`. The default checksum type is `block`. For more information about checksums for array LUNs, see the *V-Series Installation Requirements and Reference Guide*.
  For disks, the checksum type can be `block` or `zoned` (AZCS). Only the default checksum type is supported for disks.

You use the following options to specify the system to own the disks and array LUNs you are assigning.

Note: If you do not specify a system to own the disks and array LUNs, they are assigned to the local system.

- `-o` *`owner_name`* specifies the name of the system to which you want to assign the disks and array LUNs.
- `-s` *`sysid`* specifies the ID of the system that the disks and array LUNs are assigned to.
  This is an alternative to specifying the system name by using the `-o` option.
- `-f` is used for changing ownership for a disk or array LUN that is already owned by a system.
  For array LUNs, `-f` is also used to enable assignment of an array LUN to a system when there is a redundancy error.

`-p` specifies which SyncMirror pool the disks and array LUNs are assigned to. Its value is either `0` or `1`. If you do not specify a pool, the disks and array LUNs are assigned to pool0. You need to specify the pool only if SyncMirror is in use on your system.

3. Use the `disk show -v` command to verify the assignments that you just made.

After you finish

If you assigned ownership for array LUNs, you should verify that two paths exist for each array LUN and verify path failover to ensure that you have path redundancy.

Related concepts

How ownership for disks and array LUNs works on page 56
How disks and array LUNs become available for use on page 58
How Data ONTAP reports drive types on page 18

How you use the wildcard character with the disk ownership commands on page 72
Verifying the existence of two paths to an array LUN on page 69

Configuring automatic ownership assignment of disks

If you have unowned disks on a stack, loop, or shelf, you can configure Data ONTAP to automatically assign disk ownership at the stack or shelf level.

Before you begin

If you have multiple loops, one disk must have been manually assigned on each loop so that automatic ownership assignment will work on each loop. Automatic disk ownership assignment must have been set up (`auto` option) so that the system can automatically assign any unowned disks to the same system at the stack or loop level.

About this task

For most system configurations, you can use automatic assignment at the stack or loop level; for smaller configurations, you can use automatic assignment at shelf level. The first automatic assignment occurs after 10 minutes.

You can use automatic assignment at the shelf level for the following system configurations:

- Your storage system supports only one stack or loop.
- You cannot assign an entire stack or loop to a single system or pool.
- In a MetroCluster configuration consisting of one stack or loop for each node and two shelves, you can have one shelf for one pool and one shelf for the other pool.
- On storage systems that support two controllers but only a single stack with more than one shelf.

If the disks on the same shelf have home ownership split between two systems, you need to manually assign the disks to specific shelves.

You can enable or disable automatic disk ownership assignment. When the `disk.auto_assign` option is `on`, automatic assignment at the stack or loop level is enabled. This option is the default. If you want Data ONTAP to automatically assign disk ownership at the shelf level, set on both the `disk.auto_assign` and `disk.auto_assign_shelf` options.

Steps

1. Decide whether you want to set up automatic ownership assignment at the stack or shelf level.

If you want to...	Then...
Configure automatic ownership assignment at the stack or loop level	Set `options disk.auto_assign` to `on`.

If you want to...	Then...
Configure automatic ownership assignment at the shelf level	Set `options disk.auto_assign` to `on` and `options disk.auto_assign_shelf` to `on`.
Turn off automatic ownership assignment	Set `options disk.auto_assign` to `off` and `options disk.auto_assign_shelf` to `off`. Note that if you set `disk.auto_assign_shelf` to `on` while `disk.auto_assign` is set to `off`, automatic assignment does not occur.

For example, you have a shelf whose disks are owned by one system and another shelf on the same loop whose disks are owned by a different system. In this case, you would configure automatic ownership assignment at the shelf level.

Data ONTAP automatically assigns unowned disks on the stack, loop, or shelf, depending on the option settings.

2. Verify the automatic assignment settings for the disks:

   disk show -o *owner_name* -v

Modifying assignment of spare disks or array LUNs

You can change the ownership of a *spare* disk or array LUN to another storage system.

Before you begin

A disk or array LUN that is a spare has been assigned to a specific system, but it has not yet been added to an aggregate. If the disk or array LUN whose ownership you want to change is in an aggregate, you must do the following before you can change ownership of the disk or array LUN:

- For an array LUN that is part of an aggregate, you must first remove the LUN from the aggregate, which changes the state of the array LUN to spare. To remove an array LUN from an aggregate, you must destroy the aggregate.
- For a disk that is part of an aggregate, you must first perform a disk replace and make the disk a spare.

About this task

You can change ownership of disks only between nodes in an HA pair. You can change ownership of array LUNs among the systems in a V-Series neighborhood.

Steps

1. At the console of the storage system that owns the disk or array LUN that you want to reassign, enter the following to see a list of spare disks or spare array LUNs on the system:

   aggr status -s

2. On the system that owns the spare disk or array LUN you want to reassign, enter either of the following commands to reassign ownership of the disk or array LUN:

   disk assign LUN-or-disk-name -o *new_owner_name* -f
   or
   disk assign LUN-or-disk-name -s *sysID-of-receiving_system* -f

   `-o` is the name of the system that you want to be the new owner of the disk or array LUN.

   `-s` is the ID of the system that you want to be the new owner of the disk or array LUN. You can obtain the system ID of the destination system by running `sysconfig` on the destination system.

   `-f` is required to force the change.

3. Enter the following command to verify that the ownership of the spare disk or array LUN moved to the other system:

   aggr status -s

   The spare disk or array LUN that you moved should no longer appear in the list of spares.

4. On the destination system, enter the following command to verify that the spare disk or spare array LUN whose ownership you changed is listed as a spare owned by the destination system:

   aggr status -s

After you finish

If you changed ownership for array LUNs, you should verify that two paths exist for each array LUN and verify path failover to ensure that you have path redundancy. You must add the disk or array LUN to an aggregate before you can use it for storage.

Related concepts

How ownership for disks and array LUNs works on page 56
Verifying the existence of two paths to an array LUN on page 69

Verifying the existence of two paths to an array LUN

If the primary path fails, Data ONTAP automatically maps each storage system port to a secondary path. You want to ensure that there are two paths to each array LUN so that the V-Series system can continue to work when running on a single path.

Verifying the existence of two paths: storage show disk command

You should verify that your V-Series system is configured with two paths to an array LUN so that there is a secondary path in case the primary path fails or is taken offline.

Steps

1. Enter the following command to show the primary and secondary paths to LUNs:

   `storage show disk -p` ***`all`***

   The system displays information similar to the following:

```
PRIMARY     PORT             SECONDARY  PORT           SHELF BAY
ADAPTER
-------------------- ---- ------------------------- ---- ---
vnmc4500s32:4.127L1    -    vnmc4500s33:19.127L1   -     -    -   0a
vnmc4500s32:4.127L12   -    vnmc4500s33:19.127L12  -     -    -   0a
vnmc4500s33:19.127L2   -    vnmc4500s32:4.127L2    -     -    -   0c
vnmc4500s33:19.127L13  -    vnmc4500s32:4.127L13   -     -    -   0c
```

   Note: When you use the *`all`* variable, adapters are displayed, but unassigned LUNs are not visible.

2. Determine whether a primary path and a secondary path to the array LUNs are shown.

   If you do not see a primary and secondary path, check zoning, host group configuration, and cabling.

   Note: Do not continue with testing until you see two paths.

3. Look at the adapters shown to see whether all paths are on a single adapter.

   If you see both paths through one port (for example, both paths through the 0c port), this is an indication that the back-end zoning is redundantly crossed. This is not a supported configuration.

   Note: Data ONTAP changes the path to array LUNs, as necessary, for load balancing. Therefore, the primary and secondary paths for a given array LUN can change when the `storage show disk` command is issued at different times.

Verifying the existence of two paths: storage array show-config command

You should verify that your V-Series system is configured with two paths to an array LUN so that there is a secondary path in case the primary path fails or is taken offline.

Step

1. Enter the following command to show the primary and secondary paths to LUNs:

   `storage array show-config`

   You see information similar to the following.

```
LUN Group Array Name Array Target Ports    Switch Port Initiator
Group 0 (4 LUNS) HP_V210 50:00:1f:e1:50:0a:86:6d  vnmc4300s35:11 0b
                         50:00:1f:e1:50:0a:86:68  vnbr4100s31:1  0a
                         50:00:1f:e1:50:0a:86:6c  vnmc4300s35:6  0d
Group 1(50 LUNS) HP_V200 50:00:1f:e1:50:0d:14:6d  vnbr4100s31:5  0a
                         50:00:1f:e1:50:0d:14:68  vnmc4300s35:3  0d
```

   This example shows output from a V-Series system connected to two storage arrays. Each LUN group is comprised of LUNs that share the same two paths. Group 0 contains a total of 4 LUNs on the HP_V210 array and Group 1 contains 50 LUNs on the HP_V200 array.

   Array LUNs that are not configured with two paths are shown as one or more LUNs with a single path, similar to the following example.

```
LUN Group Array Name Array Target Ports    Switch Port  Initiator
        (4 LUNS) HP_V210 50:00:1f:e1:50:0a:86:6d  vnmc4300s35:11 0b
```

Verifying path failover for array LUNs

You want to demonstrate that the V-Series system continues to work when running with a single path, for example, when a switch or array port is taken offline. You can test path failover by physically removing fibre cables or taking ports offline using Data ONTAP commands.

The procedure you use to test path failover differs slightly, depending on whether you are testing a stand-alone system or an HA pair.

Verifying path failover for array LUNs in a stand-alone system

It is important to demonstrate that a stand-alone V-Series system continues to operate on a single path.

Steps

1. Set your privilege level to advanced:

```
priv set advanced
```

2. Set port *0a* offline by using the following command:

   ```
   fcadmin offline 0a
   ```

3. Show the number of disks seen on each adapter using the following command:

   ```
   sysconfig
   ```

   No disks will be assigned to adapter *0a*.

4. Show the primary and secondary paths by using the following command:

   ```
   storage show disk -p
   ```

5. Return port *0a* to online:

   ```
   fcadmin online 0a
   ```

Verifying path failover for array LUNs in an HA pair

It is important to demonstrate that controller failover and then path failover occur in an HA pair so that the system can to continue to operate on a single path.

Steps

1. Set your privilege level to advanced:

   ```
   priv set advanced
   ```

   You need to enter this command on the local and partner node.

2. On the local node, enter the following command to set port 0a offline (assuming the redundant port pair is 0a and 0c):

   ```
   fcadmin offline 0a
   ```

3. Verify that only one path is available on the port pair:

   ```
   storage show disk -p
   ```

4. Enter the following command to initiate HA pair takeover:

   ```
   cf takeover
   ```

5. On the partner node, enter the following command:

   ```
   cf giveback
   ```

6. After the partner node is back online, repeat Steps 1, 2, and 3 on the partner node.

Guidelines for assigning disks to SyncMirror pools

Assigned disks and array LUNs are associated with a pool, either pool0 or pool1. Keeping all disks on a loop or stack in the same pool ensures redundancy and supports automatic disk assignment.

Typically, pool0 is assigned to the local pool and pool1 is assigned to the remote pool.

On storage system models that only support one loop or stack, you can configure your system to have both pools on a single loop or stack. Automatic assignment at the stack or loop level does not work, but automatic assignment at the shelf level works for this configuration.

For more information about configuring SyncMirror with disks or array LUNs, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.

How you use the wildcard character with the disk ownership commands

You can use the wildcard character ("*") with some commands, including commands to manage disk ownership. However, you should understand how Data ONTAP expands the wildcard character.

You can use the wildcard character with the following disk ownership commands:

- `disk show`
- `disk assign`
- `disk remove_ownership`

When you use the wildcard character with these commands, Data ONTAP expands it with zero or more characters to create a list of disk names that will be operated on by the command. This can be very useful when you want to assign all of the disks attached to a particular port or switch, for example.

Note: Be careful when you use the wildcard character. It is accepted anywhere in the disk name string, and is a simple string substitution. Therefore, you might get unexpected results.

For example, to assign all disks on port 1 of the switch brocade23 to node03, you would use the following command:

```
disk assign brocade23:1.* -o node03
```

However, if you left off the second ".", as in the following command, you would assign all disks attached to ports 1, 10, 11, 12, and so on:

```
disk assign brocade23:1* -p 0
```

Assigning multiple disks attached to an HBA

To assign all of the disks attached to the B port of the HBA in expansion slot 5 to pool0, you would use the following command:

```
disk assign 5b.* -p 0
```

Managing array LUNs using Data ONTAP

For Data ONTAP to be able to use storage on a storage array, some tasks must be done on the storage array and some tasks must be done in Data ONTAP.

For example, the storage array administrator must create array LUNs for Data ONTAP use and map them to Data ONTAP. You can then assign them to systems running Data ONTAP.

If the storage array administrator wants to make configuration changes to an array LUN after it is assigned to a system, for example to resize it, you might need to perform some activities in Data ONTAP before it is possible to reconfigure the LUN on the storage array.

Array LUN name format

The array LUN name is a path-based name that includes the devices in the path between the V-Series system and the storage array, ports used, and the SCSI LUN ID on the path that the storage array presents externally for mapping to hosts.

On a V-Series system operating in 7-Mode, there are two names for each array LUN because there are two paths to each LUN, for example, *brocade3:6.126L1* and *brocade15:6.126L1*.

Array LUN format for systems operating in 7-Mode and releases prior to 8.0

Configuration	Array LUN name format	Component descriptions
Direct-attached	*adapter.idlun-id*	*adapter* is the adapter number on the V-Series system. *id* is the channel adapter port on the storage array. *lun-id* is the array LUN number that the storage array presents to hosts. Example: *0a.0L1*

Configuration	Array LUN name format	Component descriptions
Fabric-attached	*switch-name:port.idlun-id*	*switch-name* is the name of the switch. *port* is the switch port that is connected to the target port (the end point). *id* is the device ID. *lun-id* is the array LUN number that the storage array presents to hosts. Example: *brocade3:6.126L1* *brocade3:6.126* is the path component and *L1* is the SCSI LUN ID.

Checking the checksum type of spare array LUNs

If you plan to add a spare array LUN to an aggregate by specifying its name, you need to make sure that the checksum type of the array LUN you want to add is the same as the aggregate checksum type.

About this task

You cannot mix array LUNs of different checksum types in an array LUN aggregate. The checksum type of the aggregate and the checksum type of the array LUNs added to it must be the same.

If you specify a number of spare array LUNs to be added to an aggregate, by default Data ONTAP selects array LUNs of the same checksum type as the aggregate.

Note: Data ONTAP 8.1.1 and later supports a new checksum scheme called *advanced zoned checksum* (AZCS). Existing zoned checksum aggregates are still supported. The checksum type of all newly created aggregates using zoned checksum array LUNS is AZCS, which provides more functionality than the "version 1" zoned checksum type that was supported in previous releases and continues to be supported for existing zoned aggregates. Zoned checksum spare array LUNs added to an existing zoned checksum aggregate continue to be zoned checksum array LUNs. Zoned checksum spare array LUNs added to an AZCS checksum type aggregate use the AZCS checksum scheme for managing checksums.

Step

1. Check the checksum type of the spare array LUNs by entering the following command:

For...	The command is...
7-Mode	**`aggr status -s`** The output shows information about the spare disks or array LUNs on the system, including the checksum type of each. You can add a block checksum array LUN to a block checksum aggregate and a zoned array LUN to either a zoned aggregate or an AZCS checksum aggregate.

Changing the checksum type of an array LUN

You need to change the checksum type of an array LUN if you want to add it to an aggregate that is a different checksum type than the checksum type of the LUN.

Before you begin

Before changing the checksum type of an array LUN, you should have reviewed the tradeoffs between performance in certain types of workloads and storage capacity utilization of each checksum type. The *V-Series Installation Requirements and Reference Guide* contains information about checksum use for array LUNs. You can also contact your Sales Engineer for details about using checksums.

About this task

You need to assign a `zoned` checksum type to an array LUN that you plan to add to a zoned checksum aggregate or an advanced zoned checksum (AZCS) aggregate. When a zoned checksum array LUN is added to an AZCS aggregate, it becomes an advanced zoned checksum array LUN. Similarly, when a zoned checksum array LUN is added to a zoned aggregate, it is a zoned checksum type.

Step

1. Enter the following command:

   `disk assign` ***`LUN-name`*** **`-c`** ***`new_checksum_type`***

   `LUN-name` is the name of the array LUN whose checksum type you want to change.

   `new_checksum_type` can be `block` or `zoned`.

   The checksum type of the array LUN is changed to the new checksum type you specified.

Prerequisites to reconfiguring an array LUN on the storage array

If an array LUN has already been assigned (through Data ONTAP) to a particular V-Series system, the information Data ONTAP wrote to the array LUN must be removed before the storage administrator attempts to reconfigure the array LUN on the storage array.

When the storage array presents an array LUN to Data ONTAP, Data ONTAP collects information about the array LUN (for example, its size) and writes that information to the array LUN. Data ONTAP cannot dynamically update information that it wrote to an array LUN. Therefore, before the storage array administrator reconfigures an array LUN, you must use Data ONTAP to change the state of the array LUN to *unused*. (The array LUN is unused from the perspective of Data ONTAP.)

While changing the state of the array LUN to unused, Data ONTAP does the following:

- Terminates I/O operations to the array LUN.
- Removes the label for RAID configuration information and the persistent reservations from the array LUN, which makes the array LUN unowned by any V-Series system.

After this process finishes, no Data ONTAP information remains in the array LUN.

You can do the following after the array LUN's state is unused:

- Remove the mapping of the array LUN to Data ONTAP and make the array LUN available to other hosts.
- Resize the array LUN or change its composition.

If you want Data ONTAP to use the array LUN again after its size or composition is changed, you must present the array LUN to Data ONTAP again and assign the array LUN to a V-Series system again. Data ONTAP is aware of the new array LUN size or composition.

Changing array LUN size or composition

Reconfiguring the size or composition of an array LUN must be done on the storage array. If an array LUN has already been assigned to a V-Series system, you must use Data ONTAP to change the state of the array LUN to unused before the storage array administrator can reconfigure it.

Before you begin

The array LUN must be a spare array LUN before you can change its state to unused.

Steps

1. On the V-Series system, enter the following command to remove ownership information:
   `disk remove -w LUNfullname`
2. On the storage array, complete the following steps:

a) Unmap (unpresent) the array LUN from the V-Series systems in the neighborhood so that they can no longer see the array LUN.
b) Change the size or composition of the array LUN.
c) If you want Data ONTAP to use the array LUN again, present the array LUN to the V-Series systems again.

   At this point, the array LUN is visible to the FC initiator ports to which the array LUN was presented, but it cannot be used by any V-Series systems yet.

3. Enter the following command on the V-Series system that you want to be the owner of the array LUN:

   ```
   disk assign {disk_list | all | [-T storage_type] -n count|auto} [-c
   block | zoned] [-o owner_name] [-s sysid] [-f] [-p pool]
   ```

   After the ownership information is removed, the array LUN cannot be used by any V-Series system until the array LUN is assigned again to a system. You can leave the array LUN as a spare or add it to an aggregate. You must add the array LUN to an aggregate before the array LUN can be used for storage.

Removing one array LUN from use by Data ONTAP

If the storage array administrator no longer wants to use a particular array LUN for Data ONTAP, you must remove the information that Data ONTAP wrote to the LUN (for example, size and ownership) before the administrator can reconfigure the LUN for use by another host.

Before you begin

If the LUN that the storage array administrator no longer wants Data ONTAP to use is in an aggregate, you must take the aggregate to which the LUN belongs offline and destroy the aggregate before starting this procedure. Taking an aggregate offline and destroying it changes the LUN from a data LUN to a spare LUN.

Step

1. Enter the following command:

   ```
   disk remove -w LUNfullname
   ```

   LUNfullname is the full name of the array LUN.

Preparing array LUNs before removing a V-Series system from service

You must release the persistent reservations on all array LUNs assigned to a V-Series system before removing the system from service.

About this task

When you assign Data ONTAP ownership of an array LUN, Data ONTAP places persistent reservations (ownership locks) on that array LUN to identify which V-Series system owns the LUN. If you want the array LUNs to be available for use by other types of hosts, you must remove the persistent reservations that Data ONTAP put on those array LUNs. The reason is that some arrays do not allow you to destroy a reserved LUN if you do not remove the ownership and persistent reservations that Data ONTAP wrote to that LUN.

For example, the Hitachi USP storage array does not have a user command for removing persistent reservations from LUNs. If you do not remove persistent reservations through Data ONTAP before removing the V-Series system from service, you must call Hitachi technical support to remove the reservations.

Contact Technical Support for instructions about how to remove persistent reservations from LUNs before removing a V-Series system from service.

Commands to display information about your storage

Data ONTAP provides commands to display information about disks, array LUNs, disk space, and storage subsystems.

Commands to display drive and array LUN information

You can see information about your drives and array LUNs using several commands, including the `aggr`, `disk`, `fcstat`, `sasadmin`, `storage`, `sysconfig`, and `sysstat` commands.

Use this Data ONTAP command...	**To display information about..**
`aggr status -f`	Drives or array LUNs in your storage system that have failed, or that have been preemptively failed by Data ONTAP.
`aggr status -m`	Drives in your storage system that are currently in the maintenance center, that have been or are being sanitized, and that are being checked by Data ONTAP due to poor response time.
`aggr status -r`	RAID layout and status for all aggregates; how each drive is used in its aggregate.
`aggr status -s`	Hot spares available in your storage system.
`disk encrypt show`	Disks that have Storage Encryption enabled.
`disk maint status`	The status of drive maintenance tests that are in progress.
`disk sanitize status`	The status of the sanitization process, after the `disk sanitize start` command has been executed.
`disk show`	List of drives and array LUNs owned by a storage system, or unowned drives and array LUNs.
`disk show -v`	All of the information provided by the `disk show` command, plus the checksum of each drive.
`fcstat device_map`	A physical representation of where FC-AL attached drives reside in a loop and a mapping of the disks to the disk shelves.
`fcstat fcal_stats`	Error and exception conditions, and handler code paths executed.

Use this Data ONTAP command...	**To display information about..**
`fcstat link_stats`	Link event counts.
`sasadmin devstats`	Statistics for SAS-connected drives: command completion counts, frame in and out counts, error and timeout counts.
`sasadmin shelf[short]`	Logical view of SAS shelf (long and short view).
`storage array show-config`	The connectivity between the storage system and individual array LUN groups on a storage array by showing specific storage system FC initiator ports, switch ports, and storage array ports on the path to each array LUN group.
`storage array show-luns`	The array LUNs that the storage array presented to Data ONTAP over the storage array World Wide Port Names (WWPNs) you specify.
`storage show acp`	The Alternate Control Path (ACP) module. Specifies whether the mode is enabled and displays connectivity and configuration information.
`storage show disk -a`	Detailed information about drives, including SSD life-cycle reporting, presented in a report form that is easily interpreted by scripts. This content also appears in the STORAGE section of an AutoSupport report.
`storage show disk -p`	Primary and secondary paths to all disks and array LUNs.
`storage show disk -T -x`	The drive type (FCAL, LUN, SATA, and so on) along with the drive and array LUN information.
`storage show disk -x`	The drive ID, shelf, bay, serial number, vendor, model, and revision level of all drives and array LUNs.
`sysconfig -d`	The drive name in the Device column, followed by the expansion slot number, shelf, bay, channel, and serial number.
`sysconfig -h`	Each drive, along with the size displayed in appropriate units (KB, GB, or TB) as calculated using the powers of two. (GB = 1024 × 1024 × 1024)
`sysstat`	The number of kilobytes per second (kB/s) of data being read and written.

Commands to display space information

Seeing information about how space is being used in your aggregates and volumes and their Snapshot copies enables you to manage your storage more effectively.

Use this Data ONTAP command...	To display information about...
`aggr status -S`	Disk space usage for aggregates
`vol status -F`	Disk space usage by volumes within an aggregate
`vol status -S`	Disk space usage for volumes
`df`	Disk space usage for volumes or aggregates
`snap delta`	The estimated rate of change of data between Snapshot copies in a volume
`snap reclaimable`	The estimated amount of space freed if you delete the specified Snapshot copies

For more information about the `snap` commands, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*. For more information about the `df` and `aggr status -S` commands, see the appropriate man page.

Commands to display storage subsystem information

You can use the `acpadmin`, `environment`, `fcadmin`, `sasadmin`, `storage show`, and `sysconfig` commands to display information about your storage subsystems.

Note: For detailed information about these commands and their options, see the appropriate man pages.

Use this Data ONTAP command...	To display information about...
`acpadmin list_all`	Alternative Control Path (ACP) processors (SAS shelves only).
`environment shelf`	Environmental information for each host adapter, including SES configuration and SES path.

Use this Data ONTAP command...	**To display information about...**
`environment shelf_log`	Shelf-specific module log file information, for shelves that support this feature. Log information is sent to the `/etc/log/shelflog` directory and included as an attachment on AutoSupport reports.
`fcadmin channels`	WWPN information.
`fcadmin device_map`	What disks are on each loop and shelf.
`fcadmin link_state`	How the ports are connected.
`sasadmin expander`	What disks are attached to expander PHYs.
`sasadmin expander_phy_state`	Expander PHY state, dongle state and event counters, PHY statistics.
`sasadmin shelf [short]`	The disks on each shelf (or a specific disk shelf), including a pictorial representation of disk placement (long or short view).
`storage show`	All disks and host adapters on the system.
`storage show acp`	Connectivity and status information for the Alternate Control Path (ACP) module (SAS shelves only).
`storage show adapter`	FC host adapter attributes, including (as appropriate for the adapter type) a description, firmware revision level, Peripheral Component Interconnect (PCI) bus width, PCI clock speed, FC node name, cacheline size, FC packet size, link data rate, static random access memory (SRAM) parity, state, in use, redundant.
`storage show bridge [-v]`	Information about the bridges connected to this controller. The `-v` option includes information about the bridge SAS ports.
`storage show disk -p`	How many paths are available to each disk.

Use this Data ONTAP command...	To display information about...
`storage show expander`	SAS expander attributes, including shelf name, channel, module, shelf ID, shelf UID, IOM state, and the following information for the disks attached to the expander: disk ID, port state, partial path timeout, link rate, invalid word count, running disparity count, PHY reset problem, CRC error count, and PHY change count.
`storage show hub`	Hub attributes: hub name, channel, loop, shelf ID, shelf user ID (UID), term switch, shelf state, ESH state, and hub activity for each disk ID: loop up count, invalid cyclic redundancy check (CRC) count, invalid word count, clock delta, insert count, stall count, util.
`storage show mc`	All media changer devices that are installed in the system.
`storage show port`	Switch ports connected to the system.
`storage show switch`	Switches connected to the system.
`storage show tape`	All tape drive devices attached to the system.
`storage show tape supported [-v]`	All tape drives supported. With -v, information about density and compressions settings is also displayed.
`storage stats tape`	Statistics for all tape drives attached to the system.
`sysconfig -A`	All sysconfig reports, including configuration errors, disks, array LUNs, media changers, RAID details, tape devices, and aggregates.
`sysconfig -m`	Tape libraries.
`sysconfig -t`	Tape drives.

Enabling or disabling a host adapter

A host adapter can be enabled or disabled by using the `storage` command. You disable an adapter to replace hardware components or modules.

About this task

You might want to disable an adapter for the following reasons:

- You are replacing any of the hardware components connected to the adapter.
- You are replacing a malfunctioning I/O module.

You can disable an adapter only if all disks connected to it can be reached through another adapter. After an adapter connected to dual-connected disks has been disabled, the other adapter is not considered redundant; thus, the other adapter cannot be disabled.

Steps

1. Identify the name of the adapter whose state you want to change by entering the following command:

   `storage show adapter`

   The field that is labeled “Slot” lists the adapter name.

2. Enter the following command.

If you want to...	Then use this command
Enable the adapter	**`storage enable adapter`** ***`adapter_name`***
Disable the adapter	**`storage disable adapter`** ***`adapter_name`***

Introduction to Storage Encryption

Overview of Storage Encryption concepts, functionality, benefits, and limitations.

What Storage Encryption is

Storage Encryption is an optional feature that you can enable for additional data protection. It is available on certain supported storage controllers and disk shelves that contain disks with built-in encryption functionality.

In a standard storage environment, data is written to disk in cleartext format. This makes the data vulnerable to potential exposure to unauthorized users when disks removed from a storage system are lost or stolen.

When you enable Storage Encryption, the storage system protects your data at rest by storing it on self-encrypting disks.

The authentication keys used by the self-encrypting disks are stored securely on external key management servers.

Purpose of the external key management server

An external key management server is a third-party system in your storage environment that securely manages authentication keys used by the self-encrypting disks in the storage system. You link the external key management server to other systems that use authentication or encryption keys such as your storage system.

The storage system uses a secure SSL connection to connect to the external key management server to store and retrieve authentication keys. The communication between the storage system and key management server uses the Key Management Interoperability Protocol (KMIP).

The external key management server securely stores authentication or encryption keys entrusted to it and provides them upon demand to authorized linked systems. This provides an additional level of security by storing authentication keys separate from the storage system. Additionally, authentication keys are always handled and stored securely. The keys are never displayed in cleartext.

You must link at least one key management server to the storage system during the Storage Encryption setup and configuration process. You should link multiple key management servers for redundancy. If the only key management server in the environment becomes unavailable, access to protected data might become unavailable until the key management server is available again. For example, when the storage system needs to unlock self-encrypting disks but cannot retrieve the authentication key from the key management server because it is unavailable.

You can specify up to four key servers during or after setup for redundancy.

For a list of supported key management servers, see the Interoperability Matrix.

How Storage Encryption works

Storage Encryption occurs at the firmware level of disks that are equipped with special firmware and hardware to provide the additional security, also known as *self-encrypting disks (SEDs)*. SEDs can operate either in unprotected mode like regular disks, or in protected mode requiring authentication after the power-on process.

SEDs always encrypt data for storage. In unprotected mode, the encryption key needed to decrypt and access the data is freely available. In protected mode, the encryption key is protected and requires authentication to be used.

When you first enable and configure Storage Encryption on a storage system using SEDs, you create an authentication key that the storage system uses to authenticate itself to the SEDs. You configure the storage system with the IP address to one or more external key management servers that securely stores the authentication key.

The storage system communicates with the key management servers at boot time to retrieve the authentication keys. Data ONTAP requires the authentication keys to authenticate itself to the SEDs any time after the SEDs are power-cycled.

If the authentication is successful, the SEDs are unlocked. The SEDs use the authentication key to decrypt the data encryption keys stored inside the disk. When presented with a read request, SEDs automatically decrypt the stored data before passing it on to the storage system. When presented with a write request from the storage system, SEDs automatically encrypt the data before writing the data to the disk's storage platters. When the SEDs are *locked*, Data ONTAP must successfully authenticate itself to the disk before the SEDs allow data to be read or written. When locked, SEDs require authentication each time the disk is powered on.

Encryption and decryption happens without a perceptible disk performance decrease or boot time increase. Storage Encryption does not require a separate license key. The only additional required component is an external key management server.

When you halt and power down the storage system, including the disk shelves containing SEDs, the disks are locked again and the data becomes inaccessible.

Disk operations with SEDs

Most of the disk-related operations are identical for SEDs and regular disks.

Because storage encryption happens at a very low level, specifically the disk firmware, it does not affect any higher level functionality. The storage controller sees SEDs the same as regular disks, and all functionality remains the same.

There are some additional options and requirements with SEDs:

- Sanitizing disks
  There are additional options to sanitize disks when using SEDs.

- Moving aggregates
  Additional steps are required when moving aggregates that contain SEDs.
- Destroying disks
  An additional option enables you to make the disks permanently inaccessible.

Benefits of using Storage Encryption

There are several scenarios where using Storage Encryption provides significant benefits by protecting data from unauthorized access when disks removed from a storage system have fallen into the wrong hands.

Data protection in case of disk loss or theft

Storage Encryption protects your data if disks are lost or stolen.

Someone who comes into possession of disks that store data using Storage Encryption cannot access the data. Without the authentication key that is required to authenticate and unlock the disks, all attempts to read or write data result in an error message returned by the SEDs.

Circumventing the disk authentication by moving the platters into another disk without encryption firmware would be unsuccessful as well. The data stored on the platters appears as ciphertext and is fully protected from unauthorized access.

Data protection when returning disks to vendors

Storage Encryption protects your data when you return disks to vendors.

The following three options are available to protect data on disks that are removed from a storage system and returned to a vendor:

- If the SED is owned by a storage system, it requires authentication to access the data. Since the vendor does not know, or have access to, the authentication key, the vendor cannot access data on the disk.
- If you sanitize the disk before returning it to a vendor, it changes the encryption key to a new unknown key. Any subsequent attempts to read data from the disk result in random data.
- If you "destroy" the disk, it changes the encryption key to a random unknown key, it changes the authentication key to a random unknown key, and permanently locks the disk, preventing any further decryption of the data and access to the disk.

Related tasks

Sanitizing disks using Storage Encryption before return to vendor on page 101

Data protection when moving disks to end-of-life

Storage Encryption protects your data when moving a disk to an end-of-life state.

You can protect data on a disk by changing the authentication key to a random value that is not stored and permanently locking the drive. This prevents any further decryption of the data and access to the disk.

Related tasks

Setting the state of disks using Storage Encryption to end-of-life on page 102

Data protection through emergency data shredding

Storage Encryption protects your data in emergency situations by allowing you to instantaneously prevent access to the data on the disk.

This might include extreme scenarios where power to the storage system or the key management server (or both) is not available, or one or both have fallen into possession of a hostile third-party.

Related tasks

Emergency shredding of data on disks using Storage Encryption on page 103

Limitations of Storage Encryption

You must keep certain limitations in mind when using Storage Encryption.

- Storage Encryption is not supported with SnapLock.
  If a SnapLock license is installed on the storage system, Storage Encryption functionality is unavailable. If Storage Encryption is enabled on a storage system, you cannot add a SnapLock license.
- For the latest information about which storage systems, disk shelves, and key management servers are supported with Storage Encryption, see the Interoperability Matrix.
- All disks in the storage system and optional attached disk shelves must have encryption functionality to be able to use Storage Encryption. You cannot mix regular non-encrypting disks with self-encrypting disks.
- Storage Encryption is not supported with Flash Pool aggregates.
- Storage Encryption `key_manager` commands are only available for local nodes.
  They are not available in takeover mode for partner nodes.
- Do not configure Storage Encryption to use 10 Gigabit network interfaces for communication with key management servers. This limitation does not apply to serving data.
- Storage Encryption supports a maximum of 128 authentication keys per key management server.

You receive a warning when the number of stored authentication keys reaches 100. You cannot create new authentication keys when the number of stored authentication keys reaches the limit of 128. You must then delete unused authentication keys before you can create new ones.

Related information

Interoperability Matrix: support.netapp.com/matrix

Managing Storage Encryption

You can perform various tasks to manage Storage Encryption, including viewing and removing key management servers, and creating, deleting, restoring and synchronizing authentication keys.

Displaying Storage Encryption disk information

You can display information about self-encrypting disks by using the `disk encrypt show` command. This command displays the key ID and lock status for each self-encrypting disk.

About this task

The key ID displayed in the command output is an identifier used by Storage Encryption and key management servers as a reference to the authentication key. It is not the actual authentication key or the data encryption key.

Step

1. To display information about SEDs, enter the following command:

   `disk encrypt show`

   The `disk encrypt show`, `lock`, and `rekey` commands support extended wildcard matching. For more information, see the `disk encrypt show` man page.

Example

The following command displays the status of each self-encrypting disk:

```
storage-system> disk encrypt show
Disk       Key
ID
Locked?
0c.00.1
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
0c.00.0
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
0c.00.3
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
0c.00.4
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
0c.00.2
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
```

```
0c.00.5
080CF0C8000000000100000000000000A948EE8604F4598ADFFB185B5BB7FED3
Yes
```

Displaying key management servers

You can display information about the external key management servers associated with the storage system by using the `key_manager show` command.

Step

1. To display external key management servers, enter the following command:

   `key_manager show`

   All external key management servers associated with the storage system are listed.

Example

The following command displays all external key management servers associated with the storage system:

```
storage-system> key_manager show
  172.18.99.175
```

Verifying key management server links

You use the `key_manager status` or `key_manager query` commands to verify that all key management servers are successfully linked to the storage system. These commands are useful for verifying proper operation and troubleshooting.

About this task

Both commands display whether key management servers are responding or not.

Step

1. Perform one of the following actions:

If you want to...	Then enter the following command:
Check the status of a specific key management server	**`key_manager status -key_server`** ***`key_server_ip_address`***

If you want to...	Then enter the following command:
Check the status of all key management servers	**`key_manager status`**
Check the status of all key management servers and view additional server details.	**`key_manager query`** The `key_manager query` command displays additional information about key tags and key IDs.

Examples

The following command checks the status of all key management servers linked to the storage system:

```
storage-system> key_manager status
Key server                      Status
172.16.132.118                   Server is responding
172.16.132.211                   Server is responding
```

The following command checks the status of all key management servers linked to the storage system and displays additional information:

```
storage-system> key_manager query
Key server 172.16.132.118 is responding.
Key server 172.16.132.211 is responding.

Key server 172.16.132.118 reports 4 keys.

Key tag                          Key ID
--------                         -------
storage-system                      080CDCB20...
storage-system                      080CDCB20...
storage-system                      080CDCB20...
storage-system                      080CDCB20...

Key server 172.16.132.211 reports 4 keys.

Key tag                          Key ID
--------                         -------
storage-system                    *080CDCB20...
storage-system                     080CDCB20...
storage-system                     080CDCB20...
storage-system                    *080CDCB20...
```

After you finish

The output of the `key_manager query` command might display key IDs marked with an asterisk (*). This indicates that keys exist on a key server but are not currently available in the Data ONTAP

key table. Run the `key_manager restore` command to import those keys from the key management server into the key table.

Adding key management servers

You can use the `key_manager add` command to link key management servers to the storage system. This allows you to add additional key management servers for redundancy after initial setup or to replace existing key management servers.

Before you begin

You must first install the required storage system and key management server SSL certificates. If they are not present, the command fails.

You must know the IP address for each key management server you want to link.

Step

1. To add a key management server, enter the following command:

   `key_manager add -key_server` ***`key_server_ip_address`***

Example

The following command adds a link from the storage system to the key management server with the IP address 172.16.132.118:

```
storage-system> key_manager add -key_server 172.16.132.118
Found client certificate file client.pem.
Registration successful for client.pem.
Found client private key file client_private.pem.
Is this file protected by a passphrase? [no]: no
Registration successful for client_private.pem.
Registering 1 key servers...
Found client CA certificate file 172.16.132.118_CA.pem.
Registration successful for 172.16.132.118_CA.pem.
Registration complete.
```

Removing key management servers

You can remove a key management server linked to the storage system by using the `key_manager remove` command.

Before you begin

Storage Encryption requires at least one key management server linked to the storage appliance to operate. If you want to replace a single key management server with another one, first add the new one before removing the old one.

You must know the IP address for each key management server you want to remove.

Step

1. To remove key management servers, enter the following command:

   `key_manager remove -key_server` ***`key_server_ip_address`***

   `-key_server` *`key_server_ip_address`* specifies the IP address of the key management server you want to remove.

Example

The following command removes the link between the storage system and the key management server with the IP address 172.18.99.175:

```
storage-system> key_manager remove -key_server 172.18.99.175
Key server 172.18.99.175 will be unregistered from service.
Unregistration successful.
```

What happens when key management servers are not reachable during the boot process

Data ONTAP takes certain precautions to avoid undesired behavior in the event that the storage system cannot reach any of the specified key management servers during the boot process.

If the storage system is configured for Storage Encryption, the SEDs have been rekeyed and locked, and the SEDs are powered on, the storage system must retrieve the required authentication keys from the key management servers to authenticate itself to the SEDs before it can access the data.

The storage system attempts to contact the specified key management servers for up to three hours. If the storage system cannot reach any of them after that time, the boot process stops and the storage system halts.

If the storage system successfully contacts any specified key management server, it then attempts to establish an SSL connection for up to 15 minutes. If the storage system cannot establish an SSL connection with any specified key management server, the boot process stops and the storage system halts.

While the storage system attempts to contact and connect to key management servers, it displays detailed information about the failed contact attempts at the CLI. You can interrupt the contact attempts at any time by pressing Ctrl-C.

As a security measure, SEDs allow only a limited number of unauthorized access attempts, after which they permanently disable access to the existing data. If the storage system cannot contact any specified key management servers to obtain the proper authentication keys, it can only attempt to authenticate with the default key which leads to a failed attempt and a panic. If the storage system is configured to automatically reboot in case of a panic, it would enter a boot loop which results in continuous failed authentication attempts on the SEDs.

Halting the storage system in these scenarios is by design to prevent the storage system from entering a boot loop and unintended data loss as a result of the SEDs locked permanently due to exceeding the safety limit of 1024 consecutive failed authentication attempts.

If you encounter this scenario where the storage system is halted due to failure to reach any specified key management servers, you must first identify and correct the cause for the communication failure before you attempt to continue booting the storage system.

Changing the authentication key

You can change the authentication key at any time by using the `key_manager rekey` command. You might want to change the authentication key as part of your security protocol or when moving an aggregate to another storage system.

Step

1. Perform one of the following actions:

If you want to...	Then...
Change the authentication key and enter a new one manually	**a.** Enter the following command at the storage system prompt: **`key_manager rekey -manual -key_tag key_tag`** **b.** When prompted, enter the new authentication key. It must be 20 to 32 characters long.
Change the authentication key and have the system generate a new one automatically	Enter the following command at the storage system prompt: **`key_manager rekey -key_tag key_tag`**

`key_tag` is the label used to associate keys with a particular storage system. If you do not specify a key tag, the storage system uses the key tag specified when you set up Storage Encryption. If

you did not specify this key tag during setup, it uses the parent key tag as the default. Each node has a parent key tag. HA pair members share the same parent key tag.

Example

The following command changes the authentication key and prompts you to enter a new one manually. You can run the `disk encrypt show` command after completion to verify the results.

```
storage-system> key_manager rekey -manual
Please enter a new passphrase:
Please reenter the new passphrase:

New passphrase generated.
Key ID:
080CDCB200000000010000000000000B0A11CBF3DDD20EFB0FBB5EE198DB22A
Key tag: storage-system

Notice: Remember to store the passphrase and the Key ID in a secure
location.

Passphrase, key ID, and key tag synchronized with the following key
server(s):
 172.16.132.118
 172.16.132.211
Completed rekey on 4 disks: 4 successes, 0 failures, including 0
unknown key and 0 authentication failures.
```

Retrieving authentication keys

You can use the `key_manager restore` command to retrieve authentication keys from a key management server to a storage system. For example, when you created authentication keys on a node, you use this command to retrieve the keys for use on the partner node.

Before you begin

You must know the IP address for each key management server you want to retrieve authentication keys from.

Step

1. To retrieve authentication keys from a key management server to the storage system, enter the following command:
   `key_manager restore -key_server` *`key_server_ip_address`* `-key_tag` *`key_tag`*

   If all specified key management servers are available, you can use the `-all` option instead of the `-key_server` option to clear out the current Data ONTAP key table and retrieve all keys matching the specified key tag from all specified key management servers.

Examples

The following command restores keys with the key tag storage-system from the key management server with the IP address 172.18.99.175:

```
storage-system> key_manager restore -key_server 172.18.99.175 -
key_tag storage-system
```

The following command restores all keys with the key tag storage-system from all key management servers linked to the storage system:

```
storage-system> key_manager restore -all -key_tag storage-system
```

Deleting an authentication key

You can delete an authentication key that is no longer needed by removing it from the external key management server.

Before you begin

Verify that the authentication key is no longer needed before deleting it. Deleting an authentication key that is still in use can permanently prevent access to data on a storage system.

Step

1. Refer to the documentation for the external key management server for details on how to delete stored authentication keys.

SSL issues due to expired certificates

If the SSL certificates used to secure key management communication between the storage system and key management servers expire, the storage system can no longer retrieve authentication keys from the key management server at bootup. This issue can cause data on SEDs to be unavailable. You can prevent this issue by updating all SSL certificates before their individual expiration dates.

SSL certificates have a limited lifespan because they have an expiration date. After the SSL certificates reach their expiration dates, the certificates are no longer valid. When this happens, SSL connections that use expired certificates fail.

For Storage Encryption, this means that the SSL connections between the storage system and the key management servers fail, the storage system no longer can retrieve authentication keys when needed, and data access to the SEDs fails, resulting in storage system panic and downtime.

To prevent this issue from occurring, you must keep track of the expiration dates of all installed SSL certificates so that you can obtain new SSL certificates before they expire.

After you have obtained the new certificate files, you must first remove the existing certificate files from the storage system, and then install the new certificates on the storage system.

Steps

1. *Removing old SSL certificates before installing new ones* on page 99
   If you want to update or reinstall the SSL certificates used by Storage Encryption, you must first manually remove the old ones to ensure that the new ones are used.
2. *Installing replacement SSL certificates on the storage system* on page 99
   After you remove the old certificates, you create the new replacement SSL certificates, save them with the proper file name and format, and then install them on the storage system.

Removing old SSL certificates before installing new ones

If you want to update or reinstall the SSL certificates used by Storage Encryption, you must first manually remove the old ones to ensure that the new ones are used.

Steps

1. Remove the IP addresses of all key management servers by entering the following command for each key management server:

   `key_manager remove -key_server` ***`key_server_ip_address`***

2. Remove the storage system's client certificates by entering the following commands:

   `keymgr delete cert client_private.pem`

   `keymgr delete cert client.pem`

3. Remove all installed key management server certificates by entering the following commands for each key management server:

   `keymgr delete cert` ***`key_server_ip_address`*****`_CA.pem`**

Installing replacement SSL certificates on the storage system

After you remove the old certificates, you create the new replacement SSL certificates, save them with the proper file name and format, and then install them on the storage system.

Before you begin

You must have removed the old certificates that are about to expire from the storage system.

You must have obtained the replacement public and private certificates for the storage system and the public certificate for the key management server and named them as required. For more information, see the *Data ONTAP Software Setup Guide for 7-Mode*.

You must have installed the appropriate new certificates on the key management server. For more information, see the documentation for your key management server.

Steps

1. Copy the certificate files to a temporary location on the storage system.
2. Install the public certificate of the storage system by entering the following command:

   keymgr install cert *<path>*client.pem
3. Install the private certificate of the storage system by entering the following command:

   keymgr install cert *<path>*client_private.pem
4. Install the public certificate of all key management servers by entering the following command for each key management server:

   keymgr install cert *<path>/key_management_server_ipaddress*_CA.pem
5. Add all key management servers by entering the following command for each key management server:

   key_manager add -key_server *key_server_ip_address*
6. Verify connectivity between the storage system and key management servers by entering the following command:

   key_manager query

   You should see a list of existing key IDs retrieved from the key management servers.

Destroying data on disks using Storage Encryption

You can destroy data stored on disks using Storage Encryption for security reasons, including sanitizing the disks, setting the disk state to end-of-life, and emergency shredding of the data.

Related tasks

Using disk sanitization to remove data from disks on page 46

Sanitizing disks using Storage Encryption before return to vendor

If you want to return a disk to a vendor but do not want anyone to access sensitive data on the disk, you can sanitize it first by using the `disk encrypt sanitize` command. This renders the data on the disk inaccessible, but the disk can be reused. This command only works on spare disks.

Steps

1. Migrate any data that needs to be preserved to a different aggregate.
2. Destroy the aggregate.
3. Identify the disk ID for the disk to be sanitized by entering the following command:
   `disk encrypt show`
4. Enter the following command:
   `disk encrypt sanitize` *`disk_ID`*

Example

The following command sanitizes a self-encrypting disk with the disk ID 0c.00.3. You can run the `sysconfig -r` command before and after the operation to verify the results.

```
storage-system> sysconfig -r
Aggregate aggr0 (online, raid_dp) (block checksums)
  Plex /aggr0/plex0 (online, normal, active)
    RAID group /aggr0/plex0/rg0 (normal)

      RAID Disk    Device     HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)    Phys (MB/blks)
      ---------    ------     ------------- ---- ---- ---- ----- --------------    --------------
      dparity     0c.00.0    0c    0   0   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
      parity      0c.00.1    0c    0   1   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
      data        0c.00.2    0c    0   2   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688


Spare disks

RAID Disk    Device     HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)    Phys (MB/blks)
---------    ------     ------------- ---- ---- ---- ----- --------------    --------------
Spare disks for block or zoned checksum traditional volumes or aggregates
```

```
spare       0c.00.3    0c    0   3   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
spare       0c.00.4    0c    0   4   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
spare       0c.00.5    0c    0   5   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688

storage-system> disk encrypt sanitize 0c.00.3
storage-system> Wed Jun 30 17:49:16 PDT [disk.failmsg:error]: Disk 0c.00.3 (3SL04F3V00009015WTHU):
message received.
Wed Jun 30 17:49:16 PDT [raid.disk.unload.done:info]: Unload of Disk 0c.00.3 Shelf 0 Bay 3 [SYSTEM
X415_S15K7560A15 NQS3] S/N [3SL04F3V00009015WTHU] has completed successfully

storage-system> Wed Jun 30 17:49:25 PDT [disk.sanit.complete:info]: Disk 0c.00.3 [S/N
3SL04F3V00009015WTHU] has completed sanitization.

storage-system> sysconfig -
r
Aggregate aggr0 (online, raid_dp) (block checksums)
  Plex /aggr0/plex0 (online, normal, active)
    RAID group /aggr0/plex0/rg0 (normal)

      RAID Disk    Device     HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)    Phys (MB/blks)
      ---------    ------     ------------- ---- ---- ---- ----- --------------    --------------
      dparity     0c.00.0    0c    0   0   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
      parity      0c.00.1    0c    0   1   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
      data        0c.00.2    0c    0   2   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688


Spare disks

RAID Disk    Device     HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)    Phys (MB/blks)
---------    ------     ------------- ---- ---- ---- ----- --------------    --------------
Spare disks for block or zoned checksum traditional volumes or aggregates
spare       0c.00.4    0c    0   4   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
spare       0c.00.5    0c    0   5   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688

Maintenance disks

RAID Disk    Device     HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)    Phys (MB/blks)
---------    ------     ------------- ---- ---- ---- ----- --------------    --------------
sanitized    0c.00.3    0c    0   3   SA:B   -  SAS  15000 560000/1146880000 560208/1147307688
storage-system>
```

Setting the state of disks using Storage Encryption to end-of-life

If you want to render a disk permanently unusable and the data on it inaccessible, you can set the state of the disk to end-of-life by using the `disk encrypt destroy` command. This command only works on spare disks.

Steps

1. Remove any data from the aggregate containing the disk.
2. Migrate any data that needs to be preserved to a different aggregate.
3. Destroy the aggregate.
4. Enter the following command:
   disk encrypt destroy *disk_ID*

Result

The disk's encryption key is set to an unknown random value and the disk is irreversibly locked. The disk is now completely unusable and can be safely disposed of without risk of unauthorized data access.

Emergency shredding of data on disks using Storage Encryption

In case of a security emergency, you can instantly prevent access to data on disks using Storage Encryption, even if power is not available to the storage system or the external key server.

Before you begin

You must configure the external key server so that it only operates if an easily destroyed authentication item (for example, a smart card or USB drive) is present. Refer to the documentation for the external key management server for more details.

About this task

The steps for emergency shredding vary depending on whether power is available to the storage system and the external key server.

Step

1. Perform one of the following actions:

If...	Then...
Power is available to the storage system and you have time to gracefully take the storage system offline	**a.** If the storage system is a node in an HA pair, disable takeover. **b.** Take all aggregates offline and destroy them. **c.** Halt the storage system. **d.** Boot into maintenance mode. **e.** Enter the following command: `disk encrypt sanitize -all` This leaves the storage system in a permanently disabled state with all data erased. To use the storage system again, you must set it up from the beginning. For more information, see the *Data ONTAP Software Setup Guide for 7-Mode*.

If...	Then...
Power is available to the storage system and you must shred the data immediately; time is critical	**a.** If the storage system is a node in an HA pair, disable takeover. **b.** Set the privilege level to advanced. **c.** Enter the following command: `disk encrypt sanitize -all` The storage system panics which is expected due to the abrupt nature of the procedure. It leaves the storage system in a permanently disabled state with all data erased. To use the storage system again, you must set it up from the beginning. For more information, see the *Data ONTAP Software Setup Guide for 7-Mode*.
Power is available to the external key server but not to the storage system	**a.** Log in to the external key server. **b.** Destroy all keys associated with the disks containing data to protect.
Power is not available to the external key server or the storage system	Destroy the authentication item for the key server (for example, the smart card). If power to the systems is restored, the external key server cannot operate due to the missing authentication item. This prevents access to the disk encryption keys by the storage system, and therefore access to the data on the disks.

How Data ONTAP uses RAID to protect your data and data availability

Understanding how RAID protects your data and data availability can help you administer your storage systems more effectively.

For native storage, Data ONTAP uses RAID-DP (double-parity) or RAID Level 4 (RAID4) protection to ensure data integrity within a RAID group even if one or two of those drives fail. Parity drives provide redundancy for the data stored in the data drives. If a drive fails (or, for RAID-DP, up to two drives), the RAID subsystem can use the parity drives to reconstruct the data in the drive that failed.

For array LUNs, Data ONTAP stripes data across the array LUNs using RAID0. The storage arrays, not Data ONTAP, provide the RAID protection for the array LUNs that they make available to Data ONTAP.

Related tasks

Controlling the performance impact of RAID-level scrubbing on page 123

RAID protection levels for disks

Data ONTAP supports two levels of RAID protection for aggregates composed of disks in native disk shelves: RAID-DP and RAID4. RAID-DP is the default RAID level for new aggregates.

For more information about configuring RAID, see *Technical Report 3437: Storage Subsystem Resiliency Guide*.

Related information

TR 3437: Storage Subsystem Resiliency Guide

What RAID-DP protection is

If an aggregate is configured for RAID-DP protection, Data ONTAP reconstructs the data from one or two failed disks within a RAID group and transfers that reconstructed data to one or two spare disks as necessary.

RAID-DP provides double-parity disk protection when the following conditions occur:

- There is a single-disk failure or double-disk failure within a RAID group.
- There are media errors on a block when Data ONTAP is attempting to reconstruct a failed disk.

The minimum number of disks in a RAID-DP group is three: at least one data disk, one regular parity disk, and one double-parity (dParity) disk.

If there is a data-disk failure or parity-disk failure in a RAID-DP group, Data ONTAP replaces the failed disk in the RAID group with a spare disk and uses the parity data to reconstruct the data of the failed disk on the replacement disk. If there is a double-disk failure, Data ONTAP replaces the failed disks in the RAID group with two spare disks and uses the double-parity data to reconstruct the data of the failed disks on the replacement disks.

RAID-DP is the default RAID type for all aggregates.

What RAID4 protection is

RAID4 provides single-parity disk protection against single-disk failure within a RAID group. If an aggregate is configured for RAID4 protection, Data ONTAP reconstructs the data from a single failed disk within a RAID group and transfers that reconstructed data to a spare disk.

The minimum number of disks in a RAID4 group is two: at least one data disk and one parity disk.

If there is a single data or parity disk failure in a RAID4 group, Data ONTAP replaces the failed disk in the RAID group with a spare disk and uses the parity data to reconstruct the failed disk's data on the replacement disk. If no spare disks are available, Data ONTAP goes into degraded mode and alerts you of this condition.

Attention: With RAID4, if there is a second disk failure before data can be reconstructed from the data on the first failed disk, there will be data loss. To avoid data loss when two disks fail, you can select RAID-DP. This provides two parity disks to protect you from data loss when two disk failures occur in the same RAID group before the first failed disk can be reconstructed.

Related concepts

How Data ONTAP handles a failed disk with a hot spare on page 116
How Data ONTAP handles a failed disk that has no available hot spare on page 116
About degraded mode on page 115

RAID protection for array LUNs

Storage arrays provide the RAID protection for the array LUNs that they make available to Data ONTAP; Data ONTAP does not provide the RAID protection.

Data ONTAP uses RAID0 (striping) for array LUNs. Data ONTAP supports a variety of RAID types on the storage arrays, except RAID0 because RAID0 does not provide storage protection.

When creating *RAID groups* on storage arrays, you need to follow the best practices of the storage array vendor to ensure that there is an adequate level of protection on the storage array so that disk failure does not result in loss of data or loss of access to data.

Note: A *RAID group* on a storage array is the arrangement of disks that together form the defined RAID level. Each RAID group supports only one RAID type. The number of disks that you select for a RAID group determines the RAID type that a particular RAID group supports. Different

storage array vendors use different terms to describe this entity—RAID groups, parity groups, disk groups, Parity RAID groups, and other terms.

Data ONTAP supports RAID4 and RAID-DP on the native disk shelves connected to a V-Series system but does not support RAID4 and RAID-DP with array LUNs.

RAID protection for Data ONTAP-v storage

Because Data ONTAP-v storage is connected to the host server, rather than a storage system running Data ONTAP, the host server provides the RAID protection for the physical disks. Data ONTAP uses RAID0 for the virtual disks to optimize performance.

See the *Data ONTAP Edge Installation and Administration Guide* for more information.

Protection provided by RAID and SyncMirror

Combining RAID and SyncMirror provides protection against more types of drive failures than using RAID alone.

You can use RAID in combination with the SyncMirror functionality, which also offers protection against data loss due to drive or other hardware component failure. SyncMirror protects against data loss by maintaining two copies of the data contained in the aggregate, one in each plex. Any data loss due to drive failure in one plex is repaired by the undamaged data in the other plex.

For more information about SyncMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

The following tables show the differences between using RAID alone and using RAID with SyncMirror:

Table 5: RAID-DP and SyncMirror

Criteria	RAID-DP alone	RAID-DP with SyncMirror
Failures protected against	• Single-drive failure • Double-drive failure within a single RAID group • Multiple-drive failures, as long as no more than two drives within a single RAID group fail	• All failures protected against by RAID-DP alone • Any combination of failures protected against by RAID-DP alone in one plex, concurrent with an unlimited number of failures in the other plex • Storage subsystem failures (HBA, cables, shelf), as long as only one plex is affected
Failures *not* protected against	• Three or more concurrent drive failures within a single RAID group • Storage subsystem failures (HBA, cables, shelf) that lead to three or more concurrent drive failures within a single RAID group	• Three or more concurrent drive failures in a single RAID group on both plexes
Required resources per RAID group	*n* data drives + 2 parity disks	2 x (*n* data drives + 2 parity drives)
Performance cost	Almost none	Low mirroring overhead; can improve performance
Additional cost and complexity	None	SyncMirror license and configuration

Table 6: RAID4 and SyncMirror

Criteria	RAID4 alone	RAID4 with SyncMirror
Failures protected against	• Single-disk failure • Multiple-disk failures, as long as no more than one disk within a single RAID group fails	• All failures protected against by RAID4 alone • Any combination of failures protected against by RAID4 alone in one plex, concurrent with an unlimited number of failures in the other plex • Storage subsystem failures (HBA, cables, shelf), as long as only one plex is affected
Failures *not* protected against	• Two or more concurrent drive failures within a single RAID group • Storage subsystem failures (HBA, cables, shelf) that lead to two or more concurrent drive failures within a single RAID group	• Two or more concurrent drive failures in a single RAID group on both plexes
Required resources per RAID group	*n* data drives + 1 parity drive	2 x (*n* data drives + 1 parity drive)
Performance cost	None	Low mirroring overhead; can improve performance
Additional cost and complexity	None	SyncMirror configuration and extra storage requirement

Table 7: RAID0 and SyncMirror

Criteria	RAID0 alone	RAID0 with SyncMirror
Failures protected against	No protection against any failures RAID protection is provided by the RAID implemented on the storage array.	Any combination of array LUN, connectivity, or hardware failures, as long as only one plex is affected

Criteria	RAID0 alone	RAID0 with SyncMirror
Failures *not* protected against	No protection against any failures RAID protection is provided by the RAID implemented on the storage array.	Any concurrent failures that affect both plexes
Required array LUN resources per RAID group	No extra array LUNs required other than *n* data array LUNs	2 x *n* data array LUNs
Performance cost	None	Low mirroring overhead; can improve performance
Additional cost and complexity	None	SyncMirror configuration and extra storage requirement

Understanding RAID disk types

Data ONTAP classifies disks as one of four types for RAID: data, hot spare, parity, or dParity. The RAID disk type is determined by how RAID is using a disk; it is different from the Data ONTAP disk type.

Data disk Holds data stored on behalf of clients within RAID groups (and any data generated about the state of the storage system as a result of a malfunction).

Spare disk Does not hold usable data, but is available to be added to a RAID group in an aggregate. Any functioning disk that is not assigned to an aggregate but is assigned to a system functions as a hot spare disk.

Parity disk Stores row parity information that is used for data reconstruction when a single disk drive fails within the RAID group.

dParity disk Stores diagonal parity information that is used for data reconstruction when two disk drives fail within the RAID group, if RAID-DP is enabled.

How Data ONTAP RAID groups work

A RAID group consists of one or more data disks or array LUNs, across which client data is striped and stored, and up to two parity disks, depending on the RAID level of the aggregate that contains the RAID group.

RAID-DP uses two parity disks to ensure data recoverability even if two disks within the RAID group fail.

RAID4 uses one parity disk to ensure data recoverability if one disk within the RAID group fails.

RAID0 does not use any parity disks; it does not provide data recoverability if any disks within the RAID group fail.

How RAID groups are named

Within each aggregate, RAID groups are named rg0, rg1, rg2, and so on in order of their creation. You cannot specify the names of RAID groups.

About RAID group size

A RAID group has a maximum number of disks or array LUNs that it can contain. This is called its maximum size, or its size. A RAID group can be left partially full, with fewer than its maximum number of disks or array LUNs, but storage system performance is optimized when all RAID groups are full.

Related references

Storage limits on page 342

Considerations for sizing RAID groups for drives

Configuring an optimum RAID group size for an aggregate made up of drives requires a trade-off of factors. You must decide which factor—speed of recovery, assurance against data loss, or maximizing data storage space—is most important for the aggregate that you are configuring.

You change the size of RAID groups on a per-aggregate basis. You cannot change the size of an individual RAID group.

HDD RAID groups

You should follow these guidelines when sizing your RAID groups composed of HDDs:

- All RAID groups in an aggregate should have the same number of disks.
  If this is impossible, any RAID group with fewer disks should have only one less disk than the largest RAID group.
- The recommended range of RAID group size is between 12 and 20.
  The reliability of SAS and FC disks can support a RAID group size of up to 28, if needed.
- If you can satisfy the first two guidelines with multiple RAID group sizes, you should choose the larger size.

SSD RAID groups in Flash Pool aggregates

The SSD RAID group size can be different from the RAID group size for the HDD RAID groups in a Flash Pool aggregate. Usually, you should ensure that you have only one SSD RAID group for a Flash Pool aggregate, to minimize the number of SSDs required for parity.

SSD RAID groups in SSD-only aggregates

You should follow these guidelines when sizing your RAID groups composed of SSDs:

- All RAID groups in an aggregate should have the same number of drives.
  If this is impossible, any RAID group with fewer drives should have only one less drive than the largest RAID group.
- The recommended range of RAID group size is between 20 and 28.

Related concepts

How Flash Pool aggregates work on page 130

Considerations for Data ONTAP RAID groups for array LUNs

Setting up Data ONTAP RAID groups for array LUNs requires planning and coordination with the storage array administrator so that the administrator makes the number and size of array LUNs you need available to Data ONTAP.

For array LUNs, Data ONTAP uses RAID0 RAID groups to determine where to allocate data to the LUNs on the storage array. The RAID0 RAID groups are not used for RAID data protection. The storage arrays provide the RAID data protection.

Note: Data ONTAP RAID groups are similar in concept to what storage array vendors call RAID groups, parity groups, disk groups, Parity RAID groups, and other terms.

Follow these steps when planning your Data ONTAP RAID groups for array LUNs:

1. Plan the size of the aggregate that best meets your data needs.
2. Plan the number and size of RAID groups that you need for the size of the aggregate.
   Follow these guidelines:
   - RAID groups in the same aggregate should be the same size with the same number of LUNs in each RAID group. For example, you should create four RAID groups of 8 LUNs each, not three RAID groups of 8 LUNs and one RAID group of 6 LUNs.
   - Use the default RAID group size for array LUNs, if possible. The default RAID group size is adequate for most organizations.

     Note: The default RAID group size is different for array LUNs and disks.
3. Plan the size of the LUNs that you need in your RAID groups.
   - To avoid a performance penalty, all array LUNs in a particular RAID group should be the same size.
   - The LUNs should be the same size in all RAID groups in the aggregate.
4. Ask the storage array administrator to create the number of LUNs of the size you need for the aggregate.
   The LUNs should be optimized for performance, according to the instructions in the storage array vendor documentation.

5. Create all the RAID groups in the aggregate at the same time.

   Note: Do not mix array LUNs from storage arrays with different characteristics in the same Data ONTAP RAID group.

   Note: If you create a new RAID group for an existing aggregate, be sure that the new RAID group is the same size as the other RAID groups in the aggregate, and that the array LUNs are the same size as the LUNs in the other RAID groups in the aggregate.

How Data ONTAP works with hot spare disks

A hot spare disk is a disk that is assigned to a storage system but is not in use by a RAID group. It does not yet hold data but is ready for use. If a disk failure occurs within a RAID group, Data ONTAP automatically assigns hot spare disks to RAID groups to replace the failed disks.

How many hot spares you should have

Having insufficient spares increases the risk of a disk failure with no available spare, resulting in a degraded RAID group. The number of hot spares you should have depends on the Data ONTAP disk type.

MSATA disks, or disks in a multi-disk carrier, should have four hot spares during steady state operation, and you should never allow the number of MSATA hot spares to dip below two.

For RAID groups composed of SSDs, you should have at least one spare disk.

For all other Data ONTAP disk types, you should have at least one matching or appropriate hot spare available for each kind of disk installed in your storage system. However, having two available hot spares for all disks provides the best protection against disk failure. Having at least two available hot spares provides the following benefits:

- When you have two or more hot spares for a data disk, Data ONTAP can put that disk into the maintenance center if needed.
  Data ONTAP uses the maintenance center to test suspect disks and take offline any disk that shows problems.
- Having two hot spares means that when a disk fails, you still have a spare available if another disk fails before you replace the first failed disk.

A single spare disk can serve as a hot spare for multiple RAID groups.

Related concepts

Spare requirements for multi-disk carrier disks on page 37

What disks can be used as hot spares

A disk must conform to certain criteria to be used as a hot spare for a particular data disk.

For a disk to be used as a hot spare for another disk, it must conform to the following criteria:

- It must be either an exact match for the disk it is replacing or an appropriate alternative.
- If SyncMirror is in use, the spare must be in the same pool as the disk it is replacing.
- The spare must be owned by the same system as the disk it is replacing.

What a matching spare is

A matching hot spare exactly matches several characteristics of a designated data disk. Understanding what a matching spare is, and how Data ONTAP selects spares, enables you to optimize your spares allocation for your environment.

A matching spare is a disk that exactly matches a data disk for all of the following criteria:

- Effective Data ONTAP disk type
  The effective disk type can be affected by the value of the `raid.mix.hdd.performance` and `raid.mix.hdd.capacity` options, which determine the disk types that are considered to be equivalent.
- Size
- Speed (RPM)
- Checksum type (BCS or AZCS)

Related concepts

How Data ONTAP reports drive types on page 18

What an appropriate hot spare is

If a disk fails and no hot spare disk that exactly matches the failed disk is available, Data ONTAP uses the best available spare. Understanding how Data ONTAP chooses an appropriate spare when there is no matching spare enables you to optimize your spare allocation for your environment.

Data ONTAP picks a non-matching hot spare based on the following criteria:

- If the available hot spares are not the correct size, Data ONTAP uses one that is the next size up, if there is one.
  The replacement disk is downsized to match the size of the disk it is replacing; the extra capacity is not available.
- If the available hot spares are not the correct speed, Data ONTAP uses one that is a different speed.
  Using drives with different speeds within the same aggregate is not optimal. Replacing a disk with a slower disk can cause performance degradation, and replacing a disk with a faster disk is not cost-effective.
- If the failed disk is part of a mirrored aggregate and there are no hot spares available in the correct pool, Data ONTAP uses a spare from the other pool.
  Using drives from the wrong pool is not optimal because you no longer have fault isolation for your SyncMirror configuration.

If no spare exists with an equivalent disk type or checksum type, the RAID group that contains the failed disk goes into degraded mode; Data ONTAP does not combine effective disk types or checksum types within a RAID group.

About degraded mode

When a disk fails, Data ONTAP can continue to serve data, but it must reconstruct the data from the failed disk using RAID parity. When this happens, the affected RAID group is said to be in *degraded mode*. The performance of a storage system with one or more RAID groups in degraded mode is decreased.

A RAID group goes into degraded mode in the following scenarios:

- A single disk fails in a RAID4 group.
  After the failed disk is reconstructed to a spare, the RAID group returns to normal mode.
- One or two disks fail in a RAID-DP group.
  If two disks have failed in a RAID-DP group, the RAID group goes into *double-degraded mode*.
- A disk is taken offline by Data ONTAP.
  After the offline disk is brought back online, the RAID group returns to normal mode.

Note: If another disk fails in a RAID-DP group in double-degraded mode or a RAID4 group in degraded mode, data loss could occur (unless the data is mirrored). For this reason, always minimize the amount of time a RAID group is in degraded mode by ensuring that appropriate hot spares are available.

About low spare warnings

By default, Data ONTAP issues warnings to the console and logs if you have fewer than one hot spare disk that matches the attributes of each disk in your storage system. You can change the threshold value for these warning messages by using the `raid.min_spare_count` option.

To make sure that you always have two hot spares for every disk (a best practice), you can set the `raid.min_spare_count` option to 2.

Setting the `raid.min_spare_count` option to 0 disables low spare warnings. You might want to do this if you do not have enough disks to provide hot spares (for example if your storage system does not support external disk shelves). You can disable the warnings only if the following requirements are met:

- Your system has 16 or fewer disks.
- You have no RAID groups that use RAID4.

  Note: You cannot create aggregates that use RAID4 protection while the `raid.min_spare_count` option is set to 0. If either of these requirements is no longer met after this option has been set to 0, the option is automatically set back to 1.

How Data ONTAP handles a failed disk with a hot spare

Using an available matching hot spare, Data ONTAP can use RAID to reconstruct the missing data from the failed disk onto the hot spare disk with no data service interruption.

If a disk fails and a matching or appropriate spare is available, Data ONTAP performs the following tasks:

- Replaces the failed disk with a hot spare disk.
  If RAID-DP is enabled and double-disk failure occurs in the RAID group, Data ONTAP replaces each failed disk with a separate spare disk.
- In the background, reconstructs the missing data onto the hot spare disk or disks.

  Note: During reconstruction, the system is in degraded mode, and file service might slow down.
- Logs the activity in the `/etc/messages` file.
- Sends an AutoSupport message.

Attention: Replace the failed disk or disks with new hot spare disks as soon as possible, so that hot spare disks are always available in the storage system.

Note: If the available spare disks are not the correct size, Data ONTAP chooses a disk of the next larger size and restricts its capacity to match the size of the disk it is replacing.

Related concepts

How Data ONTAP handles a failed disk that has no available hot spare on page 116

Related tasks

Removing a failed disk on page 43
Adding disks to a storage system on page 39

How Data ONTAP handles a failed disk that has no available hot spare

When a failed disk has no appropriate hot spare available, Data ONTAP puts the affected RAID group into degraded mode indefinitely and the storage system automatically shuts down within a specified time period.

If the maximum number of disks have failed in a RAID group (two for RAID-DP, one for RAID4), the storage system automatically shuts down in the period of time specified by the `raid.timeout` option. The default timeout value is 24 hours.

To ensure that you are aware of the situation, Data ONTAP sends an AutoSupport message whenever a disk fails. In addition, it logs a warning message in the `/etc/message` file once per hour after a disk fails.

Attention: If a disk fails and no hot spare disk is available, contact technical support.

Related concepts

About degraded mode on page 115
How Data ONTAP handles a failed disk with a hot spare on page 116

Considerations for changing the timeout RAID option

The `raid.timeout` option controls how long a storage system runs after a RAID group goes into degraded mode or the NVRAM battery malfunctions or loses power. You can change the value of this option, but you should understand the implications of doing so.

The purpose for the system shutdown is to avoid data loss, which can happen if an additional disk failure occurs in a RAID group that is already running in degraded mode, or if a stand-alone system encounters a catastrophic error and has to shut down without NVRAM. You can extend the number of hours the system operates in these conditions by increasing the value of this option (the default value is 24). You can even disable the shutdown by setting the option to zero, but the longer the system operates with one or both of these conditions, the greater the chance of incurring data loss.

How RAID-level disk scrubs verify data integrity

RAID-level scrubbing means checking the disk blocks of all disks in use in aggregates (or in a particular aggregate, plex, or RAID group) for media errors and parity consistency. If Data ONTAP finds media errors or inconsistencies, it uses RAID to reconstruct the data from other disks and rewrites the data.

RAID-level scrubs help improve data availability by uncovering and fixing media and checksum errors while the RAID group is in a normal state (for RAID-DP, RAID-level scrubs can also be performed when the RAID group has a single-disk failure).

RAID-level scrubs can be scheduled or run manually.

How you schedule automatic RAID-level scrubs

By default, Data ONTAP performs a weekly RAID-level scrub starting on Sunday at 1:00 a.m. for a duration of six hours. You can change the start time and duration of the weekly scrub, or add more automatic scrubs.

To schedule an automatic RAID-level scrub, you use the `raid.scrub.schedule` option.

To change the duration of automatic RAID-level scrubbing without changing the start time, you use the `raid.scrub.duration` option, specifying the number of minutes you want automatic RAID-level scrubs to run. If you set this option to `-1`, all automatic RAID-level scrubs run to completion.

Note: If you specify a duration using the `raid.scrub.schedule` option, that value overrides the value you specify with this option.

Scheduling example

The following command schedules two weekly RAID scrubs. The first scrub is for 240 minutes (four hours) every Tuesday starting at 2 a.m. The second scrub is for eight hours every Saturday starting at 10 p.m.

```
options raid.scrub.schedule 240m@tue@2,8h@sat@22
```

Verification example

The following command displays your current RAID-level automatic scrub schedule.

```
options raid.scrub.schedule
```

Reverting to the default schedule example

The following command reverts your automatic RAID-level scrub schedule to the default (Sunday at 1:00 a.m., for six hours):

```
options raid.scrub.schedule ""
```

Related tasks

Controlling the performance impact of RAID-level scrubbing on page 123

How you run a manual RAID-level scrub

You can manually run a RAID-level scrub on individual RAID groups, plexes, aggregates, or all aggregates using the `aggr scrub` command. You can also stop, suspend, and resume manual RAID-level scrubs.

If you try to run a RAID-level scrub on a RAID group that is not in a normal state (for example, a group that is reconstructing or degraded), the scrub returns errors and does not check that RAID group. You can run a RAID-level scrub on a RAID-DP group with one failed disk.

Scrubbing all aggregates

The following command starts a RAID-level scrub on all of the aggregates in the storage system:

```
aggr scrub start
```

Scrubbing a particular RAID group

The following command starts a RAID-level scrub on rg0 in plex1 of aggregate aggr2:

```
aggr scrub start aggr2/plex1/rg0
```

Stopping a manual RAID-level scrub

The following command stops a manual RAID-level scrub currently running on plex1 or aggr0:

```
aggr scrub stop aggr0/plex1
```

If you do not specify a name of an aggregate, plex, or RAID group, Data ONTAP stops all manual RAID-level scrubs. After you stop a scrub, it cannot be resumed.

Suspending a manual RAID-level scrub

The following command suspends a manual RAID-level scrub currently running on aggregate aggr3:

```
aggr scrub suspend aggr3
```

You can resume this scrub later by using the `aggr scrub resume` command.

Viewing RAID-level scrub status

The following command displays the status of all currently running RAID-level scrubs, along with the date and time when the last full scrub completed:

```
aggr scrub status -v
```

Customizing the size of your RAID groups

You can customize the size of your RAID groups based on your requirements for data availability, performance, and disk utilization.

About this task

For standard aggregates, you change the size of RAID groups on a per-aggregate basis. For Flash Pool aggregates, you can change the RAID group size for the SSD RAID groups and the HDD RAID groups independently. You cannot change the size of individual RAID groups.

The following list outlines some facts about changing the RAID group size:

- If you increase the RAID group size, more disks or array LUNs will be added to the most recently created RAID group until it reaches the new size.
- All other existing RAID groups in that aggregate remain the same size, unless you explicitly add disks to them.
- You cannot decrease the size of already created RAID groups.
- The new size applies to all subsequently created RAID groups in that aggregate (or, in the case of a Flash Pool, all subsequently created RAID groups for the affected RAID group type-- SSD or HDD).

Step

1. Use the appropriate command as outlined in the following table:

If you want to...	Enter the following command...
Change the RAID group size for the SSD RAID groups of a Flash Pool aggregate	**aggr options *aggr_name* cache_raidsize *size***
Change the size of any other RAID groups	**aggr options *aggr_name* raidsize *size***

Examples

The following command changes the maximum RAID group size of the aggregate n1_a4 to 20 disks or array LUNs:

```
aggr options n1_a4 raidsize 20
```

The following command changes the maximum RAID group size of the SSD cache RAID groups of the Flash Pool aggregate n1_cache_a2 to 24:

```
aggr options n1_cache_a2 cache_raidsize 24
```

Related concepts

How Data ONTAP uses RAID to protect your data and data availability on page 105
Considerations for sizing RAID groups for drives on page 111
Considerations for Data ONTAP RAID groups for array LUNs on page 112
How Data ONTAP RAID groups work on page 110

Related tasks

Increasing the size of an aggregate on page 147

Controlling the impact of RAID operations on system performance

You can reduce the impact of RAID operations on system performance by decreasing the speed of the RAID operations.

You can control the speed of the following RAID operations with RAID options:

- RAID data reconstruction
- Disk scrubbing
- Plex resynchronization
- Synchronous mirror verification

The speed that you select for each of these operations might affect the overall performance of the storage system. However, if the operation is already running at the maximum speed possible and it is fully utilizing one of the three system resources (the CPU, disks, or the disk-to-controller connection bandwidth), changing the speed of the operation has no effect on the performance of the operation or the storage system.

If the operation is not yet running, you can set a speed that minimally slows storage system network operations or a speed that severely slows storage system network operations. For each operation, use the following guidelines:

- If you want to reduce the performance impact on client access to the storage system, change the specific RAID option from medium to low. Doing so also causes the operation to slow down.
- If you want to speed up the operation, change the RAID option from medium to high. Doing so might decrease the performance of the storage system in response to client access.

Controlling the performance impact of RAID data reconstruction

Because RAID data reconstruction consumes CPU resources, increasing the speed of data reconstruction sometimes slows storage system network and disk operations. You can control the speed of data reconstruction with the `raid.reconstruc.perf_impact` option.

About this task

When RAID data reconstruction and plex resynchronization are running at the same time, Data ONTAP limits the combined resource utilization to the greater impact set by either operation. For example, if `raid.resync.perf_impact` is set to `medium` and `raid.reconstruct.perf_impact` is set to `low`, the resource utilization of both operations has a medium impact.

The setting for this option also controls the speed of Rapid RAID recovery.

Step

1. Enter the following command:

   options raid.reconstruct.perf_impact *impact*

   impact can be high, medium, or low.

   high means that the storage system uses most of the system resources available for RAID data reconstruction; this setting can heavily affect storage system performance, but reconstruction finishes sooner, reducing the time that the RAID group is in degraded mode.

   low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance. However, reconstruction takes longer to complete, increasing the time that the storage system is running in degraded mode.

   The default impact is medium.

Controlling the performance impact of RAID-level scrubbing

When Data ONTAP performs a RAID-level scrub, it checks the disk blocks of all disks on the storage system for media errors and parity consistency. You can control the impact this operation has on system performance with the raid.verify.perf_impact option.

About this task

When RAID-level scrubbing and mirror verification are running at the same time, Data ONTAP limits the combined resource utilization to the greater impact set by either operation. For example, if raid.verify.perf_impact is set to medium and raid.scrub.perf_impact is set to low, the resource utilization by both operations has a medium impact.

If there are times during the day when the load on your storage system is decreased, you can also limit the performance impact of the automatic RAID-level scrub by changing the start time or duration of the automatic scrub.

Step

1. Enter the following command:

   options raid.scrub.perf_impact *impact*

   impact can be high, medium, or low.

   high means that the storage system uses most of the system resources available for scrubbing; this setting can heavily affect storage system performance, but the scrub finishes sooner.

   low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance, but the scrub takes longer to complete.

   The default impact is low.

Related concepts

How you schedule automatic RAID-level scrubs on page 117
How Data ONTAP uses RAID to protect your data and data availability on page 105

Controlling the performance impact of plex resynchronization

Plex resynchronization ensures that both plexes of a mirrored aggregate are identical. You can control the performance impact of plex resynchronization by using the `raid.resync.perf_impact` option.

About this task

When plex resynchronization and RAID data reconstruction are running at the same time, Data ONTAP limits the combined resource utilization to the greatest impact set by either operation. For example, if `raid.resync.perf_impact` is set to `medium` and `raid.reconstruct.perf_impact` is set to `low`, the resource utilization by both operations has a medium impact.

You should set this option to the same value for both nodes in an HA configuration.

Step

1. Enter the following command:

   `options raid.resync.perf_impact` ***`impact`***

   `impact` can be `high`, `medium`, or `low`.

   `high` means that the storage system uses most of the system resources available for plex resynchronization; this setting can heavily affect storage system performance, but the resynchronization finishes sooner.

   `low` means that the storage system uses very little of the system resources; this setting lightly affects storage system performance, but the resynchronization takes longer to complete.

   The default impact is `medium`.

Controlling the performance impact of mirror verification

You use mirror verification to ensure that the two plexes of a synchronous mirrored aggregate are identical. You can control the speed of mirror verification, and its effect on system resources, by using the `raid.verify.perf_impact` option.

About this task

When mirror verification and RAID-level scrubbing are running at the same time, Data ONTAP limits the combined resource utilization to the greatest impact set by either operation. For example, if `raid.verify.perf_impact` is set to `medium` and `raid.scrub.perf_impact` is set to `low`, the resource utilization of both operations has a medium impact.

For more information about synchronous mirroring, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.

Step

1. Enter the following command:

   `options raid.verify.perf_impact` ***`impact`***

   `impact` can be `high`, `medium`, or `low`.

   `high` means that the storage system uses most of the system resources available for mirror verification; this setting can heavily affect storage system performance, but the mirror verification finishes sooner.

   `low` means that the storage system uses very little of the system resources; this setting lightly affects storage system performance, but the mirror verification takes longer to complete.

   The default impact is `low`.

How you use aggregates to provide storage to your volumes

To support the differing security, backup, performance, and data sharing needs of your users, you can group the physical data storage resources on your storage system into one or more aggregates. You can then design and configure these aggregates to provide the appropriate level of performance and redundancy.

Each aggregate has its own RAID configuration, plex structure, and set of assigned disks or array LUNs. The aggregate provides storage, based on its configuration, to its associated FlexVol volumes.

Aggregates have the following characteristics:

- They can be composed of disks or array LUNs.
- They can be mirrored or unmirrored.
- They can be in 64-bit or 32-bit format.
- If they are composed of disks, they can be single-tier (composed of only HDDs or only SSDs) or they can be Flash Pools, which include both of those storage types in two separate tiers.

For information about best practices for working with aggregates, see *Technical Report 3437: Storage Subsystem Resiliency Guide*.

Related concepts

Disk speeds supported by Data ONTAP on page 22

Related references

Storage limits on page 342

Related information

Technical Report 3437: Storage Subsystem Resiliency Guide

Introduction to 64-bit and 32-bit aggregate formats

Aggregates are either 64-bit or 32-bit format. 64-bit aggregates have much larger size limits than 32-bit aggregates. 64-bit and 32-bit aggregates can coexist on the same storage system.

32-bit aggregates have a maximum size of 16 TB; 64-bit aggregates' maximum size depends on the storage system model. For the maximum 64-bit aggregate size of your storage system model, see the *Hardware Universe* at *hwu.netapp.com*.

You can expand 32-bit aggregates to 64-bit aggregates by increasing their size beyond 16 TB. 64-bit aggregates, including aggregates that were previously expanded, cannot be converted to 32-bit aggregates.

You can see whether an aggregate is a 32-bit aggregate or a 64-bit aggregate by using the `aggr status` command.

Related references

Storage limits on page 342

Best practices for expanding a 32-bit aggregate to 64-bit

You should be aware of certain best practices before expanding an aggregate from 32-bit to 64-bit format.

Following these suggestions ensures a smooth expansion operation:

- If the aggregate you are adding storage to contains FlexCache volumes, destroy the FlexCache volumes before initiating the expansion and re-create them after the operation is complete.
- If you are expanding aggregates that contain volumes in a SnapMirror relationship, expand the aggregate containing the source volume first whenever possible.

  Otherwise, expand the source aggregate as soon as possible after expanding the destination aggregate.
- If you are creating a FlexClone volume from a SnapMirror destination volume and you are expanding the aggregates containing the source and destination volumes, expand both source and destination, and use a base Snapshot copy that was created after the source volume was expanded.
- When you add storage to any aggregate, add an entire RAID group at a time to keep the size of your RAID groups homogeneous.

For more information about expanding a 32-bit aggregate to 64-bit, see TR-3978, *In-Place Expansion of 32-bit Aggregates to 64-bit Overview and Best Practices*.

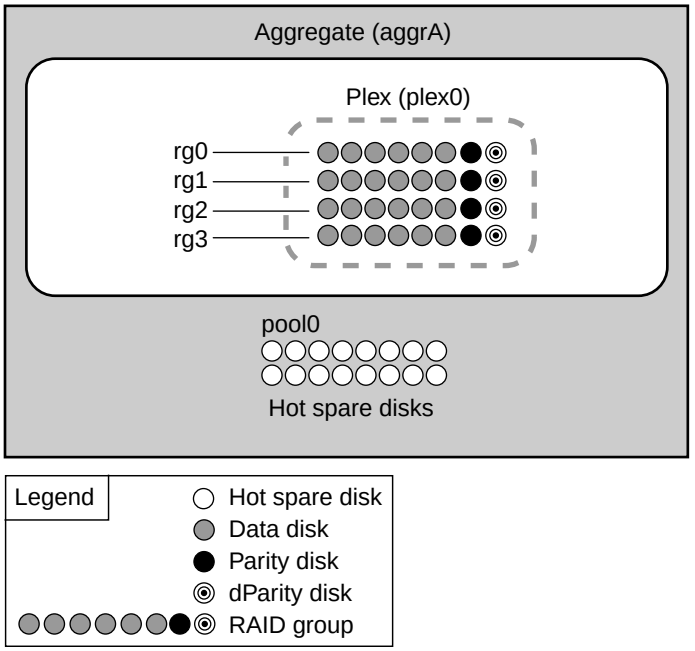
Related information

TR 3978: In-Place Expansion of 32-bit Aggregates to 64-bit Overview and Best Practices

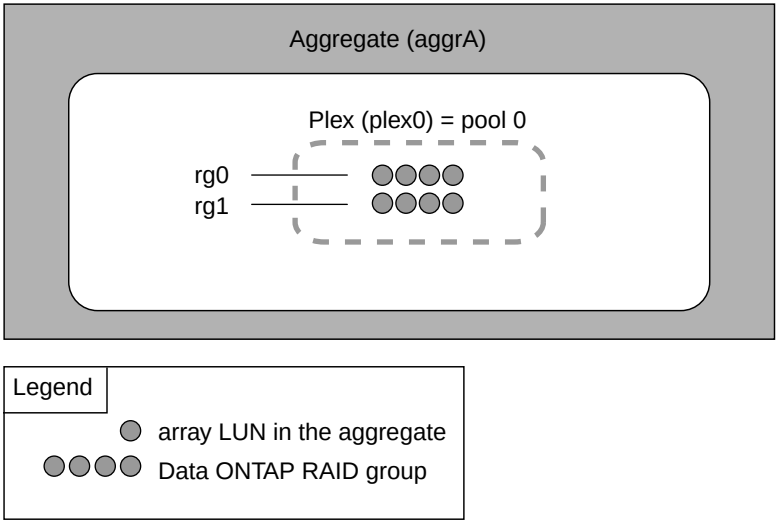
How unmirrored aggregates work

Unless you are using SyncMirror, all of your aggregates are unmirrored. Unmirrored aggregates have only one *plex* (copy of their data), which contains all of the RAID groups belonging to that aggregate.

The following diagram shows an unmirrored aggregate with disks, with its one plex.



The following diagram shows an unmirrored aggregate with array LUNs, with its one plex.

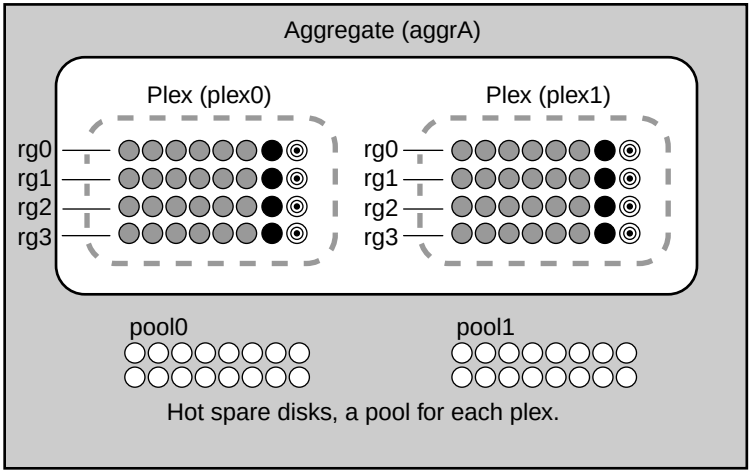


How mirrored aggregates work

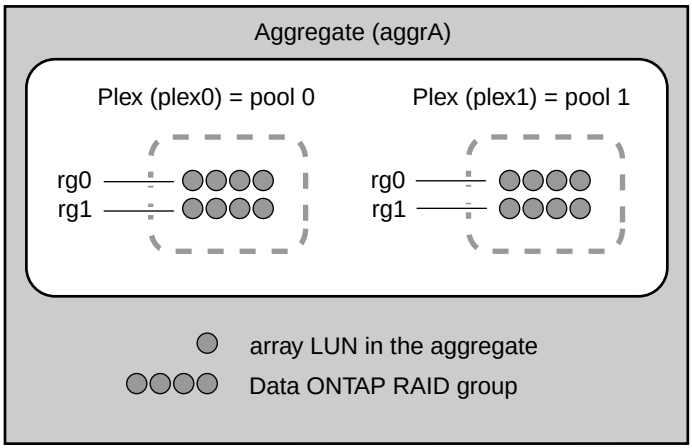
Mirrored aggregates have two *plexes* (copies of their data), which use the SyncMirror functionality to duplicate the data to provide redundancy.

When SyncMirror is enabled, all the disks or array LUNs are divided into two pools, and a copy of the plex is created. The plexes are physically separated (each plex has its own RAID groups and its own pool), and the plexes are updated simultaneously. This provides added protection against data loss if more disks fail than the RAID level of the aggregate protects against or there is a loss of connectivity, because the unaffected plex continues to serve data while you fix the cause of the failure. After the plex that had a problem is fixed, you can resynchronize the two plexes and reestablish the mirror relationship.

In the following diagram of a storage system using disks, SyncMirror is enabled and implemented, so Data ONTAP copies plex0 and automatically names the copy plex1. Plex0 and plex1 contain copies of one or more file systems. In this diagram, 32 disks were available prior to the SyncMirror relationship being initiated. After initiating SyncMirror, the spare disks are allocated to pool0 or pool1.



The following diagram shows a storage system using array LUNs with SyncMirror enabled and implemented.



How Flash Pool aggregates work

The Flash Pool aggregate technology enables you to add one or more RAID groups composed of SSDs to an aggregate that consists of RAID groups of HDDs.

The SSDs provide a high-performance cache for the active data set of the data volumes provisioned on the Flash Pool aggregate, which offloads I/O operations from the HDDs to the SSDs. For random workloads, this can increase the performance of the volumes associated with the aggregate by improving the response time and overall throughput for I/O-bound data access operations. (The performance increase is not seen for predominantly sequential workloads.)

The SSD cache does not contribute to the size of the aggregate as calculated against the maximum aggregate size. For example, even if an aggregate is at the maximum aggregate size, you can add an SSD RAID group to it. The SSDs *do* count toward the overall spindle limit.

The HDD RAID groups in a Flash Pool aggregate behave the same as HDD RAID groups in a standard aggregate, following the same rules for mixing disk types, sizes, speeds, and checksums.

The checksum type, RAID type, and RAID group size values can be configured for the SSD cache RAID groups and HDD RAID groups independently.

There is a platform-dependent maximum size for the SSD cache. For information about this limit for your platform, see the *Hardware Universe*.

Requirements for using Flash Pool aggregates

The Flash Pool aggregate technology has some configuration requirements that you should be aware of before planning to use it in your storage architecture.

Flash Pool aggregates cannot be used in the following configurations:

- 32-bit aggregates

- Aggregates composed of array LUNs
- Aggregates that use the ZCS checksum type
- SnapLock aggregates
- Traditional volumes
- A storage system that uses Storage Encryption

FlexShare is not supported for volumes associated with Flash Pool aggregates.

Read-only volumes, such as SnapMirror destinations, are not cached in the Flash Pool cache.

If you are not mirroring the Flash Pool and it is not used for a MetroCluster configuration, you should disable automatic aggregate Snapshot creation for the Flash Pool. You do this by using the `aggr options nosnap` command. For information about automatic aggregate Snapshot copy creation, see the *Data ONTAP System Administration Guide for 7-Mode*.

Flash Pool aggregates can be created from mirrored aggregates; however, the SSD configuration must be kept the same for both plexes.

For a list of the platforms that support Flash Pool aggregates, and minimum numbers of SSDs, see the *Hardware Universe*.

If you create a Flash Pool aggregate using an aggregate that was created using Data ONTAP 7.1 or earlier, the volumes associated with that Flash Pool aggregate will not support write caching.

For more information about the types of workloads and resource constraints that benefit from using Flash Pool aggregates, see *Technical Report 4070: NetApp Flash Pool Design and Implementation Guide*.

Related information

TR 4070: NetApp Flash Pool Design and Implementation Guide

How Flash Pool aggregates and Flash Cache compare

Both the Flash Pool aggregate technology and the family of Flash Cache modules (Flash Cache and Flash Cache 2) provide a high-performance cache to increase storage performance. However, there are differences between the two technologies that you should understand before choosing between them.

You can employ both technologies on the same system. However, data stored in volumes associated with a Flash Pool aggregate (or an SSD aggregate) is not cached by Flash Cache.

Criteria	Flash Pool aggregate	Flash Cache
Scope	A specific aggregate	All aggregates assigned to a controller
Caching types supported	Read and write	Read

Criteria	Flash Pool aggregate	Flash Cache
Cached data availability during and after takeover events	Cached data is available and unaffected by either planned or unplanned takeover events.	Cached data is not available during takeover events. After giveback for a planned takeover, previously cached data that is still valid is re-cached automatically.
PCIe slot on storage controller required?	No	Yes
Supported with array LUNs?	No	Yes
Supported with Storage Encryption?	No	Yes. Data in the cache is not encrypted.
Supported with SnapLock?	No	Yes

For more information about Flash Cache, see the *Data ONTAP System Administration Guide for 7-Mode*.

About read and write caching for Flash Pools

The Flash Pool aggregate technology provides both read caching and write caching for random I/O workloads. You can configure Flash Pool aggregate caching on the volume, but for most workloads, the default caching policies result in optimal performance.

Some volumes cannot be enabled for write caching. When you attempt to use an aggregate associated with one or more of these volumes as a Flash Pool aggregate, you must force the operation. In this case, writes to that volume would not be cached in the SSD cache, but otherwise the Flash Pool aggregate would function normally. You can get more information about why a volume cannot be enabled for write caching by using the `vol status -v` command.

For more information about read and write caching policies, see *Technical Report 4070: NetApp Flash Pool Design and Implementation Guide*.

Related information

TR 4070: NetApp Flash Pool Design and Implementation Guide

How Flash Pool aggregate cache capacity is calculated

Flash Pool aggregate cache capacity cannot exceed a platform-dependent limit for the system. Knowing the available cache capacity enables you to determine how many data SSDs you can add before reaching the limit.

The current cache capacity is the sum of the “used size” capacity of all of the data SSDs used in Flash Pool aggregates on the system. Parity SSDs do not count toward the limit. For systems using

SyncMirror, including MetroCluster configurations, only the Flash Pool aggregate cache of one plex counts toward the cache limit.

For systems in an HA configuration, the cache size limits apply to the HA configuration as a whole, and can be split arbitrarily between the two nodes, provided that the total limit for the HA configuration is not exceeded.

If Flash Cache modules are installed in a system, the available cache capacity for Flash Pool aggregate use is the Flash Pool aggregate cache capacity limit minus the sum of the Flash Cache module cache installed on the node. (In the unusual case where the size of the Flash Cache modules is not symmetrical between the two nodes in an HA configuration, the Flash Pool aggregate available cache capacity is decreased by the size of the larger Flash Cache module.)

For information about cache size limits, see the *Hardware Universe*.

Example calculation with Flash Cache modules

For an HA configuration composed of two storage controllers with a maximum cache capacity of 12 TB and 2 TB of Flash Cache installed on each node, the maximum Flash Pool aggregate cache capacity for the HA pair would be 12 TB minus 2 TB, or 10 TB.

Example calculation with asymmetrically sized Flash Cache modules

For an HA configuration composed of two storage controllers with a maximum cache capacity of 12 TB and 2 TB of Flash Cache installed on one node and 3 TB of Flash Cache installed on the other node, the maximum Flash Pool aggregate cache capacity for the HA pair would be 12 TB minus 3 TB, or 9 TB.

Restrictions for using aggregates composed of SSDs

Aggregates composed of SSDs have some restrictions on when they can be used.

You cannot use aggregates composed of SSDs with the following configurations or technologies:

- SnapLock
- Traditional volumes
- Storage Encryption
- FlexShare

How you can use disks with mixed speeds in the same aggregate

Whenever possible, you should use disks of the same speed in an aggregate. However, if needed, you can configure Data ONTAP to allow mixed speed aggregates based on the disk type.

To configure Data ONTAP to allow mixed speed aggregates, you use the following RAID options:

- `raid.mix.hdd.rpm.performance`
- `raid.mix.hdd.rpm.capacity`

When these options are set to `on`, Data ONTAP allows mixing speeds for the designated disk type. Performance disks are FC and SAS; capacity disk types are SATA, BSAS, FSAS, MSATA, and ATA.

By default, `raid.mix.hdd.rpm.performance` is set to `off`, and `raid.mix.hdd.rpm.capacity` is set to `on`.

Even if Data ONTAP is not configured to allow mixing speeds, you can still create aggregates out of disks with different speeds by using the `-f` option of the `aggr create` or `aggr add` commands.

Related concepts

Disk speeds supported by Data ONTAP on page 22

How to control disk selection from heterogeneous storage

When disks with different characteristics coexist on the same node, or when both disks and array LUNs are attached to the same node, the system has heterogeneous storage. When you create an aggregate from heterogeneous storage, you should take steps to ensure that Data ONTAP uses the disks you expect.

If your node has heterogeneous storage and you do not explicitly specify what type of disks to use, Data ONTAP uses the disk type (including array LUNs) with the highest number of available disks. When you create or add storage to an aggregate using heterogeneous storage, you should use one of the following methods to ensure that Data ONTAP selects the correct disks or disk types:

- Through disk attributes:
  - You can specify disk size by using the `@size` option.
    Disks within 20% of the specified size are selected.
  - You can specify disk speed by using the `-R` option.
  - You can specify disk type by using the `-T` option.
- Through an explicit disk list.
  You can list the names of specific disks you want to use.

- Through disk selection preview.
  You can use the `-n` option to identify which disks Data ONTAP selects automatically. If you are happy with the disks selected, you can proceed with automatic disk selection. Otherwise, you can use one of the previous methods to ensure that the correct disks are selected.

Note: For unplanned events such as disk failures, which cause Data ONTAP to add another disk to a RAID group automatically, the best way to ensure that Data ONTAP chooses the best disk for any RAID group on your system is to always have at least one spare (and preferably two) available to match all disk types and sizes in use in your system.

Rules for mixing HDD types in aggregates

You can mix disks from different loops or stacks within the same aggregate. Depending on the value of the `raid.mix.hdd.disktype` RAID options, you can mix certain types of HDDs within the same aggregate, but some disk type combinations are more desirable than others.

When the appropriate `raid.mix.hdd.disktype` option is set to `off`, single-tier aggregates and the HDD tier of Flash Pool aggregates can be composed of only one Data ONTAP disk type. This setting ensures that your aggregates are homogeneous, and requires that you provide sufficient spare disks for every disk type in use in your system.

The default value for the `raid.mix.hdd.disktype.performance` option is `off`, to prevent mixing SAS and FC-AL disks.

The default value for the `raid.mix.hdd.disktype.capacity` option is `on`. For this setting, the SATA, BSAS, FSAS, and ATA disk types are considered to be equivalent for the purposes of creating and adding to aggregates, and spare management.

To maximize aggregate performance and for easier storage administration, you should avoid mixing FC-AL-connected and SAS-connected disks in the same aggregate. This is because of the performance mismatch between FC-AL-connected disk shelves and SAS-connected disk shelves. When you mix these connection architectures in the same aggregate, the performance of the aggregate is limited by the presence of the FC-AL-connected disk shelves, even though some of the data is being served from the higher-performing SAS-connected disk shelves.

MSATA disks cannot be mixed with any other disk type in the same aggregate.

Disks using Storage Encryption have a Data ONTAP disk type of SAS. However, they cannot be mixed with any other disk type, including SAS disks that are not using Storage Encryption. If any disks on a storage system use Storage Encryption, all of the disks on the storage system (and its high-availability partner node) must use Storage Encryption.

Note: If you set a `raid.mix.hdd.disktype` option to `off` for a system that already contains aggregates with more than one type of HDD, those aggregates continue to function normally and accept both types of HDDs. However, no other aggregates composed of the specified disk type will accept mixed HDD types as long as that option is set to `off`.

For information about best practices for working with different types of disks, see *Technical Report 3437: Storage Best Practices and Resiliency Guide*.

Related concepts

How Data ONTAP reports drive types on page 18

Related information

TR 3437: Storage Best Practices and Resiliency Guide

Rules for mixing drive types in Flash Pool aggregates

By definition, Flash Pool aggregates contain more than one drive type. However, the HDD tier follows the same disk-type mixing rules as single-tier aggregates. For example, you cannot mix SAS and SATA disks in the same Flash Pool. The SSD cache can contain only SSDs.

How disk checksum types affect aggregate and spare management

There are two checksum types available for disks used by Data ONTAP: BCS (block) and AZCS (zoned). Understanding how the checksum types differ and how they impact storage management enables you to manage your storage more effectively.

Both checksum types provide the same resiliency capabilities; BCS optimizes data access speed and capacity for disks that use 520 byte sectors. AZCS provides enhanced storage utilization and capacity for disks that use 512 byte sectors (usually SATA disks, which emphasize capacity).

Aggregates have a checksum type, which is determined by the checksum type of the disks that compose the aggregate. The following configuration rules apply to aggregates, disks, and checksums:

- Checksum types cannot be combined within RAID groups.
  This means that you must consider checksum type when you provide hot spare disks.
- When you add storage to an aggregate, if it has a different checksum type than the storage in the RAID group to which it would normally be added, Data ONTAP creates a new RAID group.
- An aggregate can have RAID groups of both checksum types.
  These aggregates have a checksum type of `mixed`.
- For mirrored aggregates, both plexes must have the same checksum type.
- Disks of a different checksum type cannot be used to replace a failed disk.
- You cannot change the checksum type of a disk.

Rules for mixing storage in aggregates for V-Series systems

When planning for aggregates, you must consider the rules for mixing storage in aggregates. You cannot mix different storage types or array LUNs from different vendors or vendor families in the same aggregate.

Adding the following to the same aggregate is not supported:

- Array LUNs and disks
- Array LUNs with different checksum types
- Array LUNs from different drive types (for example, FC and SATA) or different speeds
- Array LUNs from different storage array vendors
- Array LUNs from different storage array model families

Note: Storage arrays in the same family share the same performance and failover characteristics. For example, members of the same family all perform active-active failover, or they all perform active-passive failover. More than one factor might be used to determine storage array families. For example, storage arrays with different architectures would be in different families even though other characteristics might be the same.

How the checksum type is determined for aggregates with array LUNs

Each Data ONTAP aggregate has a checksum type associated with it. The aggregate checksum type is determined by the checksum type of the array LUNs that are added to it.

The checksum type of an aggregate is determined by the checksum type of the first array LUN that is added to the aggregate. The checksum type applies to an entire aggregate (that is, to all volumes in the aggregate). Mixing array LUNs of different checksum types in an aggregate is not supported.

- An array LUN of type *block* must be used with block checksum type aggregates.
- An array LUN of type *zoned* must be used with advanced zoned checksum (AZCS or advanced_zoned) type aggregates.

  Note: Prior to Data ONTAP 8.1.1, zoned checksum array LUNs were used with ZCS (zoned) type aggregates. Starting in 8.1.1, any new aggregates created with zoned checksum array LUNs are AZCS aggregates. You can, however, add zoned checksum array LUNs to existing ZCS aggregates.

Before you add array LUNs to an aggregate, you must know the checksum type of the LUNs you want to add, for the following reasons:

- You cannot add array LUNs of different checksum types to the same aggregate.
- You cannot convert an aggregate from one checksum type to the other.

When you create an aggregate you can specify the number of array LUNs to be added, or you can specify the names of the LUNs to be added. If you want to specify a number of array LUNs to be added to the aggregate, the same number or more array LUNs of that checksum type must be available.

Understanding the root aggregate

The root aggregate contains the root volume, which contains special directories and configuration files that help you administer the storage system.

The following facts apply to the root aggregate:

- Starting with Data ONTAP 8.1, new systems are shipped with the root volume in a 64-bit root aggregate.
- By default, the storage system is set up to use a hard disk drive (HDD) aggregate for the root aggregate.

  When no HDDs are available, the system is set up to use a solid-state drive (SSD) aggregate for the root aggregate. If you want to change the root aggregate, you can choose either an HDD aggregate or an SSD aggregate to be the root aggregate (`aggr options` *`aggr_name`* `root`), provided that the corresponding type of disk drives is available on the system.
- A Flash Pool aggregate (an aggregate that contains both HDDs and SSDs) can be used as the root aggregate.

  Attention: If you revert or downgrade to Data ONTAP 8.1 or earlier with a Flash Pool aggregate configured as your root aggregate, your system will not boot.

Managing aggregates

You manage aggregates by creating them, increasing their size, setting their RAID level, and managing their state. In addition, you can destroy, undestroy and move aggregates.

You cannot reduce the number of disks in an aggregate by removing data disks. The only way to reduce the number of data disks in an aggregate is to copy the data and transfer it to a new aggregate that has fewer data disks.

Creating an aggregate

You create an aggregate to provide storage to one or more FlexVol volumes.

Before you begin

You should know what drives or array LUNs will be used in the new aggregate.

If you have multiple drive types in your system (heterogeneous storage), you should understand how you can ensure that the correct drive type is selected.

About this task

You can display a list of the available spares by using the `aggr status -s` command.

Aggregate names must conform to the following requirements:

- Begin with either a letter or an underscore (_).
- Contain only letters, digits, and underscores.
- Contain no more than 250 characters.

Steps

1. Create the aggregate by entering the following command:
   `aggr create` *`aggr_name`* `[-f] [-m] [-n] [-t {raid0 | raid4 | raid_dp}] [-r` *`raidsize`*`] [-T` *`disk-type`*`] -R` *`rpm`*`] [-L] [-p]` *`disk-list`*

   `aggr_name` is the name for the new aggregate.

   `-f` overrides the default behavior that does not permit drives in a plex to belong to different pools. This option also enables you to mix drives with different RPM speeds even if the appropriate `raid.rpm` option is not off.

   `-m` specifies the optional creation of a SyncMirror-replicated volume if you want to supplement RAID protection with SyncMirror protection.

   `-n` displays the results of the command but does not execute it. This is useful for displaying the drives that would be automatically selected prior to executing the command.

`-t {raid0 | raid4 | raid_dp}` specifies the level of RAID protection you want to provide for this aggregate. If no RAID level is specified for an aggregate composed of disks, the default value (`raid_dp`) is applied. `raid0` is used only for array LUNs.

`-r` *`raidsize`* is the maximum size of the RAID groups for this aggregate. If no size is specified, the default is used.

`-T` *`disk-type`* specifies the Data ONTAP drive type. This option is needed when creating aggregates on systems that have mixed drive types or both drives and array LUNs.

`-R` *`rpm`* specifies the type of disk to use based on its speed. Valid values for *`rpm`* include `5400`, `7200`, `10000`, and `15000`.

`-L` creates a SnapLock aggregate. For more information about the SnapLock feature, see the na_aggr(1) man page or the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.

`-p` specifies the pool from which the drives are selected.

`disk-list` is one of the following values:

- *`ndisks`*[`@`*`disk-size`*]
  `ndisks` is the number of drives to use.
  `disk-size` is the drive size to use, in gigabytes.
- `-d` *`disk_name1 disk_name2... disk_nameN`*
  `disk_name1`, *`disk_name2`*, and *`disk_nameN`* are drive IDs of available drives; use a space to separate drive IDs.

2. Verify the RAID group and drives of your new aggregate by entering the following command:

   `aggr status -r` ***`aggr_name`***

Examples

The following command creates a 64-bit aggregate called newfastaggr, with 20 drives, the default RAID group size, and all drives with 15K RPM:

```
aggr create newfastaggr -R 15000 20
```

The following command creates a 64-bit aggregate called newFCALaggr.

```
aggr create newFCALaggr -T FCAL 15
```

Related concepts

Considerations for sizing RAID groups for drives on page 111
Considerations for Data ONTAP RAID groups for array LUNs on page 112
Protection provided by RAID and SyncMirror on page 107
How you use aggregates to provide storage to your volumes on page 126
How Data ONTAP uses RAID to protect your data and data availability on page 105

How you can use disks with mixed speeds in the same aggregate on page 134

Related references

Storage limits on page 342

Creating a Flash Pool aggregate

You create a Flash Pool aggregate by enabling the feature on an existing 64-bit aggregate composed of HDD RAID groups, and then adding one or more SSD RAID groups to that aggregate. This results in two tiers for that aggregate: an SSD tier and an HDD tier.

Before you begin

- You must have identified a valid 64-bit aggregate composed of HDDs to convert to a Flash Pool aggregate.
- You must have determined write-caching eligibility of the volumes associated with the aggregate, and completed any required steps to resolve eligibility issues.
- You must have determined the SSDs you will be adding, and these SSDs must be owned by the node on which you are creating the Flash Pool aggregate.
- You must have determined the checksum types of both the SSDs you are adding and the HDDs already in the aggregate.
- You must have determined the number of SSDs you are adding and the optimal RAID group size for the SSD RAID groups.
  Fewer RAID groups in the SSD cache reduces the number of parity disks required.
- You must have determined the RAID level you want to use for the SSD tier.
- You must have familiarized yourself with the configuration requirements for Flash Pools.

About this task

After you add an SSD cache to an aggregate to create a Flash Pool aggregate, you cannot remove the SSD cache to convert the aggregate back to its original configuration.

You can change the RAID group size of the SSD cache, but you cannot make this change until after SSDs have been added. Once drives have been added to a RAID group, they cannot be removed. If you know that you want to use a different RAID group size than the default SSD RAID group size, you can add three SSDs at first. Then, after you update the RAID group size, you can add the rest of the SSDs.

By default, the RAID level of the SSD cache is the same as the RAID level of the HDD RAID groups. You can override this default selection by specifying the `-t` option when you add the first SSD RAID groups. Although the SSD cache is providing caching for the HDD tier, the SSD RAID groups are integral to the health of the aggregate as a whole. An SSD RAID group that experiences a failure that exceeds the RAID protection capability of the RAID level in use takes the aggregate offline. For this reason, it is a best practice to keep the RAID level of the SSD cache the same as that of the HDD RAID groups.

There are platform- and workload-specific best practices for Flash Pool aggregate SSD cache size and configuration. For information about these best practices, see *Technical Report 4070: NetApp Flash Pool Design and Implementation Guide*.

Steps

1. Mark the aggregate as hybrid-enabled:

   aggr options *aggr_name* hybrid_enabled true

   If this step does not succeed, determine write-caching eligibility for the target aggregate.

2. Add the SSDs to the aggregate by using the `aggr add` command.

   You can specify the SSDs by ID or by using the `disk_type` and `ndisks` parameters. You do not need to specify a new RAID group; Data ONTAP automatically puts the SSDs into their own RAID group.

   If you plan to change the RAID group size for the SSD tier, you should add only three SSDs in this step.

   If the HDDs and the SSDs do not have the same checksum type, or if the aggregate is a mixed-checksum aggregate, then you must use the `-c` parameter to specify the checksum type of the disks you are adding to the aggregate.

   You can specify a different RAID type for the SSD tier by using the `-t` option.

3. If you want a different RAID group size for the SSD tier than for the HDD tier, change the SSD RAID group size:

   aggr options *aggr_name* cache_raid_group_size *size*

4. If you did not add all of the required SSDs in the previous step, add the rest of the SSDs by using the `aggr add` command again.

Related concepts

How Flash Pool aggregates work on page 130
Requirements for using Flash Pool aggregates on page 130

Related tasks

Determining and enabling volume write-caching eligibility on page 142

Related information

TR 4070: NetApp Flash Pool Design and Implementation Guide

Determining and enabling volume write-caching eligibility

Understanding whether the FlexVol volumes associated with an aggregate are eligible for write caching can help you ensure that the volumes with high performance requirements can get the

maximum performance improvement from having their associated aggregate converted to a Flash Pool aggregate.

About this task

You can use this procedure to determine write caching eligibility to help you decide which aggregates are good candidates to become Flash Pool aggregates. You do not need any SSDs to complete this procedure.

Flash Pool aggregates employ two types of caching: *read caching* and *write caching*. Read caching is available for all volumes. Write caching is usually available for most volumes, but might be disabled due to the following reasons:

- The associated aggregate was created using Data ONTAP 7.1 or earlier
  The error message for this condition refers to an `incompatible file system version`.
- An internal ID collision
  The error message for this condition says that `one or more volumes in the aggregate is not eligible for write caching`.

If the aggregate you want to convert to a Flash Pool aggregate was created using Data ONTAP 7.1 or earlier and you want write caching to be enabled for its associated volumes, you must copy each FlexVol volume associated with the aggregate to a new aggregate and convert the new aggregate to a Flash Pool aggregate. If some volumes are ineligible due to an ID collision, you can resolve the ID collision by moving the ineligible volumes to another aggregate (and back again if needed) or simply proceed with the conversion, using the force option. In this case, the ineligible volumes do not benefit from the Flash Pool aggregate cache for write operations.

Steps

1. Attempt to enable the Flash Pool aggregate capability on the aggregate:

   `aggr options` *`aggr_name`* `hybrid_enabled true`

2. Take the applicable action based on the result of Step 1:

If...	Then...
The Flash Pool aggregate capability is successfully enabled	Disable the Flash Pool aggregate capability again: **`aggr options` *`aggr_name`* `hybrid_enabled false`** You have completed this task. All of the volumes associated with the aggregate are eligible for write caching.
The command displays an error message telling you that the aggregate cannot be converted to a Flash Pool aggregate	Go to the next step.

3. Take the applicable action based on why the volumes are ineligible and your requirements for those volumes:

If the volumes are ineligible because...	Then...
The aggregate was created using Data ONTAP 7.1 or earlier, and you do not need to enable write caching for its associated volumes	You can convert the aggregate by using the force option.
The aggregate was created using Data ONTAP 7.1 or earlier, and you want to enable write caching for its associated volumes	Copy all of the volumes associated with this aggregate to a new aggregate, and then repeat this procedure on the new aggregate.
There are ID collisions for some volumes in the aggregate, and you do not need to enable write caching for those volumes	You can convert the aggregate by using the force option.
There are ID collisions for some volumes in the aggregate, and you need to enable write caching for those volumes	Complete the following steps: **a.** Determine which volumes have ID collisions by running the `vol status -v` command for each volume, and noting the ones that do not show `read-write` eligibility. **b.** Copy or move all but one of each set of volumes with the same conflicting ID to a different aggregate, and then move them back until no more volumes show an ID conflict. If the volumes are accessed using only SAN protocols, you can use DataMotion for Volumes to move the volumes nondisruptively. For more information, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

Example with ID collision

The following example shows the system output when volumes within the aggregate have ID collisions:

```
sys1> aggr options aggr1 hybrid_enabled on
aggr options: Failed to set caching policies on Aggregate "aggr1"
because one or more volumes
in the aggregate is not eligible for write caching.

sys1> vol status -v vol8
...........

                                    svo_allow_rman=off,
svo_reject_errors=off,
                                    no_i2p=off,
fractional_reserve=100, extent=off,
                                    try_first=volume_grow,
read_realloc=off,
                                    snapshot_clone_dependency=off,
                                    no_delete_log=off,
```

```
dlog_hole_reserve=off,
                                    nbu_archival_snap=off
                Volume has clones: vol8_clone
                         Volume UUID: 4a7afb51-47cd-11e1-
bebd-123478563412
                Containing aggregate: 'aggr1'
                        Volinfo mode: 7-mode
                Volume Instance UUID: 4a7afb94-47cd-11e1-
bebd-123478563412
              Volume Provenance UUID: 4a7afb94-47cd-11e1-
bebd-123478563412

                Plex /aggr1/plex0: online, normal, active
                    RAID group /aggr1/plex0/rg0: normal, block
checksums
                    RAID group /aggr1/plex0/rg1: normal, block
checksums

        Snapshot autodelete settings for vol8:
                                        state=off
                                        commitment=try
                                        trigger=volume
                                        target_free_space=20%
                                        delete_order=oldest_first
                                        defer_delete=user_created
                                        prefix=(not specified)
                                        destroy_list=none
        Volume autosize settings:
                                state=off
        Hybrid Cache:
                Eligibility=read
                Write caching ineligibility reason=ID
Collision(25443)
```

Changing the RAID type of RAID groups in a Flash Pool aggregate

The SSD cache of a Flash Pool aggregate can have a different RAID type than the HDD RAID groups. You can change the RAID type of the SSD cache or HDD RAID groups independently of one another. All of the HDD RAID groups must have the same RAID type.

About this task

If the SSD cache RAID group goes into a failed state, the Flash Pool aggregate goes offline, just as it would if an HDD RAID group goes into a failed state. For this reason, you should use RAID-DP as the RAID type for the SSD cache whenever possible, and adhere to good hot spare practices for the SSD cache.

If the SSD cache has a different RAID type than the HDD RAID groups, the Flash Pool aggregate is considered to have a mixed RAID type, displayed as `mixed_raid_type` for the aggregate.

Steps

1. Change the RAID type of the SSD cache or HDD RAID groups of the Flash Pool aggregate:

   aggr options *aggr_name* -T *disk-type*

   To change the RAID type of the SSD cache, use **-T SSD**. To change the RAID type of the HDD RAID groups, specify any disk type included in the HDD RAID groups.

2. Verify the RAID groups in your Flash Pool aggregate:

   aggr status *aggr_name* -r

   You also can use the `aggr status -r` command to obtain more details about the RAID types of the HDD RAID groups and SSD cache of the Flash Pool aggregate.

Example

In this example, the HDD RAID groups and SSD cache of a Flash Pool aggregate named "test" initially have a RAID type of RAID4. The following command changes the RAID type of the SSD cache to RAID-DP, and converts the Flash Pool aggregate to the mixed RAID type:

aggr options test raidtype raid_dp -T SSD

The output from the `aggr status -r` command shows that the aggregate has a mixed RAID type, the HDD RAID groups have a RAID type of RAID4, and the SSD cache has a RAID type of RAID-DP.

```
aggr status -r test

Aggregate test (online, mixed_raid_type, hybrid) (block checksums)
  Plex /test/plex0 (online, normal, active)
    RAID group /test/plex0/rg0 (normal, block checksums, raid4)

RAID Disk Device   HA SHELF BAY CHAN Pool Type RPM  Used (MB/blks)    Phys (MB/blks)
--------- ------   -- ----- --- ---- ---- ---- ---  --------------    --------------
parity    6b.43.1  6b  43   1   SA:B -    BSAS 7200 847555/1735794176 847884/1736466816
data      6b.43.2  6b  43   2   SA:A -    BSAS 7200 847555/1735794176 847884/1736466816
data      6b.43.3  6b  43   3   SA:B -    BSAS 7200 847555/1735794176 847884/1736466816
data      6b.43.6  6b  43   6   SA:A -    BSAS 7200 847555/1735794176 847884/1736466816
data      6b.43.9  6b  43   9   SA:B -    BSAS 7200 847555/1735794176 847884/1736466816
data      6b.43.10 6b  43  10   SA:A -    BSAS 7200 847555/1735794176 847884/1736466816

    RAID group /test/plex0/rg1 (normal, block checksums, raid_dp)

RAID Disk Device   HA SHELF BAY CHAN Pool Type RPM  Used (MB/blks)    Phys (MB/blks)
--------- ------   -- ----- --- ---- ---- ---- ---  --------------    --------------
dparity   6b.43.15 6b  43   15  SA:B -    SSD  N/A  84574/173208064   84796/173663616
parity    6b.43.4  6b  43   4   SA:A -    SSD  N/A  84574/173208064   84796/173663616
data      6b.43.13 6b  43   13  SA:B -    SSD  N/A  84574/173208064   84796/173663616
data      6b.43.14 6b  43   14  SA:A -    SSD  N/A  84574/173208064   84796/173663616
```

Increasing the size of an aggregate

You can add disks or array LUNs to an aggregate so that it can provide more storage to its associated volumes. If you need to add enough storage to a 32-bit aggregate to increase its size beyond 16 TB, you can do so; this operation expands the aggregate to 64-bit format.

Before you begin

You must understand the following concepts:

- The requirement to add disks or array LUNs owned by the same system and pool
- For aggregates composed of disks:
  - Benefits of keeping your RAID groups homogeneous for disk size and speed.
  - Which types of disks can be used together.
  - Checksum rules when disks of more than one checksum type are in use.
  - How to ensure that the correct disks are added to the aggregate (the disk addition operation cannot be undone).
  - How to add disks to aggregates from heterogeneous storage.
  - The minimum number of disks to add for best performance.
  - The number of hot spares you need to provide for protection against disk failures.
  - Requirements for adding disks from multi-disk carrier disk shelves
  - The requirement to add storage to both plexes of a mirrored aggregate at the same time to ensure that the plexes are the same size and contain the same disk types

About this task

When you add HDDs to an aggregate, you should add a complete RAID group. For information about adding SSDs to a Flash Pool aggregate, see *Technical Report 4070: NetApp Flash Pool Design and Implementation Guide*.

Steps

1. Verify that appropriate spare disks or array LUNs are available for you to add by entering the following command:

   aggr status -s

   For disks, make sure that enough of the spares listed are of the correct type, size, speed, and checksum type for the target RAID group in the aggregate to which you are adding the disks.

2. Add the disks or array LUNs by entering the following command:

   aggr add *aggr_name* [-T] [-c] [-f] [-n] [-g {*raid_group_name* | new | all}] *disk_list*

If you are adding disks with a different checksum than the aggregate, as when creating a Flash Pool aggregate, or if you are adding disks to a mixed checksum aggregate, you must either specify the disks to be added with a disk list or use the `-c` option to specify the checksum.

If you are adding disks to a Flash Pool aggregate, you must either specify the disks to be added with a disk list or use the `-T` option to specify the disk type.

`-f` enables you to add disks or array LUNs from a different pool or, for disks, of a different speed.

`-n` displays the results of the command but does not execute it. This is useful for displaying the disks or array LUNs that Data ONTAP would automatically select. You can then decide whether to accept the selection provided by Data ONTAP or to add different disks or array LUNs.

If you specify the `-g` option, the storage is added to the RAID group you specify. *`raid_group_name`* is the name that Data ONTAP gave to the group—for example, rg0. If you are adding SSDs to the SSD tier of a Flash Pool aggregate, you do not need to specify the RAID group name; the SSD RAID group is selected by default based on the type of the disks you are adding. You should always fill up the existing SSD RAID group before creating a new one.

To add the storage to a new RAID group, use the `new` value instead of the group name.

To fill all existing RAID groups to the current value of the `raidsize` option before creating a new RAID group, use the `all` keyword instead of the group name. When you specify the `all` keyword, Data ONTAP adds disks to an existing RAID group until it reaches the maximum size, and then moves on to the next existing RAID group. If there are more disks to be added and all existing RAID groups are full, Data ONTAP creates a new RAID group.

`disk_list` is one of the following parameters:

- *`ndisks`*[`@`*`disk_size`*]
- `-d` *`disk1`* [*`disk2`*...]

The *`disk_size`* parameter is the approximate size of the disk in GB. Disks that are within approximately 20 percent of the specified size are selected.

3. If the previous step was unsuccessful because you are adding disks to a 32-bit aggregate and the additional disks would cause its size to exceed 16 TB, complete the following steps to expand the aggregate to 64-bit:

a) Repeat the `aggr add` command you entered before, with the `-64bit-upgrade normal` parameter added.

Example

For example, if you entered the **`aggr add 10@600G`** command, you would enter the following command:

`aggr add -64bit-upgrade normal 10@600G`

Data ONTAP checks each volume associated with the aggregate to ensure that it has enough free space to be expanded to 64-bit. If all of the volumes have enough free space, the disks are

added and the aggregate is expanded to the 64-bit format. If any of the volumes are too full to be expanded, the command fails.

b) If the previous command failed, run the command again, replacing the `-64-bit-upgrade normal` parameter with the `-64-bit-upgrade check` parameter. Follow the instructions in the output of that command.

c) If you had to add more space to any volume, repeat the `aggr add` command again, this time with the `-64bit-upgrade normal` parameter.

d) If you want to ensure that the disk usage quota accounting for this aggregate is exactly correct, reinitialize quotas on all of its volumes.

   If you do not reinitialize quotas, quotas on volumes associated with this aggregate will remain active, but the disk usage accounting will be slightly lower than the actual usage until the next time quotas are reinitialized.

Examples

The following command adds four 300-GB disks to the aggr1 aggregate:

```
aggr add aggr1 4@300G
```

The following command adds the disks 5a.17, 5a.19, 5a.20, and 5a.26 to the rg1 RAID group of the aggr2 aggregate:

```
aggr add aggr2 -g rg1 -d 5a.17 5a.19 5a.20 5a.26
```

The following command adds four disks to each plex of a mirrored aggregate aggr_mir:

```
aggr add aggr_mir -d 5a.18 5a.19 5a.20 5a.21 -d 8b.14 8b.15 8b.16 8b.
17
```

After you finish

After you add storage to an aggregate, you should run a full reallocation job on each FlexVol volume contained in that aggregate. For information about reallocation, see the *Data ONTAP System Administration Guide for 7-Mode*.

Related concepts

What happens when you add storage to an aggregate on page 150
How to control disk selection from heterogeneous storage on page 134
How you can use disks with mixed speeds in the same aggregate on page 134
How many hot spares you should have on page 113
Best practices for expanding a 32-bit aggregate to 64-bit on page 127
How you use aggregates to provide storage to your volumes on page 126
How Flash Pool aggregates work on page 130

Related references

Storage limits on page 342

Related information

TR 4070: NetApp Flash Pool Design and Implementation Guide

What happens when you add storage to an aggregate

By default, Data ONTAP adds new drives or array LUNs to the most recently created RAID group until it reaches its maximum size. Then Data ONTAP creates a new RAID group. Alternatively, you can specify a RAID group that you want to add storage to.

When you create an aggregate or add storage to an aggregate, Data ONTAP creates new RAID groups as each RAID group is filled with its maximum number of drives or array LUNs. The last RAID group formed might contain fewer drives or array LUNs than the maximum RAID group size for the aggregate. In that case, any storage added to the aggregate is added to the last RAID group until the specified RAID group size is reached.

If you increase the RAID group size for an aggregate, new drives or array LUNs are added only to the most recently created RAID group; the previously created RAID groups remain at their current size unless you explicitly add storage to them.

If you add a drive to a RAID group that is larger than the parity drives, the drive is capacity-limited to the size of the smallest parity drive.

Note: You are advised to keep your RAID groups homogeneous when possible. If needed, you can replace a mismatched drive with a more suitable drive later.

Related concepts

How to control disk selection from heterogeneous storage on page 134

How you can use disks with mixed speeds in the same aggregate on page 134

How mirrored aggregates work on page 129

Related tasks

Replacing disks that are currently being used in an aggregate on page 41

Forcibly adding disks to aggregates

You might want to override some of the restrictions on what disks can be added to an aggregate if you do not have disks of the right speed or enough disks in the correct pool. You can do so by using the `aggr add -f` command.

About this task

Forcibly adding disks can be useful in the following situations:

- You need to add disks from two different spare disk pools to a mirrored aggregate.

  Note: Using disks from the wrong pool in a mirrored aggregate removes an important fault isolation property of the SyncMirror functionality. You should do so only when absolutely necessary, and you should return to a supported configuration as soon as possible.
- You need to add disks of a different speed than that of existing disks in the aggregate.

Step

1. Add the disks by entering the following command:

   aggr add *aggr_name* -f [-n] [-g {*raid_group_name* | new | all}] *disk_list*

Taking an aggregate offline

You use the `aggr offline` command to take an aggregate offline to perform maintenance on the aggregate, move it, or destroy it.

Steps

1. If the aggregate you want to take offline contains FlexVol volumes, boot into maintenance mode.

   Note: This step is not necessary for traditional volumes.
2. Enter the following command:

   aggr offline *aggr_name*
3. If you previously booted into maintenance mode, return to normal mode.

Result

The aggregate is now offline. You cannot access any data in the aggregate's volumes.

Related tasks

Taking a volume offline on page 179

Bringing an aggregate online

After you restrict an aggregate or take it offline, you can use the `aggr online` command to make it available to the storage system again by bringing it back online.

Step

1. Enter the following command:

   aggr online *aggr_name*

If the aggregate is inconsistent, the command prompts you for confirmation.

Attention: If you bring an inconsistent aggregate online, it might suffer further file system corruption. If you have an inconsistent aggregate, contact technical support.

Result

The aggregate is online and available for use.

Related tasks

Bringing a volume online on page 180

Putting an aggregate into restricted state

Putting an aggregate into the restricted state prevents data access to volumes associated with the aggregate but does not take the aggregate completely offline. You must put an aggregate into the restricted state to make it the target of an aggregate copy or SnapMirror operation.

Steps

1. If the aggregate you want to restrict contains FlexVol volumes, boot into maintenance mode.
2. Enter the following command:

   `aggr restrict aggr_name`
3. If you previously booted into maintenance mode, return to normal mode.

Result

The aggregate is now restricted. Data in the aggregate's volumes is unavailable to clients.

Related tasks

Putting a volume into restricted state on page 179

Changing the RAID level of an aggregate

When you change an aggregate’s RAID level (from RAID4 to RAID-DP, for example), Data ONTAP reconfigures existing RAID groups to the new level and applies the new level to subsequently created RAID groups.

You cannot change the Data ONTAP RAID level of aggregates containing array LUNs. Aggregates that contain array LUNs must have a Data ONTAP RAID level of RAID0. RAID protection for aggregates that contain array LUNs is provided by the storage array.

Changing an aggregate's RAID level from RAID4 to RAID-DP

You can change an existing aggregate's RAID level from RAID4 to RAID-DP if you want the increased protection that RAID-DP provides.

Steps

1. Determine the number of RAID groups and the size of their parity disks in the aggregate in question by entering the following command:

   aggr status *aggr_name* -r

2. List the available hot spares on your system by entering the following command:

   aggr status -s

3. Make sure that at least one, and preferably two hot spare disks exist for each RAID group listed. If necessary, add additional hot spare disks.

4. Enter the following command:

   aggr options *aggr_name* raidtype raid_dp

Result

When you change the RAID level of an aggregate from RAID4 to RAID-DP, Data ONTAP makes the following changes:

- Adds an additional disk to each existing RAID group from the storage system's hot spare disks; assigns the new disk the dParity disk function for the RAID-DP group. A reconstruction begins for each RAID group to populate the dParity disk.
- Changes the `raidsize` option for the aggregate to the appropriate RAID-DP default value.

  Note: You can change the `raidsize` option after the RAID level change is complete.

After you finish

You can verify the new RAID level by using the `aggr options` command.

Related concepts

How you use aggregates to provide storage to your volumes on page 126
How Data ONTAP works with hot spare disks on page 113

Related tasks

Customizing the size of your RAID groups on page 120

Related references

Storage limits on page 342

Changing an aggregate's RAID level from RAID-DP to RAID4

When you change an aggregate's RAID level from RAID-DP to RAID4, the extra parity disks are converted to spares. In addition, the `raidsize` option is changed.

Step

1. Enter the following command:

   `aggr options` *`aggr_name`* `raidtype raid4`

Result

When you change the RAID level of an aggregate from RAID4 to RAID-DP, Data ONTAP makes the following changes:

- In each of the aggregate's existing RAID groups, the RAID-DP second parity disk (dParity) is removed and designated as a hot spare, thus reducing each RAID group's size by one parity disk.
- Data ONTAP changes the setting for the aggregate's `raidsize` option to the size of the largest RAID group in the aggregate, except in the following situations:
  - If the aggregate's largest RAID group is larger than the maximum RAID4 group size, then the aggregate's `raidsize` option is set to the maximum.
  - If the aggregate's largest RAID group is smaller than the default RAID4 group size, then the aggregate's `raidsize` option is set to the default group size.
  - If the aggregate's `raidsize` option is already below the default value for RAID4, it is reduced by 1.

After you finish

You can verify the new RAID level by using the `aggr options` command.

Related concepts

How you use aggregates to provide storage to your volumes on page 126

Related tasks

Customizing the size of your RAID groups on page 120

Related references

Storage limits on page 342

Destroying an aggregate

You destroy an aggregate when you no longer need the data in that aggregate or when you have copied the content of the aggregate to another location.

Before you begin

Before you can destroy an aggregate, you must have destroyed all of the FlexVol volumes associated with that aggregate.

About this task

When you destroy an aggregate, Data ONTAP converts its parity disks and its data disks back into hot spares. You can then use the spares in other aggregates and other storage systems.

Attention: If you destroy an aggregate, the data in the aggregate is no longer accessible by normal access methods, unless you restore it before any of its disks are zeroed or reused in another aggregate.

Note: If you want to make the data in the aggregate inaccessible by any means, you can sanitize its disks.

Note: You cannot destroy a SnapLock Compliance aggregate until the retention periods for all data contained in it have expired. For more information about the SnapLock functionality, see the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.

Steps

1. Take the aggregate offline by entering the following command:

   `aggr offline` ***`aggr_name`***

2. Destroy the aggregate by entering the following command:

   `aggr destroy` ***`aggr_name`***

   The following message is displayed:

   ```
   Are you sure you want to destroy this aggregate ?
   ```

3. Enter the following command to confirm that you want to destroy the aggregate:

   `y`

   The following message is displayed:
   `Aggregate '`*`aggr_name`*`' destroyed.`

Related concepts

How disk sanitization works on page 26

Restoring a destroyed aggregate

If you previously destroyed an aggregate and have changed your mind, you can restore the aggregate if the data is still intact and the aggregate was not SnapLock-compliant.

Before you begin

You must know the name of the aggregate you want to restore, because there is no Data ONTAP command available to display destroyed aggregates, and they do not appear in the management tools.

Steps

1. Restore the aggregate by entering the following command:

   aggr undestroy *aggr_name*

   Example

   aggr undestroy aggr1

   The following message is displayed:

   ```
   To proceed with aggr undestroy, select one of the following options [1]
   abandon the command [2] undestroy aggregate aggr1 ID:
   0xf8737c0-11d9c001-a000d5a3-bb320198 Selection (1-2)?
   ```

   If you select **2**, a message is displayed with a date and time stamp for each disk that is restored to the aggregate.

2. Run the `wafliron` program with the privilege level set to `advanced`.

Physically moving an aggregate composed of disks

To move an aggregate composed of disks from one storage system (the source) to another (the target), you need to physically move disks, disk shelves, or entire loops or stacks. You might move an aggregate to move data to a new storage system model or remove data from an impaired storage system.

Before you begin

The *target* storage system must meet the following requirements:

- It must be running a version of Data ONTAP that is the same or later than the version running on the source system.

  Note: You cannot move an aggregate between a system running Data ONTAP operating in 7-Mode and a system running Data ONTAP operating in Cluster-Mode.

- It must support the shelf, I/O module, and disk types being moved.

- It must support the size of the aggregate being moved.

About this task

The procedure described here does *not* apply to aggregates composed of array LUNs.

If you are moving an aggregate composed of disks using Storage Encryption, you need to take some extra steps before and after moving the aggregate. If the physical security of the disks during the move is a concern and your key management server supports the creation of a trust relationship between two storage systems, then you should use that capability to retain secure encryption on the disks during the move. Otherwise, you must set the encryption key to a known value before moving the disks and give them a new authentication key after the disks are installed in the destination storage system. This is the method described in the steps below.

Steps

1. Enter the following command at the source storage system to locate the disks that contain the aggregate:

   aggr status *aggr_name* -r

   The locations of the data, parity, and dParity disks in the aggregate appear under the HA, SHELF, and BAY columns (dParity disks appear for RAID-DP aggregates only).

2. If you are moving disks using Storage Encryption, reset their authentication key to their MSID (the default security ID set by the manufacturer) by entering the following command on the source system:

   disk encrypt rekey 0x0 *disk_list*

   You can also use the wildcard character (*) to specify the disks to be rekeyed. For example, to rekey all disks in a specific shelf, you can specify *adapter-name.shelI-ID*.* as your disk list.

3. Boot the source storage system into Maintenance mode.

4. Take the aggregate offline by entering the following command:

   aggr offline *aggr_name*

   The aggregate is taken offline and its hosted volumes are unmounted.

5. Reboot into Normal mode.

6. If disk ownership autoassignment is on, turn it off by entering the following command:

   options disk.auto_assign off

   If the system is part of an HA pair, you must complete this step on each node.

7. Remove the software ownership information from the disks to be moved by entering the following command for each disk:

   disk assign *disk_name* -s unowned -f

8. Follow the instructions in the disk shelf hardware guide to remove the disks or shelves that you identified previously from the source storage system.

Note: Removing a shelf requires system downtime. You cannot "hot-remove" a shelf from a storage system.

9. If you turned off disk ownership autoassignment previously, turn it back on by entering the following command:

   options disk.auto_assign on

   If the system is part of an HA pair, you must complete this step on each node.

10. Install the disks or disk shelves in the target storage system.

11. Assign the disks that you moved to the target storage system by entering the following command for each moved disk:

    disk assign *disk_name*

    The newly relocated aggregate is offline and considered as a foreign aggregate. If the newly relocated aggregate has the same name as an existing aggregate on the target storage system, Data ONTAP renames it *aggr_name*(1), where *aggr_name* is the original name of the aggregate.

12. Confirm that the newly relocated aggregate is complete by entering the following command:

    aggr status *aggr_name*

    Attention: If the aggregate is incomplete (if it has a status of `partial`), add all missing disks before proceeding. Do not try to add missing disks after the aggregate comes online—doing so causes them to become hot spare disks. You can identify the disks currently used by the aggregate by using the `aggr status -r` command.

13. If the storage system renamed the aggregate because of a name conflict, enter the following command to rename the aggregate:

    aggr rename *aggr_name new_name*

14. Enter the following command to bring the aggregate online in the destination storage system:

    aggr online *aggr_name*

    The aggregate comes online and is no longer considered to be a foreign aggregate.

15. Enter the following command to confirm that the added aggregate came online:

    aggr status *aggr_name*

16. If you moved disks using Storage Encryption, set them back to a secure encryption key.

If you...	Then...
Want all of the disks on your storage system to have the same authentication key and you have the authentication key for the other disks on the storage system	Rekey the disks that were moved using the existing authentication key by entering the following command at the destination storage system prompt: **`disk encrypt rekey` *`new_key_ID disk_list`*** *`new_key_ID`* is the key ID for the existing authentication key on the destination storage system. You can use the wildcard character (*) to specify the disks to be rekeyed. For example, to rekey all disks in a specific shelf, you can specify *adapter-name.shelI-ID*.* as your disk list.
Want all of the disks on your storage system to have the same authentication key and you do not have the authentication key for the other disks on the storage system	Rekey all of the disks in the destination storage system by entering the following command at the destination storage system prompt: **`key_manager rekey -keytag` *`key_tag`*** You can allow Data ONTAP to generate a new authentication key automatically or provide your own by using the `-manual` parameter.
Do not need all of the disks on your storage system to have the same authentication key	Rekey the disks that were moved by entering the following command: **`disk encrypt rekey` *`new_key_ID disk_list`*** *`new_key_ID`* is the key ID for a new authentication key on the destination storage system. You can use the wildcard character (*) to specify the disks to be rekeyed. For example, to rekey all disks in a specific shelf, you can specify *adapter-name.shelI-ID*.* as your disk list.

After you finish

After you move the aggregate and bring it online in the destination storage system, you need to re-create the following configuration information for all volumes associated with the aggregate:

- Client connections (CIFS shares or NFS exports)
- Scheduled tasks (for example, deduplication or reallocation)
- Quotas
- Relationships between volumes (for example, SnapMirror or SnapVault)
- FlexCache volumes
- LUN connection information

Moving an aggregate composed of array LUNs

You might want to move an aggregate composed of array LUNs to a less loaded system in the V-Series neighborhood to balance the load processing over the systems.

Before you begin

- You should plan the number and size of your aggregates ahead of time so that you have flexibility in the amount of the workload that you can shift from one system in the V-Series neighborhood to another.
- You should ensure that the *target* system meets the following requirements:
  - The target system must be running a version of Data ONTAP that is the same as or later than the version running on the source system.
  - The target system must support the size of the aggregate being moved.

About this task

To move the aggregate composed of array LUNs from one storage system (the source) to another (the target), you need to change the ownership of each array LUN in the aggregate from the source system to the target system. You can move both aggregates and traditional volumes using this procedure.

Note: If there are vFiler units in the aggregate you want to move, you might prefer to use SnapMover to move the aggregate. When SnapMover is used to move a vFiler unit, all aggregates in the vFiler unit are moved with the vFiler unit. To use vFiler units, you must have MultiStore software and SnapMover. See the *Data ONTAP MultiStore Management Guide for 7-Mode* for more information.

Steps

1. Enter the following commands on the target system:
   a) Obtain the system ID of the target system by entering either of the following commands:

      `disk show`
      or

      `sysconfig`

      You need to provide the target system's ID on the source system when you assign each of the array LUNs to the target system.
2. Enter the following commands on the source system:
   a) Enter the following command to display the array LUNs that the aggregate contains:

      `aggr status` ***`aggr_name`*** **`-r`**

The array LUNs that are displayed are the LUNs that you need to reassign to the target system to be able to move the aggregate.

b) Write down the names of the array LUNs in the aggregate that you want to move.

c) Enter the following command to shut down the source system:

halt

d) At the boot environment prompt, enter the following command to boot the source system:

bye

e) Interrupt the boot process by pressing Ctrl-C when you see the following message on the console:
`Press Ctrl-C for Boot menu`

f) Enter Maintenance mode.

g) When prompted whether you want to continue with booting, enter the following:

y

h) Enter the following command to take the aggregate offline:

aggr offline *aggr_name*

aggr_name is the name of the traditional volume or aggregate.

i) Enter the following and confirm that the aggregate is offline:

aggr status

j) In Maintenance mode, enter the following command *separately* for each array LUN in the aggregate that you are moving to the target system:

disk assign -s *system_id_target* *disk_id* -f

system_id_target is the system ID of the target system (the system to which you want to move the array LUN.)

disk_id is the ID of the array LUN you want to move.

Note: Entering this command automatically removes ownership of the array LUN from the source system and assigns it to the target system.

3. Enter the following commands on the target system.

a) Enter the following command to start a scan so that the target system can recognize the LUNs you moved to it as its own:

disk show

b) Enter the following command:

aggr status

The display shows the *foreign* aggregate as offline. (The aggregate you are moving is a foreign aggregate to the target system.) If the foreign aggregate has the same name as an existing aggregate on the system, Data ONTAP renames it *aggr_name(1)*, where *aggr_name* is the original name of the aggregate.

Attention: If the foreign aggregate is incomplete, that is, if you have not moved all the array LUNs in the aggregate, go back to the source system to add the missing array LUNs to the aggregate you moved to the target system. (Enter the following on the source system:

disk assign -s *system_id_target disk_id* -f

c) If Data ONTAP renamed the foreign aggregate because of a name conflict and you want to change the name, enter the following command to rename the aggregate :

aggr rename *aggr_name new_name*

aggr_name is the name of the aggregate you want to rename.

new_name is the new name of the aggregate.

Example

The following command renames the users(1) aggregate as newusers:

```
aggr rename users(1) newusers
```

d) Enter the following command to confirm that the aggregate you moved came online:

aggr status *aggr_name*

aggr_name is the name of the aggregate.

4. On the source system, reboot the system out of Maintenance mode.

Using volumes

Volumes are data containers. In a NAS environment, they contain file systems that hold user data that is accessible using one or more of the access protocols supported by Data ONTAP, including NFS, CIFS, HTTP, FTP, FC, and iSCSI. In a SAN environment, they contain LUNs.

Volumes depend on their associated aggregate for their physical storage.

FlexCache volumes, FlexClone volumes, and SnapLock volumes are types of volumes that share some characteristics with FlexVol volumes, but also have some special capabilities and requirements.

Related references

Storage limits on page 342

How FlexVol volumes work

A FlexVol volume is a volume that is loosely coupled to its containing aggregate. A FlexVol volume can share its containing aggregate with other FlexVol volumes. Thus, a single aggregate can be the shared source of all the storage used by all the FlexVol volumes contained by that aggregate.

Because a FlexVol volume is managed separately from the aggregate, you can create small FlexVol volumes (20 MB or larger), and you can increase or decrease the size of FlexVol volumes in increments as small as 4 KB.

When a FlexVol volume is created, it reserves a small amount of extra space (approximately 0.5 percent of its nominal size) from the free space of its containing aggregate. This space is used to store volume metadata. Therefore, upon creation, a FlexVol volume with a space guarantee of `volume` uses free space from the aggregate equal to its size × 1.005. A newly-created FlexVol volume with a space guarantee of `none` or `file` uses free space equal to 0.005 × its nominal size.

Related tasks

FlexVol volume operations on page 186

Related references

Storage limits on page 342

Differences between 64-bit and 32-bit FlexVol volumes

FlexVol volumes are one of two formats: 64-bit or 32-bit. A 64-bit volume has a larger maximum size than a 32-bit volume.

A newly created FlexVol volume is the same format as its associated aggregate. However, a volume can be a different format from its associated aggregate in certain cases.

The maximum size of a 64-bit volume is determined by the size of its associated aggregate, which depends on the storage system model.

A 32-bit volume has a maximum size of 16 TB.

Note: For both volume formats, the maximum size for each LUN or file is 16 TB.

Related references

Storage limits on page 342

Interoperability between 64-bit and 32-bit FlexVol volumes

Some Data ONTAP features use two FlexVol volumes; those volumes can be different formats. Some of these features can interoperate between different volume formats, but some cannot.

Data ONTAP feature	Interoperates between 64-bit and 32-bit format?
FlexCache	Y
`ndmpcopy`	Y
Qtree SnapMirror	Y
Synchronous SnapMirror	Y
`volume copy`	Y
Volume SnapMirror	Y
`volume move` (DataMotion for Volumes)	32-bit to 64-bit only
DataMotion for vFiler	N

How traditional volumes work

A traditional volume is a volume that is contained by a single, dedicated, aggregate. It is tightly coupled with its containing aggregate. No other volumes can get their storage from this containing aggregate.

The only way to increase the size of a traditional volume is to add entire disks to its containing aggregate. You cannot decrease the size of a traditional volume. The smallest possible traditional volume uses all the space on two disks (for RAID4) or three disks (for RAID-DP).

Traditional volumes and their containing aggregates are always of type 32-bit. You cannot grow a traditional volume larger than 16 TB.

You cannot use SSDs to create a traditional volume.

How the volume language attribute affects data visibility and availability

Every volume has a language. The language of the volume determines the character set Data ONTAP uses to display file names and data for that volume.

Attention: You are strongly advised to set all volumes to have the same language as the root volume, and to set the volume language at volume creation time. Changing the language of an existing volume can cause some files to become inaccessible.

The language of the root volume has special significance, because it affects or determines the following items:

- Default language for all volumes
- System name
- Domain name
- Console commands and command output
- NFS user and group names
- CIFS share names
- CIFS user account names
- Access from CIFS clients that don't support Unicode
- How configuration files in /etc are read
- How the home directory definition file is read

Note: Regardless of the language you specify for the root volume, names of the following objects must be in ASCII characters:

- Qtrees
- Snapshot copies

- Volumes
- Aggregates

For more information about the root volume, see the *System Administration Guide*.

How file access protocols affect what language to use for your volumes

Your choice of file access protocol (CIFS or NFS) affects the languages you should choose for your volumes.

Protocols in use	Volume language
NFS Classic (v2 or v3) only	Language setting does not matter
NFS Classic (v2 or v3) and CIFS	Language of the clients
NFS v4, with or without CIFS	*cl_lang*.UTF-8, where *cl_lang* is the language of the clients. **Note:** If you use NFS v4, all NFS Classic clients must be configured to present file names using UTF-8.

How to manage duplicate volume names

Volume names must be unique for a storage system. When Data ONTAP detects a duplicate volume name, it renames one of the volumes, but the substitute names it uses can cause problems, so you need to take corrective action if that happens.

When Data ONTAP detects a potential duplicate volume name, it appends the string "(*d*)" to the end of the name of the new volume, where *d* is a digit that makes the name unique.

For example, if you have a volume named vol1 and you copy a volume named vol1 from another storage system, Data ONTAP renames the newly copied volume to vol1(1).

You must rename any volume with an appended digit as soon as possible, for the following reasons:

- The name containing the appended digit is not guaranteed to persist across reboots.
  Renaming the volume prevents the name of the volume from changing unexpectedly later on.
- The parentheses characters, "(" and ")", are not legal characters for NFS.
  Any volume whose name contains those characters cannot be exported to NFS clients.
- The parentheses characters could cause problems for client scripts.

Volume states and status

Volumes can be in one of four states—online, offline, restricted, or quiesced. In addition, they can show one or more status values, depending on how they are configured and the health of their disks.

You can determine a volume's current state and status by using the `vol status` command.

The following table displays the possible states for volumes.

State	Description
`online`	Read and write access to this volume is allowed.
`restricted`	Some operations, such as parity reconstruction, are allowed, but data access is not allowed.
`offline`	No access to the volume is allowed.
`quiesced`	The volume is in the final stages of a move. Data access is not allowed, and many volume, qtree, and quota management operations are temporarily unavailable.

The following table displays the possible status values for volumes.

Note: Although FlexVol volumes do not directly involve RAID, the state of a FlexVol volume includes the state of its containing aggregate. Thus, the states pertaining to RAID apply to FlexVol volumes as well as traditional volumes.

Status	Description
`access denied`	The origin system is not allowing access. (FlexCache volumes only.)
`connecting`	The caching system is trying to connect to the origin system. (FlexCache volumes only.)
`copying`	The volume is currently the target of an active `vol copy` or `snapmirror` operation.
`degraded`	The volume's containing aggregate contains at least one degraded RAID group that is not being reconstructed after single disk failure.
`double degraded`	The volume's containing aggregate contains at least one degraded RAID-DP group that is not being reconstructed after double disk failure.
`flex`	The volume is a FlexVol volume.

Status	Description
`flexcache`	The volume is a FlexCache volume.
`foreign`	Disks used by the volume's containing aggregate were moved to the current storage system from another storage system.
`growing`	Disks are being added to the volume's containing aggregate.
`initializing`	The volume's containing aggregate is being initialized.
`invalid`	The volume does not contain a valid file system.
`ironing`	A WAFL consistency check is being performed on the volume's containing aggregate.
`lang mismatch`	The language setting of the origin volume was changed since the caching volume was created. (FlexCache volumes only.)
`mirror degraded`	The volume's containing aggregate is mirrored and one of its plexes is offline or resynchronizing.
`mirrored`	The volume's containing aggregate is mirrored.
`needs check`	A WAFL consistency check needs to be performed on the volume's containing aggregate.
`out-of-date`	The volume's containing aggregate is mirrored and needs to be resynchronized.
`partial`	At least one disk was found for the volume's containing aggregate, but two or more disks are missing.
`raid0`	The volume's containing aggregate consists of RAID0 (no parity) groups (array LUNs only).
`raid4`	The volume's containing aggregate consists of RAID4 groups.
`raid_dp`	The volume's containing aggregate consists of RAID-DP groups.
`reconstruct`	At least one RAID group in the volume's containing aggregate is being reconstructed.
`rem vol changed`	The origin volume was deleted and re-created with the same name. (FlexCache volumes only.)
`rem vol unavail`	The origin volume is offline or has been deleted. (FlexCache volumes only.)

Status	Description
`remote nvram err`	The origin system is experiencing problems with its NVRAM. (FlexCache volumes only.)
`resyncing`	One of the plexes of the volume's containing mirrored aggregate is being resynchronized.
`snapmirrored`	The volume is in a SnapMirror relationship with another volume.
`trad`	The volume is a traditional volume.
`unrecoverable`	The volume is a FlexVol volume that has been marked unrecoverable; if you have a volume in this state, you should contact technical support.
`unsup remote vol`	The origin system is running a version of Data ONTAP the does not support FlexCache volumes or is not compatible with the version running on the caching system. (FlexCache volumes only.)
`verifying`	RAID mirror verification is running on the volume's containing aggregate.
`wafl inconsistent`	The volume or its containing aggregate has been marked corrupted. If you have a volume in this state, you should contact technical support.

Related concepts

Using FlexCache volumes to accelerate data access on page 193

How security styles affect data access

Each volume and qtree on the storage system has a security style. The security style determines what type of permissions are used for data on volumes when authorizing users. You must understand what the different security styles are, when and where they are set, how they impact permissions, how they differ between volume types, and more.

For more information about security styles, see the *Data ONTAP File Access and Protocols Management Guide for 7-Mode*.

Improving client performance with traditional and lease oplocks

Traditional oplocks (opportunistic locks) and lease oplocks enable a CIFS client in certain file-sharing scenarios to perform client-side caching of read-ahead, write-behind, and lock information. A client can then read from or write to a file without regularly reminding the server that it needs access to the file in question. This improves performance by reducing network traffic.

Lease oplocks are an enhanced form of oplocks available with the SMB 2.1 protocol and later. Lease oplocks allow a client to obtain and preserve client caching state across multiple SMB opens originating from itself.

For more information, see the *Data ONTAP File Access and Protocols Management Guide for 7-Mode*.

How Data ONTAP can automatically provide more space for full FlexVol volumes

Data ONTAP uses two methods for automatically providing more space for a FlexVol volume when that volume is nearly full: allowing the volume size to increase, and deleting Snapshot copies (with any associated storage object). If you enable both of these methods, you can specify which method Data ONTAP should try first.

Data ONTAP can automatically provide more free space for the volume by using one of the following methods:

- Increase the size of the volume when it is nearly full (known as the *autogrow* feature).
  This method is useful if the volume's containing aggregate has enough space to support a larger volume. You can configure Data ONTAP to increase the size in increments and set a maximum size for the volume. The increase is automatically triggered based on the amount of data being written to the volume in relation to the current amount of used space and any thresholds set.
- Delete Snapshot copies when the volume is nearly full.
  For example, you can configure Data ONTAP to automatically delete Snapshot copies that are not linked to Snapshot copies in cloned volumes or LUNs, or you can define which Snapshot copies you want Data ONTAP to delete first—your oldest or newest Snapshot copies. You can also determine when Data ONTAP should begin deleting Snapshot copies—for example, when the volume is nearly full or when the volume's Snapshot reserve is nearly full.
  For more information about deleting Snapshot copies automatically, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

If you enable both of these methods, you can specify which method Data ONTAP tries first when a volume is nearly full. If the first method does not provide sufficient additional space to the volume, Data ONTAP tries the other method next. By default, Data ONTAP tries to increase the size of the volume first.

Considerations for changing the maximum number of files allowed on a volume

Volumes have a maximum number of files that they can contain. You can change the maximum number of files for a volume, but before doing so you should understand how this change affects the volume.

The number of files a volume can contain is determined by how many inodes it has. An *inode* is a data structure that contains information about files. Volumes have both private and public inodes. Public inodes are used for files that are visible to the user; private inodes are used for files that are used internally by Data ONTAP. You can change only the maximum number of public inodes for a volume. You cannot affect the number of private inodes.

Data ONTAP automatically sets the maximum number of public inodes for a newly created volume based on the size of the volume: 1 inode per 32 KB of volume size. When the size of a volume is increased, either directly by an administrator or automatically by Data ONTAP through the autosize feature, Data ONTAP also increases (if necessary) the maximum number of public inodes so there is at least 1 inode per 32 KB of volume size, until the volume reaches approximately 1 TB in size. Growing the volume greater than 1 TB in size does not automatically result in more inodes, because Data ONTAP does not automatically create more than 33,554,409 inodes. If you need more files than the default number for any size volume, you can use the `maxfiles` command to increase the maximum number of inodes for the volume.

You can also decrease the maximum number of public inodes. This does not change the amount of space currently allocated to inodes, but it does lower the maximum amount of space the public inode file can consume. However, after space has been allocated for inodes, it is never returned to the volume. Therefore, lowering the maximum number of inodes below the number of inodes currently allocated does not return the space used by the allocated but unused inodes to the volume.

Cautions for increasing the maximum directory size for FlexVol volumes

The default maximum directory size for FlexVol volumes is model-dependent, and optimized for the size of system memory. Before increasing the maximum directory size, involve customer support.

You can increase the default maximum directory size for a specific volume by using the `maxdirsize` volume option, but doing so could impact system performance.

Understanding the root volume

The storage system's root volume contains special directories and configuration files that help you administer the storage system. Understanding the facts about the root volume helps you manage it.

The following facts apply to the root volume:

- How the root volume is installed and whether you need to create it yourself depend on the storage system:
  - For FAS systems and V-Series systems ordered with disk shelves, the root volume is a FlexVol volume that is installed at the factory.
  - For a V-Series system that does not have a disk shelf, you install the root volume on an array LUN.
    For more information about setting up a V-Series system, see the *Data ONTAP Software Setup Guide for 7-Mode*.
  - For systems running virtual storage, the Data ONTAP-v installation process creates a single aggregate by using all currently defined virtual disks and creates the root FlexVol volume in that aggregate.
    For more information about system setup, see the Installation and Administration Guide that came with your Data ONTAP-v system.
- The default name for the root volume is `/vol/vol0`.
  You can designate a different volume to be the new root volume. Starting in Data ONTAP 8.0.1, you can designate a 64-bit volume to be the new root volume.
- The root volume's fractional reserve must be 100%.

Related tasks

Changing the root volume on page 184

Recommendations for the root volume

There are recommendations to keep in mind when choosing what kind of volume to use for the root volume.

The following are general recommendations for root volumes:

- Root volumes can use either FlexVol or traditional volumes.
  If a root volume exists as a traditional volume, it can be a stand-alone RAID4 or RAID-DP volume. RAID4 requires a minimum of two disks and can protect against single-disk failures. RAID-DP, the default RAID type, requires a minimum of three disks and can protect against double-disk failures. Using RAID-DP for the root aggregate is recommended.
  Data ONTAP 8.0 or later allows you to create only a new FlexVol root volume, not a new traditional root volume, from the boot menu. However, preexisting traditional root volumes are still supported.

- It is recommended that the root volume be in a separate aggregate that does not include data volumes or other user data.

  However, for small storage systems where cost concerns outweigh resiliency, a FlexVol based root volume on a regular aggregate might be more appropriate.
- You should avoid storing user data in the root volume, regardless of the type of volume used for the root volume.
- For a V-Series system with a disk shelf, the root volume can reside on the disk shelf (recommended) or on the third-party storage.

  For a V-Series system that does not have a disk shelf, the root volume resides on the third-party storage. You can install only one root volume per V-Series system, regardless of the number of storage arrays or disk shelves that the V-Series system uses for storage.

The following are additional facts and considerations if the root volume is on a disk shelf:

- Smaller stand-alone root volumes offer fault isolation from general application storage; on the other hand, FlexVol volumes have less impact on overall storage utilization, because they do not require two or three disks to be dedicated to the root volume and its small storage requirements.
- If a FlexVol volume is used for the root volume, file system consistency checks and recovery operations could take longer to finish than with the two- or three-disk traditional root volume.

  FlexVol recovery commands work at the aggregate level, so all of the aggregate's disks are targeted by the operation. One way to mitigate this effect is to use a smaller aggregate with only a few disks to house the FlexVol volume containing the root volume.
- In practice, having the root volume on a FlexVol volume makes a bigger difference with smaller capacity storage systems than with very large ones, in which dedicating two disks for the root volume has little impact.
- For higher resiliency, use a separate two-disk root volume.

  Note: You should convert a two-disk root volume to a RAID-DP volume when performing a disk firmware update, because RAID-DP is required for disk firmware updates to be nondisruptive. When all disk firmware and Data ONTAP updates have been completed, you can convert the root volume back to RAID4.

  For Data ONTAP 7.3 and later, the default RAID type for traditional root volume is RAID-DP. If you want to use RAID4 as the raid type for your traditional root volume to minimize the number of disks required, you can change the RAID type from RAID-DP to RAID4 by using `vol options vol0 raidtype raid4`.

The following requirement applies if the root volume is on a storage array:

- For storage systems whose root volume is on a storage array, only one array LUN is required for the root volume regardless of whether the root volume is a traditional volume or a FlexVol volume.

Special system files

For storage systems upgraded from a release earlier than Data ONTAP 8.0, some system files exist in every volume of the system. You must not remove or modify these files unless technical support directs you to do so. These files enable you to restore LUNs in Snapshot copies if you revert to a release earlier than Data ONTAP 8.0.

The following system files are in the root level of every volume, including the root volume:

- `.vtoc_internal`
- `.bplusvtoc_internal`

General volume operations

General volume operations are operations you can perform on either a FlexVol volume or a traditional volume. They include managing a volume's language, viewing or changing its state, renaming or destroying it, increasing the number of files it can contain, and running a reallocation operation on it.

Migrating from traditional volumes to FlexVol volumes

You cannot convert directly from a traditional volume to a FlexVol volume. You must create a new FlexVol volume and then move the data to the new volume.

NetApp Professional Services staff, including Professional Services Engineers (PSEs) and Professional Services Consultants (PSCs) are trained to assist customers with migrating data between volume types, among other services. For more information, contact your local NetApp Sales representative, PSE, or PSC.

Related concepts

How FlexVol volumes work on page 163
Using volumes on page 163

Preparing your destination volume

Before migrating, you need to create and name a destination volume of the correct size and number of inodes.

Before you begin

If your destination volume is on the same storage system as the source volume, your system must have enough free space to contain both copies of the volume during the migration.

If the new FlexVol volume will be the root volume, it must meet the minimum size requirements for root volumes, which are based on your storage system. Data ONTAP prevents you from designating as root a volume that does not meet the minimum size requirement. For more information, see the *Data ONTAP System Administration Guide for 7-Mode*.

Steps

1. Enter the following command to determine the amount of space your traditional volume uses:

   df -Ah *vol_name*

Example

```
sys1> df -Ah vol0
Aggregate        total       used     avail    capacity
vol0              24GB     1434MB      22GB          7%
vol0/.snapshot  6220MB     4864MB    6215MB          0%
```

The total space used by the traditional volume is displayed as `used` for the volume name.

2. Enter the following command to determine the number of inodes your traditional volume uses:

   `df -I vol_name`

   Example

```
sys1> df -I vol0
Filesystem           iused      ifree  %iused  Mounted on
vol0               1010214   27921855     3%   /vol/vol0
```

   The number of inodes your traditional volume uses is displayed as `iused`.

3. Identify or create an aggregate to contain the new FlexVol volume.

   Note: To determine if an existing aggregate is large enough to contain the new FlexVol volume, you can use the `df -Ah` command. The space listed under `avail` should be large enough to contain the new FlexVol volume.

4. If you want the destination (FlexVol) volume to have the same name as the source (traditional) volume, and they are on the same storage system, you must rename the source volume before creating the destination volume. Do this by entering the following command:

   `aggr rename vol_name new_vol_name`

   Example

   `aggr rename vol0 vol0trad`

5. Create the destination volume in the containing aggregate.

   Example

   `vol create vol0 aggrA 90g`

   Note: For root volumes, you must use the (default) volume space guarantee, because it ensures that writes to the volume do not fail due to a lack of available space in the containing aggregate.

6. Confirm that the size of the destination volume is at least as large as the source volume by entering the following command on the target volume:

   `df -h vol_name`

7. Confirm that the destination volume has at least as many inodes as the source volume by entering the following command on the destination volume:

   df -I *vol_name*

   Note: If you need to increase the number of inodes in the destination volume, use the `maxfiles` command.

Result

You have created a destination volume with sufficient resources to accept the data from the source volume.

Related tasks

Creating an aggregate on page 139
Creating a FlexVol volume on page 186

Migrating your data

You use the `ndmpcopy` command from the Data ONTAP prompt to migrate your data to the target volume.

Steps

1. Ensure that NDMP is configured correctly by entering the following commands:

   options ndmpd.enable on

   options ndmpd.authtype challenge

   Note: If you are migrating your volume between storage systems, make sure that these options are set correctly on both systems.

2. Disable data access to the source volume.
3. Migrate the data by entering the following command at the storage system prompt:

   ndmpcopy *src_vol_name dest_vol_name*

   Example

   ndmpcopy /vol/vol0trad /vol/vol0

   Attention: Make sure that you use the storage system command-line interface to run the `ndmpcopy` command. If you run this command from a client, your data will not migrate successfully.

   For more information about the `ndmpcopy` command, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

4. Verify that the `ndmpcopy` operation completed successfully by validating the copied data.

Result

The target volume now contains the data from the source volume.

Snapshot copies on the source volume are not affected by this procedure. However, they are not replicated to the target FlexVol volume as part of the migration.

Completing the migration

After you copy your data, you need to perform some additional tasks before the migration is complete.

Steps

1. If you are migrating your root volume, complete the following steps:
   a) Make the new FlexVol volume the root volume by entering the following command:
   `vol options vol_name root`

   Example

   `vol options vol0 root`
   b) Reboot the storage system.
2. Update the clients to point to the new FlexVol volume.
   - In a CIFS environment, complete these steps:

   a) Point CIFS shares to the new FlexVol volume.
   b) Update the CIFS maps on the client machines so that they point to the new FlexVol volume.
   - In an NFS environment, complete these steps:

   a) Point NFS exports to the new FlexVol volume.
   b) Update the NFS mounts on the client machines so that they point to the new FlexVol volume.
3. Make sure that all clients can see the new FlexVol volume and read and write data:
   a) Using a CIFS or an NFS client, create a new folder or directory.
   b) Using the client, copy some temporary data into the new folder or directory and confirm that you can access that data from the client.
   c) Delete the new folder.
4. If you are migrating the root volume and you changed the name of the root volume, update the `httpd.rootdir` option to point to the new root volume.
5. If quotas were used with the traditional volume, configure the quotas on the new FlexVol volume.
6. Create a Snapshot copy of the target volume and create a new Snapshot schedule as needed.

   For more information, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.
7. Start using the migrated volume for the data source for your applications.

8. When you are confident the volume migration was successful, you can take the original volume offline or destroy it.

   Note: You should preserve the original volume and its Snapshot copies until the new FlexVol volume has been stable for some time.

Putting a volume into restricted state

You use the `vol restrict` command to put a volume into restricted state, which makes it unavailable for read or write access by clients. You might want to do this if you want the volume to be the target of a volume copy or SnapMirror replication operation.

About this task

When you restrict a FlexVol volume, it relinquishes any unused space that has been allocated for it in its containing aggregate. If this space is allocated for another volume and then you bring the volume back online, this can result in an overcommitted aggregate.

Related concepts

Using volumes on page 163

Related tasks

Putting an aggregate into restricted state on page 152

Taking a volume offline

You use the `vol offline` command to take a volume offline to perform maintenance on the volume, move it, or destroy it. When a volume is offline, it is unavailable for read or write access by clients.

About this task

When you take a FlexVol volume offline, it relinquishes any unused space that has been allocated for it in its containing aggregate. If this space is allocated for another volume and then you bring the volume back online, this can result in an overcommitted aggregate.

Note: You cannot take the root volume offline.

Note: If you attempt to take a volume offline while any files contained by that volume are open, the `volume offline` command fails and displays the names (or inodes, if i2p is disabled) of the files that are open, along with the processes that opened them.

Related concepts

Using volumes on page 163

Related tasks

Taking an aggregate offline on page 151

Bringing a volume online

After you restrict a volume or take it offline, you can make it available to the storage system again by bringing it online using the `vol online` command.

About this task

If you bring a FlexVol volume online into an aggregate that does not have sufficient free space to fulfill the space guarantee for that volume, this command fails.

Attention: If the volume you are bringing online is inconsistent, the `vol online` command prompts you for confirmation. If you bring an inconsistent volume online, it might suffer further file system corruption.

Related concepts

Using volumes on page 163

Related tasks

Bringing an aggregate online on page 151

Renaming a volume

You use the `vol rename` command to rename a volume. You can rename volumes without interrupting data service.

Step

1. Enter the following command:
   `vol rename *vol_name new-name*`

Result

The following events occur:

- The volume is renamed.
- If NFS is in use and the `nfs.export.auto-update` option is On, the `/etc/exports` file is updated to reflect the new volume name.
- If CIFS is running, shares that refer to the volume are updated to reflect the new volume name.
- The in-memory information about active exports gets updated automatically, and clients continue to access the exports without problems.

After you finish

If you access the storage system using NFS, add the appropriate mount point information to the `/etc/fstab` or `/etc/vfstab` file on clients that mount volumes from the storage system.

Destroying a volume

If you no longer need a volume and the data it contains, you can destroy the volume to free its space for other data.

About this task

When you destroy a FlexVol volume, all the disks included in its containing aggregate remain assigned to that containing aggregate, although the space associated with the volume is returned as free space to the containing aggregate.

When you destroy a traditional volume, however, you also destroy the traditional volume's dedicated containing aggregate. This converts its parity disk and all its data disks back into hot spares. After the disks have been zeroed, you can use them in other aggregates, traditional volumes, or storage systems.

Attention: If you destroy a volume, the data in the volume is no longer accessible.

Steps

1. Take the volume offline by entering the following command:

   `vol offline` *`vol_name`*

2. Enter the following command to destroy the volume:

   `vol destroy` *`vol_name`*

Result

The following events occur:

- The volume is destroyed.
- If NFS is in use and the `nfs.exports.auto-update` option is `on`, entries in the `/etc/exports` file that refer to the destroyed volume are removed.
- If CIFS is running, any shares that refer to the destroyed volume are deleted.
- If the destroyed volume was a FlexVol volume, its allocated space is freed, becoming available for allocation to other FlexVol volumes contained by the same aggregate.
- If the destroyed volume was a traditional volume, the disks it used become hot spare disks.

After you finish

If you access your storage system using NFS, update the appropriate mount point information in the `/etc/fstab` or `/etc/vfstab` file on clients that mount volumes from the storage system.

Displaying file or inode usage

FlexVol volumes have a maximum number of files that they can contain. Knowing how many files are contained by your volumes helps you determine whether you need to increase the number of (public) inodes for your volumes to prevent them from hitting their maximum file limit.

About this task

Public inodes can be either free (they are not associated with a file) or used (they point to a file). The number of free inodes for a volume is the total number of inodes for the volume minus the number of used inodes (the number of files).

Step

1. To display inode usage for a volume, enter the following command:

   df -i *volume_name*

   You can omit the volume name; in this case, Data ONTAP displays the inode usage for all volumes on the node.

Example

```
shl5> df -i
Filesystem               iused      ifree  %iused  Mounted on
/vol/root_vs0/              97      19893      0%  /vol/root_vs0/
/vol/vol0/                8455   10977711      0%  /vol/vol0/
/vol/v1/                   100       1096      8%  /vol/v1/
```

Changing the maximum number of files allowed in a volume

Volumes have a limit on the number of files they can contain. You can change this limit using the `maxfiles` command, which affects the maximum number of public inodes the volume can have.

Steps

1. Enter the following command:

   maxfiles *vol_name max_num_files*

   Note: Inodes are added in blocks. If the requested increase in the number of files is too small to require a new inode block to be added, the maxfiles value is not increased. If this happens, repeat the command with a larger value for *max_num_files*.

You cannot decrease *max_num_files* below the number of currently allocated public inodes, but the number of public inodes may be less than the current value of *max_num_files*.

2. You can confirm the new maximum number of files, as well as the number of files currently present in the volume, by entering the following command:

   maxfiles *vol_name*

   Note: The value returned reflects only the number of files that can be created by users, or public inodes; the private inodes reserved for internal use are not included in this number.

Changing the language for a volume

You should use caution when changing the language for an existing volume, because doing so could affect the system's ability to display your data. In addition, a system reboot is necessary before the language change is complete.

Before you begin

Before changing the language that a volume uses, be sure you understand how volumes use the language attribute and how this change could affect access to your data.

Steps

1. Determine the correct language code for your volume.

   You can view the possible language codes by using the `vol lang` command.

2. Enter the following command to change the volume language:

   vol lang *vol_name language*

   Note: If you are changing the NFS character set, you are asked to confirm your choice, and also whether you want to halt the system so that WAFL_check can be run to check for any files that will no longer be accessible using NFS. The default answer for this question is **yes**. If you do not want to halt the system, you must enter **n**.

3. Reboot the storage system.

   Note: Although the language change is effective for the target volume immediately, the full effect of the change is not complete until after the reboot.

After you finish

You can verify the new language by using the `vol status -l` command.

Related concepts

How the volume language attribute affects data visibility and availability on page 165

Using volumes on page 163

Changing the root volume

Every storage system must have a root volume. Therefore, you must always have one volume designated as the root volume. However, you can change which volume is used as the system's root volume.

Before you begin

The volume that you are designating to be the new root volume must meet the minimum size requirement. The required minimum size for the root volume varies, depending on the storage system model. If the volume is too small to become the new root volume, Data ONTAP prevents you from setting the root option.

In addition, the volume that you are designating to be the new root volume must have at least 2 GB of free space. It must also have a fractional reserve of 100%. The `vol status –v` command displays information about a volume's fractional reserve.

If you use a FlexVol volume for the root volume, ensure that it has a guarantee of `volume`.

Starting in Data ONTAP 8.0.1, you can designate a volume in a 64-bit aggregate to be the new root volume.

If you move the root volume outside the current root aggregate, you must also change the value of the aggregate `root` option so that the aggregate containing the root volume becomes the root aggregate.

For V-Series systems with the root volume on the storage array, the array LUN used for the root volume must meet the minimum array LUN size for the root volume. For more information about the minimum array LUN size for the root volume on V-Series systems, see the *Hardware Universe* at *hwu.netapp.com*.

About this task

You might want to change the storage system's root volume, for example, when you migrate your root volume from a traditional volume to a FlexVol volume.

Steps

1. Identify an existing volume to use as the new root volume, or create the new root volume by using the `vol create` command.
2. Use the `ndmpcopy` command to copy the `/etc` directory and all of its subdirectories from the current root volume to the new root volume.

   For more information about `ndmpcopy`, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.
3. Enter the following command to specify the new root volume:

vol options *vol_name* root

vol_name is the name of the new root volume.

If the volume does not have at least 2 GB of free space, the command fails and an error message appears.

After a volume is designated to become the root volume, it cannot be brought offline or restricted.

4. If you moved the root volume outside the current root aggregate, enter the following command to change the value of the aggregate `root` option so that the aggregate containing the root volume becomes the root aggregate:

   aggr options *aggr_name* root

   aggr_name is the name of the new root aggregate.

   For more information about the aggregate `root` option, see the na_aggr(1) man page.

5. Enter the following command to reboot the storage system:

   reboot

   When the storage system finishes rebooting, the root volume is changed to the specified volume.

   If you changed the root aggregate, a new root volume is created during the reboot when the aggregate does not already contain a FlexVol volume designated as the root volume and when the aggregate has at least 2 GB of free space.

6. Update the `httpd.rootdir` option to point to the new root volume.

FlexVol volume operations

You can create FlexVol volumes, clone them, determine the amount of space they use, resize them, and display their containing aggregate, among other tasks.

Related concepts

How FlexVol volumes work on page 163

Creating a FlexVol volume

You create FlexVol volumes to provide resizable, flexible data containers that can be mounted and accessed using all data access protocols supported by Data ONTAP.

Before you begin

Before creating a FlexVol volume, you must have determined the following attributes for the new volume:

- Name
  The volume name must conform to the following requirements:
  - Begin with either a letter or an underscore (_)
  - Contain only letters, digits, and underscores
  - Contain no more than 250 characters
  - Be different from all other volume names on the storage system
- Size
  The volume must be at least 20 MB in size. Its maximum size depends on whether it is in a 32-bit or 64-bit aggregate and the model of the storage system that hosts the volume.
- Language
  The default language is the language of the root volume.
- Space guarantee setting (optional)
  The default space guarantee is `volume`.
- If the volume is to be accessed using CIFS, the CIFS oplocks setting
- Security style setting

Steps

1. If you have not already done so, create the aggregate that will contain the FlexVol volume that you want to create.
2. Enter the following command:

vol create *vol_name* [-l *language_code*] [-s {volume|file|none}] *aggr_name* *size*{k|m|g|t}

vol_name is the name for the new FlexVol volume (without the /vol/ prefix)

language_code specifies a language other than that of the root volume.

-s {volume | file | none} specifies the space guarantee setting that is enabled for the specified FlexVol volume. If no value is specified, the default value is volume.

aggr_name is the name of the containing aggregate for the new FlexVol volume.

size{k | m | g | t} specifies the volume size in kilobytes, megabytes, gigabytes, or terabytes. For example, you would enter **20m** to indicate 20 megabytes. If you do not specify a unit, size is taken as bytes and rounded up to the nearest multiple of 4 KB.

Example

The following command creates a 200-MB volume called newvol, in the aggregate called aggr1, using the French character set:

vol create newvol -l fr aggr1 200M

The new volume is created and, if NFS is in use, an entry is added to the /etc/exports file for the new volume. The default automatic Snapshot schedule is applied to the new volume.

3. If you access the storage system using CIFS, update the share information for the new volume.
4. If you access the storage system using NFS, complete the following steps:
   a) Verify that the line added to the /etc/exports file for the new volume is correct for your security model.
   b) Add the appropriate mount point information to the /etc/fstab or /etc/vfstab file on clients that mount volumes from the storage system.

After you finish

If needed, you can verify that the CIFS oplocks and security style settings are correct, and modify them as needed.

Note: You should set these values as soon as possible after creating the volume. If you change these values after files are in the volume, the files might become inaccessible to users because of conflicts between the old and new values. For example, UNIX files available under mixed security might not be available after you change to NTFS security.

If the default automatic Snapshot schedule does not match your data protection strategies, update the Snapshot schedule for the newly created volume with a more appropriate schedule. For more information about Snapshot schedules, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related concepts

How the volume language attribute affects data visibility and availability on page 165

How volume guarantees work with FlexVol volumes on page 264
Using volumes on page 163

Related tasks

Creating an aggregate on page 139

Related references

Storage limits on page 342

Resizing a FlexVol volume

You can increase or decrease the amount of space that an existing FlexVol volume is allowed to occupy in its containing aggregate. A FlexVol volume can grow to the size you specify as long as the containing aggregate has enough free space to accommodate that growth.

Steps

1. Check the available space of the containing aggregate by entering the following command:

   df -A *aggr_name*

2. If you want to determine the current size of the volume, enter one of the following commands:

   vol size *vol_name*

   df *vol_name*

3. Enter the following command to resize the volume:

   vol size *vol_name* [+|-] *n*{k|m|g|t}

   If you include the + or -, *n*{k|m|g|t} specifies how many kilobytes, megabytes, gigabytes or terabytes to increase or decrease the volume size. If you do not specify a unit, size is taken as bytes and rounded up to the nearest multiple of 4 KB.

   If you omit the + or -, the size of the volume is set to the size you specify, in kilobytes, megabytes, gigabytes, or terabytes. If you do not specify a unit, size is taken as bytes and rounded up to the nearest multiple of 4 KB.

   Note: If you attempt to decrease the size of a FlexVol volume to less than the amount of space that it is currently using, the command fails.

   Decreasing the size of a FlexVol volume does not decrease the space reserved for metadata for the volume (it remains .5 percent of the original nominal size of the volume).

4. You can verify the success of the resize operation by entering the following command:

   vol size *vol_name*

Related references

Storage limits on page 342

Displaying the containing aggregate for a FlexVol volume

You display a FlexVol volume's containing aggregate by using the `vol container` command.

Traditional volume operations

Operations that apply exclusively to traditional volumes usually involve management of the aggregate to which that volume is closely coupled.

About this task

Additional traditional volume operations described in other chapters or other guides include the following operations:

- Configuring and managing SyncMirror replication of volume data
  See the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.
- Configuring and managing SnapLock volumes
  See the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.

Related concepts

Changing the RAID level of an aggregate on page 152
How Data ONTAP uses RAID to protect your data and data availability on page 105

Related tasks

Increasing the size of an aggregate on page 147
Physically moving an aggregate composed of disks on page 156

Creating a traditional volume

Traditional volumes do not provide the flexibility that FlexVol volumes do, because they are tightly coupled with their containing aggregate. However, if you want a single-volume aggregate, you can create a traditional volume.

Before you begin

You must have determined the name of the traditional volume. Volume names must conform to the following requirements:

- Begin with either a letter or an underscore (_)
- Contain only letters, digits, and underscores
- Contain no more than 250 characters

Note: You can change the name of a traditional volume later by using the `aggr rename` command.

You must have determined what disks will be used in the new volume. You can specify disks by listing their IDs, or by specifying a disk characteristic such as speed or type. You can display a list of the available spares on your storage system by using the `aggr status -s` command.

You must have determined the CIFS oplocks setting for the new volume.

You must have determined the security setting for the new volume.

Steps

1. Enter the following command:

   aggr create *vol_name* -v [-l *language_code*] [-f] [-m] [-n] [-v] [-t {raid4|raid_dp}] [-r *raidsize*] [-T *disk-type*] -R *rpm*] [-L] *disk-list*

   vol_name is the name for the new volume (without the `/vol/` prefix).

   language_code specifies the language for the new volume. The default is the language of the root volume.

   Note: For a description of the RAID-related parameters, see the na_aggr(1) man page or the information about creating aggregates.

   The new volume is created and, if NFS is in use, an entry for the new volume is added to the `/etc/exports` file. The default automatic Snapshot schedule is applied to the new volume.

2. Verify that the volume exists as you specified:

   aggr status *vol_name* -r

   The system displays the RAID groups and disks of the specified volume on your storage system.

3. If you access the storage system using CIFS, update your CIFS shares as necessary.
4. If you access the storage system using NFS, complete the following steps:
   a) Verify that the line added to the `/etc/exports` file for the new volume is correct for your security model.
   b) Add the appropriate mount point information to the `/etc/fstab` or `/etc/vfstab` file on clients that mount volumes from the storage system.
5. Verify that the CIFS oplocks and security style settings are correct, or modify them as needed.

   You should update these values as soon as possible after creating the volume. If you change the values after files are in the volume, the files might become inaccessible to users because of conflicts between the old and new values. For example, UNIX files available under mixed security might not be available after you change to NTFS security.

After you finish

If the default automatic Snapshot schedule does not match your data protection strategies, update the Snapshot schedule for the newly created volume with a more appropriate schedule. For more information about Snapshot schedules, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*.

Related concepts

How the volume language attribute affects data visibility and availability on page 165
How to control disk selection from heterogeneous storage on page 134
Using volumes on page 163
How you use aggregates to provide storage to your volumes on page 126

Related tasks

Creating an aggregate on page 139

Related references

Storage limits on page 342

Using FlexCache volumes to accelerate data access

A FlexCache volume is a sparsely-populated volume on a local storage system that is backed by a volume on a different, optionally remote, storage system. A sparsely-populated volume or a sparse volume provides access to data in the backing volume (also called the origin volume) without requiring that all the data be in the sparse volume.

You can use only FlexVol volumes to create FlexCache volumes. However, many of the regular FlexVol volumes features are not supported on FlexCache volumes, such as Snapshot copy creation, deduplication, compression, FlexClone volume creation, volume move, and volume copy.

You can use FlexCache volumes to speed up access to data, or to offload traffic from heavily accessed volumes. FlexCache volumes help improve performance, especially when clients need to access the same data repeatedly, because the data can be served directly without having to access the source. Therefore, you can use FlexCache volumes to handle system workloads that are read-intensive.

Cache consistency techniques help in ensuring that the data served by the FlexCache volumes remains consistent with the data in the origin volumes.

Related tasks

FlexCache volume operations on page 209

How FlexCache volumes serve read requests

A FlexCache volume directly serves read requests if it contains the data requested by the client. Otherwise, the FlexCache volume requests the data from the origin volume and stores the data before serving the client request. Subsequent read requests for the data are then served directly from the FlexCache volume.

This improves performance when the same data is accessed repeatedly, because after the first request, the data no longer has to travel across the network, or be served from an overloaded system.

FlexCache hardware and software requirements

Before you can create FlexCache volumes and use them to access data in their origin volumes, you must ensure that both your origin and caching systems meet the hardware and software requirements for the FlexCache functionality.

The requirements for the caching system and the origin system are different.

For the caching system, the following requirements must be met:

- If you want to cache a 7-Mode volume, ensure that the caching system has one of the following versions of Data ONTAP:
  - Data ONTAP 7.0.5 or later in the 7.0 release family
  - Data ONTAP 7.2.1 or later in the 7.2 release family
  - Any version in the Data ONTAP 7.3 release family
  - Data ONTAP 8.x or later operating in 7-Mode

    Note: Systems running Data ONTAP 8.0 or 8.1 clustered releases or any version in the Data ONTAP 10.0 release family cannot serve as caching systems.

    Note: The caching and origin systems do not need to have the same version of Data ONTAP.
- The caching system for a clustered Data ONTAP volume must have Data ONTAP 8.x or later operating in 7-Mode.
- The caching system must have a valid NFS license, with NFS enabled.

  Note: The NFS license is not required when the caching system is an SA system.
- The `licensed_feature.flexcache_nfs.enable` option must be set to `on`.

For the origin system, the following requirements must be met:

- The system must have one of the following versions of Data ONTAP:
  - Data ONTAP 7.0.1 or later in the 7.x release families
  - Any version in the Data ONTAP 8.x release families
  - Data ONTAP 10.0.4 or later in the 10.0 release family

  Note: If your origin system is running Data ONTAP 10.0 or Data ONTAP 8.x, your caching system must have Data ONTAP 7.2.1 or later.
- For caching a clustered Data ONTAP volume, the origin system must be running clustered Data ONTAP 8.2 or later.
- The origin system must have a valid NFS license, with NFS enabled.
- The `flexcache.access` option must be set to allow access to FlexCache volumes.

  Note: For more information about this option, see the na_protocolaccess(8) man page.

  If the origin volume is in a vFiler unit, you must set this option for the vFiler context.
- The `flexcache.enable` option must be set to `on`.

  Note: If the origin volume is in a vFiler unit, you must set this option for the vFiler context.

For information about configuring and managing FlexCache volumes in a clustered Data ONTAP environment, see the *Clustered Data ONTAP Logical Storage Management Guide*

Limitations of FlexCache volumes

You can have a maximum of 100 FlexCache volumes on a storage system. In addition, there are certain features of Data ONTAP that are not available on FlexCache volumes, and others that are not available on volumes that are backing FlexCache volumes.

You cannot use the following Data ONTAP capabilities on FlexCache volumes (these limitations do not apply to the origin volumes):

- Client access using any protocol other than NFSv2 or NFSv3
- Client access using IPv6
- Compression (compressed origin volumes are supported)
- Snapshot copy creation
- SnapRestore
- SnapMirror (qtree or volume)
- SnapVault
- FlexClone volume creation
- The `ndmp` command
- Quotas
- Qtrees
- Volume copy
- Deduplication
- Creation of FlexCache volumes in any vFiler unit other than vFiler0
- Creation of FlexCache volumes in the same aggregate as their origin volume
- Mounting the FlexCache volume as a read-only volume

If your origin volume is larger than 16 TB, the output of the `df` command on the caching system will show "---" for the size information about the origin volume. To see the size information for the origin volume, you must run the `df` command on the origin system.

You cannot use the following Data ONTAP capabilities on FlexCache origin volumes or storage systems without rendering all of the FlexCache volumes backed by that volume or storage system unusable:

Note: If you want to perform these operations on an origin system, you can destroy the affected FlexCache volumes, perform the operation, and re-create the FlexCache volumes. However, the FlexCache volumes will need to be repopulated.

- You cannot move an origin volume between vFiler units or to vFiler0 by using any of the following commands:
  - `vfiler move`
  - `vfiler add`
  - `vfiler remove`

  - `vfiler destroy`

    Note: You can use SnapMover (`vfiler migrate`) to migrate an origin volume without having to re-create FlexCache volumes backed by that volume.

    Origin volumes can be owned by any vFiler unit.

- You cannot use a FlexCache origin volume as the destination of a `snapmirror migrate` command.
- You cannot change the language of the origin volume if the change causes the underlying character set to change, or if the new language is not available on the caching system.

  For example, you can change the language of the origin volume from English to US English. However, if you want to change the language from English to a language that uses a different character set, such as Japanese, then you need to destroy and re-create all of the FlexCache volumes backed by the origin volume.
- Qtrees contained by the origin volume that belong to a vFiler unit other than the vFiler unit that owns the origin volume are not accessible to a FlexCache volume.

  For example, suppose that volume `vol1` is owned by vFiler0 but `qtree1`, which is contained by `vol1`, is owned by another vFiler unit. FlexCache volumes created with `vol1` as the backing volume will not be able to access the data contained in `qtree1`.
- If your origin volume is on a system running a version of the Data ONTAP 10.0 release family, and any node in the origin cluster is down, the FlexCache volume will not be able to establish a connection with the origin volume.
- If the origin volume contains Snapshot copies, the Snapshot data is not written to disk (cached). Snapshot data is stored only in the in-memory buffer cache of the caching filer.

Types of volumes you can use for FlexCache

A FlexCache volume must be a FlexVol volume. The origin volume can be a FlexVol or a traditional volume; it can also be a SnapLock volume. There are some restrictions on what can be used as an origin volume.

FlexCache volumes and FlexVol origin volumes can be either 32-bit or 64-bit volumes; a FlexCache volume does not need to be the same type as its origin volume (a 32-bit FlexCache volume can have a 64-bit origin volume and vice versa).

You cannot use the following storage containers as a FlexCache origin volume:

- A FlexCache volume
- A volume that contains SnapVault destinations
- A qtree

How the FlexCache Autogrow capability works

For best caching performance, you should allow Data ONTAP to control the size of your FlexCache volumes, by using the FlexCache Autogrow capability.

Making your FlexCache volume too small can negatively impact your caching performance. When the FlexCache volume begins to fill up, it flushes randomly chosen, previously cached files to make room for newly requested data. When data from the flushed files is requested again, it must be retrieved again from the origin volume.

Therefore it is best to use the Autogrow capability and allow Data ONTAP to increase the size of your FlexCache volumes as the size of the working set increases. This method has the following advantages:

- If the size of the FlexCache volume's working set increases, as long as there is space in the containing aggregate, the FlexCache volume automatically increases its size rather than ejecting data from the cache, which could affect data access performance.
- These size increases happen without operator intervention.
- If you have several FlexCache volumes sharing the same aggregate, the volumes that are getting the most data accesses will also receive the most space.
- If you increase the size of an aggregate, the FlexCache volumes contained by that aggregate will automatically take advantage of the extra space if needed.

The Autogrow capability is enabled by default in new FlexCache volumes created without specifying a size using Data ONTAP 7.3 and later. You can enable the Autogrow capability on existing FlexCache volumes by using the `vol options` command with the `flexcache_autogrow` option.

Note: Before the Autogrow capability was available, the preferred sizing strategy for FlexCache volumes was to size the FlexCache volume to the same size as its containing aggregate. If this approach is providing you with the performance and space utilization you need, you do not need to reconfigure those existing FlexCache volumes to use the Autogrow capability.

How FlexCache volumes use space management

FlexCache volumes do not use space management in the same manner as regular FlexVol volumes. The amount of disk space reserved for a FlexCache volume is determined by the value of the `flexcache_min_reserved` volume option, rather than the nominal size of the FlexCache volume.

The default value for the `flexcache_min_reserved` volume option is 100 MB. In general, you should not change the value of this option.

Attention: FlexCache volumes' space guarantees must be honored. When you take a FlexCache volume offline, the space allocated for the FlexCache becomes available for use by other volumes in the aggregate (as with all FlexVol volumes). However, unlike regular FlexVol volumes,

FlexCache volumes cannot be brought online if there is insufficient space in the aggregate to honor their space guarantee.

Related concepts

How volume guarantees work with FlexVol volumes on page 264
Using volumes on page 163

How FlexCache volumes share space with other volumes

You can have multiple FlexCache volumes in the same aggregate. You can also have regular FlexVol volumes in the same aggregate as the FlexCache volumes. If you want to set up your system efficiently, you must understand the way these volumes share space.

When you include multiple FlexCache volumes in the same aggregate, each FlexCache volume reserves only a small amount of space. The rest of the space is allocated as required. This means that a "hot" FlexCache volume (one that is being accessed heavily) is permitted to take up more space, while a FlexCache volume that is not being accessed as often will gradually be reduced in size.

Note: When an aggregate containing FlexCache volumes runs out of free space, Data ONTAP randomly selects a FlexCache volume in that aggregate to be truncated. Truncation means that files are removed from the FlexCache volume until the size of the volume is decreased to a predetermined percentage of its former size.

If you have regular FlexVol volumes in the same aggregate as your FlexCache volumes, and the aggregate starts filling up, the FlexCache volumes can lose some of their unreserved space (if it is not being used). In this case, when the FlexCache volume needs to fetch a new data block and it does not have enough free space to accommodate the data block, an existing data block is removed from one of the FlexCache volumes to accommodate the new data block.

If the ejected data is causing many cache misses, you can add more space to the aggregate or move some of the data to another aggregate.

Methods to view FlexCache statistics

Data ONTAP provides statistics about FlexCache volumes to help you understand the access patterns and administer the FlexCache volumes effectively.

You can display statistics for your FlexCache volumes by using the following methods:

- The `flexcache stats` command (client and server statistics)
- The `nfsstat` command (client statistics only)
- The `perfstat` utility
- The `stats` command

Note: You can use the fstat command for statistics about the objects cached by the FlexCache volume.

For more information about the commands, see the na_flexcache(1), na_stats(1), and nfsstat(1) man pages.

Related tasks

Displaying FlexCache client statistics on page 211
Displaying FlexCache server statistics on page 212

What happens when connectivity to the origin system is lost

You can control how the FlexCache volume functions when connectivity between the caching and origin systems is lost by using the `disconnected_mode` and `acdisconnected` volume options.

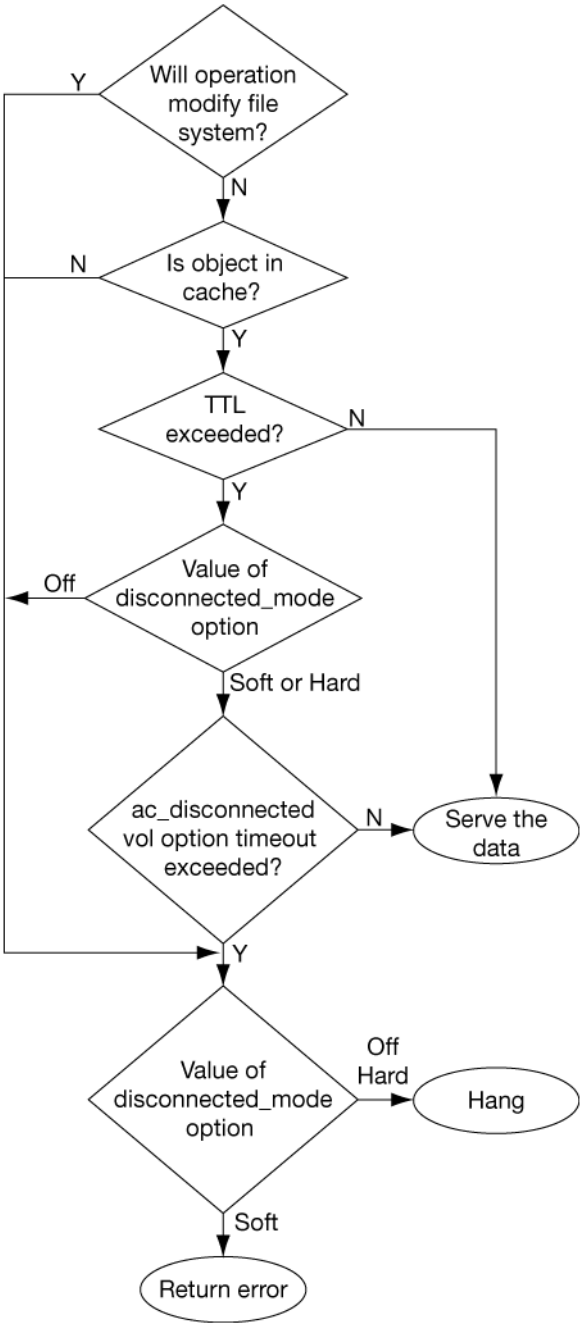
The `disconnected_mode` volume option and the `acdisconnected` timeout, combined with the regular TTL timeouts (`acregmax`, `acdirmax`, `acsymmax`, and `actimeo`), enable you to control the behavior of the FlexCache volume when contact with the origin volume is lost.

When you configure the FlexCache disconnected options, you should consider the following questions:

- Would your applications or file access protocols react better if an I/O request returned an error or if it did not return at all?
- How long can you safely serve stale data when connectivity is lost?

The following flowchart shows the multi-step decision process used by Data ONTAP to determine what happens when a FlexCache volume is disconnected from its origin volume. The possible outcomes of this process are:

- The data is served.
- An error is returned.
- The operation hangs.


Will operation modify file system?
Y
N
Is object in cache?
N
Y
TTL exceeded?
N
Y
Value of disconnected_mode option
Off
Soft or Hard
ac_disconnected vol option timeout exceeded?
N
Serve the data
Y
Value of disconnected_mode option
Off
Hard
Hang
Soft
Return error

How the NFS export status of the origin volume affects FlexCache access

A volume does not need to be exported to serve as an origin volume for a FlexCache volume. If you want to prevent a volume from being an origin volume, set the `flexcache.access` option to `none`.

How FlexCache caching works

Understanding how FlexCache determines the validity of cached data will help you determine whether your data set is a good candidate for a FlexCache.

What a cached file contains

When the client requests a data block of a specific file from a FlexCache volume, then the attributes of that file are cached, and that file is considered to be cached, even if all of its data blocks are not present in the FlexCache volume. If the requested data is cached and valid, a read request for that data is fulfilled without access to the origin volume.

How data changes affect FlexCache volumes

How data changes affect FlexCache volumes depends on where the change is made: on the FlexCache volume, the origin volume, or another FlexCache volume. If a file is directly updated on the origin volume, the cached copy of the file is invalidated. If the write request is relayed to the origin volume, only the changed blocks are invalidated in the FlexCache volume.

Writes to a file on the origin volume

When a change is made to a file on the origin system, Data ONTAP revokes the delegation for that file and invalidates the entire file for all FlexCache volumes backed by that origin volume.

Note: The FlexCache copy of the file is not invalidated until an access to that file is made on the FlexCache volume.

The cache is not affected when only the access time of a file is updated.

Writes to a file on the FlexCache volume

When a write is made to a file on the FlexCache volume, the write request is relayed to the origin volume. When the origin volume acknowledges the request, the blocks that were changed are invalidated on the FlexCache volume, but the rest of the file remains valid.

How cache consistency is achieved

Cache consistency for FlexCache volumes is achieved by using three primary techniques: *delegations*, *attribute cache timeouts*, and *write operation proxy*.

Delegations ensure that the FlexCache volumes can directly serve client read requests without having to access the origin volume. As long as the FlexCache volume continues to have a delegation for the data on a file, the origin volume does not modify the contents of the file. The origin volume must revoke all the delegations to a file from the FlexCache volumes before modifying the contents of the file.

If a FlexCache volume does not have delegations to a cached file, the data in the cache is considered valid for a duration determined by the attribute cache timeout value.

Any FlexCache volume that receives a write request from the client proxies the request to the origin volume. If the FlexCache volume contains a cached copy of the data that changes in the origin, the FlexCache volume removes the file containing the cached data and stores the updated data during a subsequent read operation.

Delegations

A delegation is a token that the origin system grants the caching volume to ensure that the caching volume can serve read requests without the need for validating the data with the origin volume. Delegations are used only in certain situations.

When the FlexCache volume retrieves data from a file on the origin volume, the origin volume provides a delegation for that file to the FlexCache volume. The FlexCache volume retains the delegation until the origin volume revokes the delegation before the particular file gets modified. After the data is modified, the FlexCache volume must fetch the data before serving it to clients.

If multiple caching volumes have delegations for a particular file, then all the delegations are revoked before the file is updated.

Note: You can use the `lock status -p flexcache` command to view the FlexCache delegations granted by the storage system.

Delegations can cause a performance decrease for writes to the origin volume, depending on the number of FlexCache volumes that are holding delegations for the file being modified.

The following list outlines situations when delegations cannot be used to guarantee that an object has not changed:

- Objects other than regular files: Directories, symbolic links, and other objects that are not regular files have no delegations.
- Origin volumes that are SnapMirror destinations: If the origin volume is a SnapMirror destination, delegations are not used.
- When the FlexCache volume is taken offline: All the delegations given to the volume are destroyed.

- When connectivity is lost: If connectivity is lost between the origin and caching systems, then delegations are considered to be revoked.
- When the maximum number of delegations is reached: If the origin volume cannot store all of its delegations, it might revoke an existing delegation to accommodate a new one.
- When the origin volume has a resource constraint: The origin volume reclaims some delegations from all its FlexCache volumes.

Attribute cache timeouts

When data is retrieved from the origin volume, the file that contains that data is considered valid in the FlexCache volume as long as a delegation exists for that file. If no delegation exists, the file is considered valid for a certain length of time, specified by the attribute cache timeout.

If a client requests data from a file for which there are no delegations, and the attribute cache timeout has been exceeded, the FlexCache volume compares the file attributes of the cached file with the attributes of the file on the origin system. Then one of the following actions is taken:

- If the two sets of file attributes match, the requested data is directly returned to the client (if it was already in the FlexCache volume) or retrieved from the origin system and then returned to the client.
- If the two sets of file attributes do not match, the file is marked as invalid in the cache. Then the requested data blocks are read from the origin system and stored in the FlexCache volume, as if it were the first time that file had been accessed from that FlexCache volume.

With attribute cache timeouts, clients can get stale data when all of the following conditions are true:

- There are no delegations for the file on the caching volume.
- The file's attribute cache timeout has not been reached.
- The file has changed on the origin volume since it was last accessed by the caching volume.

Note: Clients can get stale data when a file on the origin volume is added to or removed from a directory that is already stored on the FlexCache volume. The file addition or deletion does not become visible on the FlexCache until the length of time specified in the directory attribute cache timeout (`acdirmax`) has passed since the last time the directory was updated on the FlexCache volume.

To prevent clients from ever getting stale data, you can set the attribute cache timeout to `0`. However, this negatively affects your caching performance, because every data request for which there is no delegation causes an access to the origin system.

The attribute cache timeouts are determined by using volume options. The option names and default values are outlined in the following table.

Volume option	Description	Default value (seconds)
`acdirmax`	Attribute cache timeout for directories	`15s`

Volume option	Description	Default value (seconds)
`acregmax`	Attribute cache timeout for regular files	`15s`
`acsymmax`	Attribute cache timeout for symbolic links	`15s`
`actimeo`	Attribute cache timeout for all objects	`15s`

For more information about modifying these options, see the na_vol(1) man page.

Write operation proxy

When a client writes to a cached file, the FlexCache volume proxies the write request to the origin volume, which makes the requested changes. The FlexCache volume is then updated to reflect the changes.

When writes are relayed through the FlexCache volume, if that volume contains the portion of the file that is changing, the entire file is invalidated from the FlexCache volume. In addition, the origin volume revokes any delegation to the file. On subsequent access of the new data, the cache volume requests the data from the origin volume.

If other FlexCache volumes are mapped to the origin volume and if they contain the changed data, then the origin volume revokes all delegations to the cached data in the FlexCache volumes. If the cached data on any of these FlexCache volumes is accessed after the attribute cache timeout expires, the particular FlexCache volume verifies the file attributes with those on the origin volume, invalidates its own copy of the cached file, and stores the updated file.

What cache hits and misses are

Cache hits and misses indicate if the data requested by the client is served directly from the FlexCache volume or from the origin volume. The occurrence of a cache hit or miss depends on factors such as availability of the requested data in the cache, the attribute cache timeout values, and the difference between attributes of a file in the cache and the origin.

Cache hits

When a client sends a read request, if the relevant block is cached in the FlexCache volume, the data is read directly from the FlexCache volume. This is called a *cache hit*. Cache hits are the result of a previous request.

A cache hit can be one of the following types:

- Hit: The requested data is already cached and the FlexCache volume serves the read request without accessing the origin volume.

- Hit-Verify: The requested data is already cached but the FlexCache volume needs to verify the attributes of the cached file with those on the origin. The FlexCache does not request any data from the origin volume while performing this verification process.

Cache misses

If the requested data is not on the FlexCache volume, or if the data has changed since it was cached, the caching system loads the data from the origin volume, and then returns it to the requesting client. This is called a *cache miss*.

A cache miss can be one of the following types:

- Miss: The requested data is not cached and is read from the origin volume.
- Miss-Verify: The requested data is cached but the file attributes have changed since it was last cached. Therefore, the file is removed from the cache, and the requested data is read from the origin.

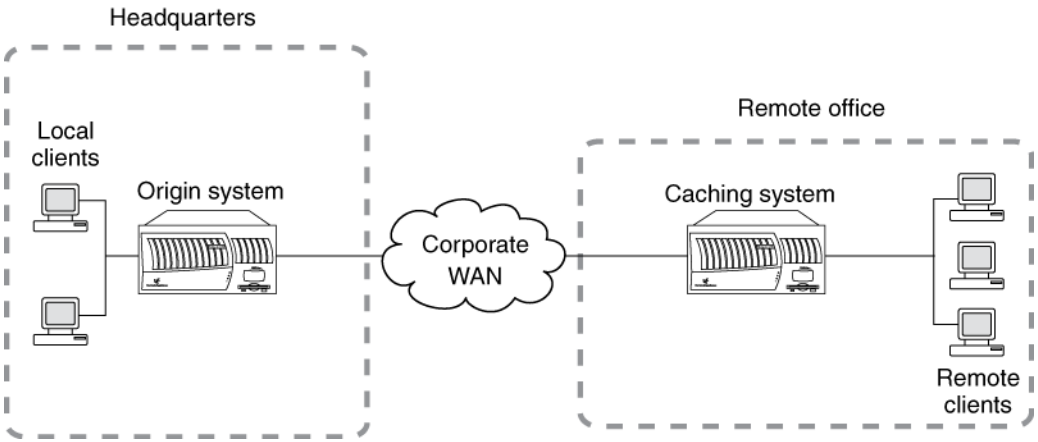
Typical FlexCache deployments

FlexCache is typically used in WAN deployments (which decrease average access time for remote clients) and LAN deployments (which reduce the workload of an overloaded storage system).

WAN deployment

In a WAN deployment, the FlexCache volume is remote from the data center. As clients request data, the FlexCache volume caches popular data, giving the end user faster access to information.

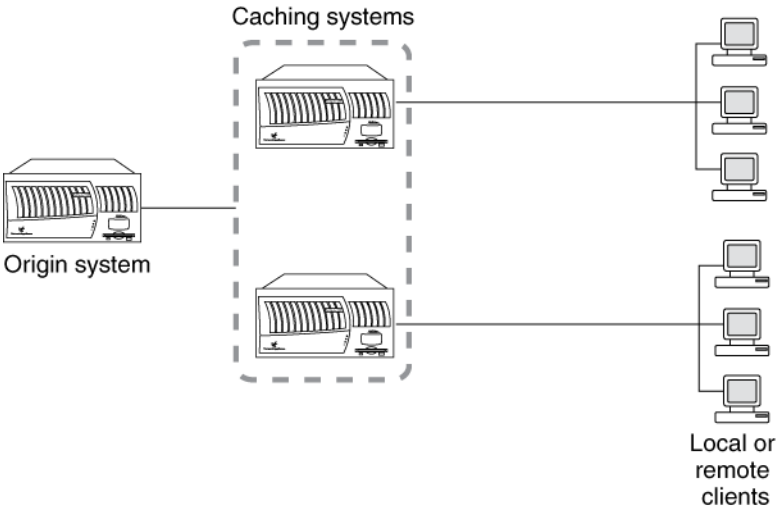
The FlexCache volume is placed as close as possible to the remote office. Client requests are then explicitly directed to the FlexCache volume. If valid data exists in the cache, that data is served directly to the client. If the data does not exist in the cache, it is retrieved across the WAN from the origin system, cached in the FlexCache volume, and returned to the client. A WAN deployment is shown in the following diagram.



LAN deployment

In a LAN deployment, or accelerator mode, the FlexCache volume is local to the administrative data center, and is used to offload work from busy file servers and free system resources.

Frequently accessed data, or "hot objects," are replicated and cached by the FlexCache volumes. This reduces network collisions and latency because the data access load is shared amongst all of the caching systems. A LAN deployment is shown in the following diagram.



Using FlexCache volumes to cache clustered Data ONTAP volumes

You can cache a clustered Data ONTAP volume outside the cluster by using a FlexCache volume in a storage system that is deployed at a remote site and connected to the cluster over a Wide Area Network (WAN). This type of configuration can help in situations such as distributing read-intensive data across multiple locations.

You must consider the following when using a FlexCache volume to cache a clustered Data ONTAP volume outside the cluster:

- The FlexCache volume must be connected directly to the clustered Data ONTAP volume and not to any namespace or junction point.
- The client applications cannot access volumes that are associated with the FlexCache volumes through junctions.

For more information about clustered Data ONTAP volumes, see the *Clustered Data ONTAP Logical Storage Management Guide*.

Related tasks

Creating FlexCache volumes on page 209

About using LUNs in FlexCache volumes

You cannot use SAN access protocols to access FlexCache volumes. You can cache a volume that contains LUNs, but this configuration can change system behavior.

When you attempt to access, in a FlexCache volume, a directory that contains a LUN, the command sometimes returns "stale NFS file handle" for the LUN. If you get that error message, you should repeat the command.

If you use the `fstat` command on a LUN, fstat always indicates that the LUN is not cached. This is expected behavior.

Note: LUNs in this context refer to the LUNs that Data ONTAP serves to clients, not to the array LUNs used for storage on a storage array.

What FlexCache status messages mean

When you enter the `vol status` command for a FlexCache volume, and the status of the FlexCache volume is not normal, you get a FlexCache status message.

The following table lists the status messages you might see for a FlexCache volume and what they mean.

FlexCache status	Description
`access denied`	The origin system is not allowing FlexCache access. Check the setting of the `flexcache.access` option on the origin system.
`connecting`	The caching system is trying to connect to the origin system.
`lang mismatch`	The language setting of the origin volume was changed since the FlexCache volume was created.
`rem vol changed`	The origin volume was deleted and re-created with the same name. Re-create the FlexCache volume to reenable the FlexCache relationship.
`rem vol unavail`	The origin volume is offline or has been deleted.
`remote nvram err`	The origin system is experiencing problems with its NVRAM.
`unsup remote vol`	The origin system is running a version of Data ONTAP that either does not support FlexCache volumes or is not compatible with the version running on the caching system.

How FlexCache volumes connect to their origin volume

FlexCache volumes use a proprietary protocol to connect to their origin volume. The protocol uses port 2050.

About SA systems

SA systems support a subset of features supported by FAS systems. SA systems are storage systems capable of storing only FlexCache volumes.

You manage an SA system the same way you manage a FAS system, with the following differences:

- Only FlexCache volumes and the root volume can be mounted, using NFSv2 or NFSv3.
- No file access protocol other than NFSv2 and NFSv3 is supported.
- Only the following licenses are supported:
  - flexcache_nfs
  - cluster
  - flex_scale
- The size of the root volume is restricted:
  - Traditional root volumes cannot be increased in size.
  - Flexible root volumes can grow only to 100 GB or the minimum root volume size, whichever is larger.
- SA systems can be configured in a standard HA configuration, but not a mirrored HA configuration or a mirrored MetroCluster configuration.

FlexCache volume operations

Operations you can perform with FlexCache volumes include creating them, displaying their status and free space, configuring the Autogrow capability, and flushing files that they are caching.

Related concepts

Using FlexCache volumes to accelerate data access on page 193

Creating FlexCache volumes

You use FlexCache volumes to speed access to remote data, or to offload traffic from heavily accessed volumes.

Before you begin

- You must have configured and enabled the FlexCache feature correctly on both the origin and caching systems.
- If the origin is a clustered Data ONTAP volume, you must have completed the following tasks:
  - The LIFs over which the FlexCache volume can access the clustered Data ONTAP volume are enabled with the `-fcache` protocol.
    For more information about creating and configuring LIFs, see the *Clustered Data ONTAP Network Management Guide*.
  - An export policy is created with `flexcache` as the protocol in the clustered Data ONTAP volume and the storage systems that host the FlexCache volumes as the clients.
    For more information about creating export policies, see the *Clustered Data ONTAP File Access and Protocols Management Guide*.

Step

1. Enter the following command:
   `vol create` ***`cache_vol aggr`*** **`[`*** ***`size`*** **`{k|m|g|t}] -S`** ***`origin`*** **`:`** ***`source_vol`***

   `origin` is the name or IP address of the origin system. If you are caching a clustered Data ONTAP volume, then specify the IP address of the clustered Data ONTAP system containing the particular volume.

   `cache_vol` is the name of the new FlexCache volume you want to create.

   `aggr` is the name of the containing aggregate for the new FlexCache volume.

   `size` `{ k | m | g | t }` specifies the FlexCache volume size in kilobytes, megabytes, gigabytes, or terabytes. If you do not specify a size, bytes is used and rounded up to the nearest multiple of 4 KB.

Note: For best performance, do not specify a size when you create a FlexCache volume. Specifying a size disables the FlexCache Autogrow capability.

source_vol is the name of the volume you want to use as the origin volume on the origin system.

Result

The new FlexCache volume is created and an entry is added to the `/etc/export` file for the new volume.

Example

The following command creates a FlexCache volume called `newcachevol`, with the Autogrow capability enabled, in the aggregate called `aggr1`, with a source volume `vol1` on storage system corp_storage:

```
vol create newcachevol aggr1 -S corp_storage:vol1
```

Related concepts

FlexCache hardware and software requirements on page 193
How the FlexCache Autogrow capability works on page 197
Using FlexCache volumes to accelerate data access on page 193
Using volumes on page 163
FlexCache hardware and software requirements on page 193
Using FlexCache volumes to cache clustered Data ONTAP volumes on page 206

Displaying free space for FlexCache volumes

When you use the `df` command on the caching storage system, you display the disk free space for the *origin* volume, rather than the local caching volume. You can display the disk free space for the local caching volume by using the `-L` option for the `df` command.

Configuring the FlexCache Autogrow capability

With the Autogrow capability enabled, Data ONTAP increases the size of a FlexCache volume when the volume starts to fill up. The Autogrow capability is enabled and disabled per FlexCache volume, and is enabled by default on new FlexCache volumes.

Step

1. Enter the command below, depending on the operation you want to perform:

If you want to..	Then enter...
Enable the Autogrow capability	`vol options vol_name flexcache_autogrow on`
Disable the Autogrow capability	`vol options vol_name flexcache_autogrow off`

Example

To enable the FlexCache Autogrow capability on the FlexCache volume fc1, enter the following command:

```
vol options fc1 flexcache_autogrow on
```

Related concepts

How the FlexCache Autogrow capability works on page 197

Flushing files from FlexCache volumes

If you know that a specific file has changed on the origin volume and you want to flush it from your FlexCache volume before it is accessed, you can use the `flexcache eject` command. For more information about this command, see the na_flexcache(1) man page.

Displaying FlexCache client statistics

You can use client statistics to see how many operations are being served by the FlexCache volume rather than the origin system. A large number of cache misses might indicate that the FlexCache volume is too small and data is being discarded and fetched again later.

Before you begin

Give the cache time to become populated before tracking cache misses.

Step

1. Depending on what statistics you want to see, enter the appropriate command.

If you want to...	Use this command:
Display FlexCache statistics	`flexcache stats -C`
Display NFS statistics for the FlexCache volume	`nfsstat -C`

Related concepts

Methods to view FlexCache statistics on page 198

Displaying FlexCache server statistics

If you are using the LAN deployment to offload an overloaded volume, you can use server statistics to get information about the origin system and ensure that the load is evenly distributed among the caching volumes.

Step

1. Depending on what statistics you want to see, enter the appropriate command.

If you want to...	Use this command:
Display overall server statistics	`flexcache stats -S`
Display server statistics per client	`flexcache stats -S -c` **Note:** To get per-client statistics, the `flexcache.per_client_stats` option must be set to `on`.

Related concepts

Methods to view FlexCache statistics on page 198

Displaying FlexCache status

You display the status for a FlexCache volume using the `vol status` command. If your FlexCache volume has a problem, a FlexCache status is displayed as the last line of the volume status output. If the status of the FlexCache is normal, no FlexCache status is displayed.

Related concepts

Using FlexCache volumes to accelerate data access on page 193

Related references

What FlexCache status messages mean on page 207

Using FlexClone volumes to create efficient copies of your FlexVol volumes

FlexClone volumes are writable, point-in-time copies of a parent FlexVol volume. FlexClone volumes are space-efficient because they share the same data blocks with their parent FlexVol volumes for common data. The Snapshot copy used to create a FlexClone volume is also shared with the parent volume.

You can clone an existing FlexClone volume to create another FlexClone volume. You can also create a clone of a FlexVol volume containing LUNs and LUN clones.

You can also split a FlexClone volume from its parent volume. As a result, the FlexClone volume becomes an independent FlexVol volume with its own disk space, instead of sharing disk space with its parent.

Understanding FlexClone volumes

FlexClone volumes can be managed similarly to regular FlexVol volumes, with a few important differences. For instance, the changes made to the parent FlexVol volume after the FlexClone volume is created are not reflected in the FlexClone volume.

The following list outlines important facts about FlexClone volumes:

- A FlexClone volume is a point-in-time, writable copy of the parent FlexVol volume.
- You must install the license for the FlexClone feature before you can create FlexClone volumes.
- A FlexClone volume is a fully functional FlexVol volume similar to its parent.
- A FlexClone volume is always created in the same aggregate as its parent.
- A traditional volume cannot be used as the parent of a FlexClone volume.

  To create a copy of a traditional volume, you must use the `vol copy` command, which creates a distinct copy that uses additional storage space equivalent to the amount of storage space used by the volume you copied.
- Because a FlexClone volume and its parent share the same disk space for common data, creating a FlexClone volume is instantaneous and requires no additional disk space (until changes are made to the FlexClone volume or its parent).
- A FlexClone volume is created with the same volume guarantee as its parent.

  The volume guarantee setting is enforced for the new FlexClone volume only if there is enough space in the containing aggregate.
- A FlexClone volume is created with the same space reservation and fractional reserve settings as its parent.
- A FlexClone volume is created with the same Snapshot schedule as its parent.
- A FlexClone volume is created with the same language setting as its parent.

- The common Snapshot copy shared between a FlexClone volume and its parent volume cannot be deleted while the FlexClone volume exists.
- While a FlexClone volume exists, some operations on its parent are not allowed, such as deleting the parent volume.
- You can sever the connection between the parent volume and the FlexClone volume.
  This is called *splitting* the FlexClone volume. Splitting removes all restrictions on the parent volume and causes the FlexClone volume to use its own additional disk space rather than sharing space with its parent.

  Attention: Splitting a FlexClone volume from its parent volume deletes all existing Snapshot copies of the FlexClone volume, and disables the creation of new Snapshot copies while the splitting operation is in progress.

  If you want to retain the Snapshot copies of the FlexClone volume, you can move the FlexClone volume to a different aggregate by using the vol move command. During the volume move operation, you can also create new Snapshot copies, if required. For more information about the volume move operation, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

- Quotas applied to the parent volume are *not* automatically applied to the FlexClone volume.
- The clone of a SnapLock volume is also a SnapLock volume, and inherits the expiry date of the parent volume. This date cannot be changed, and the volume cannot be destroyed before the expiry date. For more information about SnapLock volumes, see the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.
- When a FlexClone volume is created, any LUNs present in the parent volume are present in the FlexClone volume but are unmapped and offline.

Related concepts

How splitting a FlexClone volume from its parent works on page 216
How volume guarantees work with FlexVol volumes on page 264

FlexClone volumes and space guarantees

A FlexClone volume inherits its initial space guarantee from its parent volume. For example, if you create a FlexClone volume from a parent volume with a space guarantee of `volume`, then the FlexClone volume's initial space guarantee will be `volume` also. You can change the FlexClone volume's space guarantee.

For example, suppose that you have a 100-MB FlexVol volume with a space guarantee of `volume`, with 70 MB used and 30 MB free, and you use that FlexVol volume as a parent volume for a new FlexClone volume. The new FlexClone volume has an initial space guarantee of `volume`, but it does not require a full 100 MB of space from the aggregate, as it would if you had copied the volume. Instead, the aggregate needs to allocate only 30 MB (100 MB minus 70 MB) of free space to the clone.

If you have multiple clones with the same parent volume and a space guarantee of `volume`, they all share the same shared parent space with each other, so the space savings are even greater.

Note: The shared space depends on the existence of the shared Snapshot copy (the base Snapshot copy that was used to create the FlexClone volume). If you delete this shared Snapshot copy, you lose the space savings provided by the FlexClone volume.

Related concepts

FlexClone volumes and shared Snapshot copies on page 215
How volume guarantees work with FlexVol volumes on page 264

How to identify shared Snapshot copies in FlexClone volumes

You can identify a shared Snapshot copy by using the `snap list` command to list the Snapshot copies *in the parent volume*. Any Snapshot copy that is marked as busy in the parent volume and is also present in the FlexClone volume is a shared Snapshot copy.

FlexClone volumes and shared Snapshot copies

When volume guarantees are in effect, a new FlexClone volume uses the Snapshot copy it shares with its parent to minimize its space requirements. If you delete the shared Snapshot copy, you might increase the space requirements of the FlexClone volume.

For example, suppose that you have a 100-MB FlexVol volume that has a volume guarantee of `volume`, with 70 MB used and 30 MB free, and you use that FlexVol volume as a parent volume for a new FlexClone volume. The new FlexClone volume has an initial volume guarantee of `volume`, but it does not require a full 100 MB of space from the aggregate, as it would if you had copied the volume. Instead, the aggregate needs to allocate only 30 MB (100 MB – 70 MB) of free space to the clone.

Now, suppose that you delete the shared Snapshot copy from the FlexClone volume. The FlexClone volume can no longer optimize its space requirements, and the full 100 MB is required from the containing aggregate.

Note: If you are prevented from deleting a Snapshot copy from a FlexClone volume due to "insufficient space in the aggregate" it is because deleting that Snapshot copy requires the allocation of more space than the aggregate currently has available. You can either increase the size of the aggregate, or change the volume guarantee of the FlexClone volume.

How you use volume SnapMirror replication with FlexClone volumes

Because both volume SnapMirror replication and FlexClone volumes rely on Snapshot copies, there are some restrictions on how the two features can be used together. For instance, you can create a volume SnapMirror relationship using a FlexClone volume or its parent as the source volume. However, you cannot create a new volume SnapMirror relationship using either a FlexClone volume or its parent as the destination volume.

Considerations for creating a FlexClone volume from a SnapMirror source or destination volume

You can create a FlexClone volume from the source or destination volume in an existing volume SnapMirror relationship. However, doing so could prevent future SnapMirror replication operations from completing successfully.

Replication might not work because when you create the FlexClone volume, you might lock a Snapshot copy that is used by SnapMirror. If this happens, SnapMirror stops replicating to the destination volume until the FlexClone volume is destroyed or is split from its parent. You have two options for addressing this issue:

- If you require the FlexClone volume on a temporary basis and can accommodate a temporary stoppage of the SnapMirror replication, you can create the FlexClone volume and either delete it or split it from its parent when possible.

  The SnapMirror replication continues normally when the FlexClone volume is deleted or is split from its parent.
- If a temporary stoppage of the SnapMirror replication is not acceptable, you can create a Snapshot copy in the SnapMirror source volume, and then use that Snapshot copy to create the FlexClone volume. (If you are creating the FlexClone volume from the destination volume, you must wait until that Snapshot copy replicates to the SnapMirror destination volume.)

  This method of creating a Snapshot copy in the SnapMirror source volume allows you to create the clone without locking a Snapshot copy that is in use by SnapMirror.

How splitting a FlexClone volume from its parent works

Splitting a FlexClone volume from its parent removes any space optimizations that are currently used by the FlexClone volume. After the split, both the FlexClone volume and the parent volume require the full space allocation determined by their volume guarantees. The FlexClone volume becomes a normal FlexVol volume.

You must be aware of the following considerations related to clone-splitting operations:

- When you split a FlexClone volume from its parent, all existing Snapshot copies of the FlexClone volume are deleted. If you want to retain the Snapshot copies of the FlexClone volume, you can move the FlexClone volume to a different aggregate by using the `vol move` command. During the volume move operation, you can also create new Snapshot copies, if required. For details of the `vol move` command, see the *Data ONTAP SAN Administration Guide for 7-Mode*.
- No new Snapshot copies can be created of the FlexClone volume during the split operation.
- Because the clone-splitting operation is a copy operation that might take considerable time to complete, Data ONTAP provides the `vol clone split stop` and `vol clone split status` commands to stop or check the status of a clone-splitting operation.
- The clone-splitting operation proceeds in the background and does not interfere with data access to either the parent or the clone volume.
- The FlexClone volume must be online when you start the split operation.
- The parent volume must be online for the split operation to succeed.
- If you take the FlexClone volume offline while splitting is in progress, the operation is suspended; when you bring the FlexClone volume back online, the splitting operation resumes.
- If the FlexClone volume has a data protection or load sharing mirror, it cannot be split from its parent volume.
- If you split a FlexClone volume from a FlexVol volume that has deduplication and compression enabled, the split volume *does not* have deduplication and compression enabled.
- After a FlexClone volume and its parent volume have been split, they cannot be rejoined.

Related tasks

Splitting a FlexClone volume from its parent on page 220

FlexClone volumes and LUNs

You can clone FlexVol volumes that contain LUNs and LUN clones.

Note: LUNs in this context refer to the LUNs that Data ONTAP serves to clients, not to the array LUNs used for storage on a storage array.

When you create a FlexClone volume, LUNs in the parent volume are present in the FlexClone volume but they are not mapped and they are offline. To bring the LUNs in the FlexClone volume online, you need to map them to initiator groups. When the LUNs in the parent volume are backed by Snapshot copies, the FlexClone volume also inherits the Snapshot copies.

If the parent volume contains LUN clones (LUNs created by using the `lun clone` command), the FlexClone volume inherits the LUN clones and their base Snapshot copies. In this case, the LUN clone's base Snapshot copy in the parent volume shares blocks with the base Snapshot copy in the FlexClone volume. You cannot delete the LUN clone's base Snapshot copy in the parent volume while the base Snapshot copy in the FlexClone volume still exists.

If the parent volume contains FlexClone files or FlexClone LUNs (LUNs created by using the `clone start` command), the FlexClone volume also contains FlexClone files and FlexClone LUNs, which share storage with the FlexClone files and FlexClone LUNs in the parent volume.

FlexClone volume operations

Operations you can perform with FlexClone volumes include creating a FlexClone volume and splitting it from its parent volume.

Creating a FlexClone volume

If you need a temporary copy of your data that can be made quickly and without using a lot of disk space, you can create a FlexClone volume. FlexClone volumes save data space because all unchanged data blocks are shared between the FlexClone volume and its parent.

Before you begin

Ensure that you have the flex_clone license installed.

Step

1. Enter the following command to clone the volume:

   vol clone create *clone_name* [-s {volume|file|none}] -b *parent_name* [*parent_snap*]

   clone_name is the name of the FlexClone volume that you want to create.

   `-s` {`volume` | `file` | `none`} specifies the space guarantee setting for the new FlexClone volume. If no value is specified, the FlexClone volume is given the same space guarantee setting as its parent.

   parent_name is the name of the FlexVol volume that you intend to clone.

   parent_snap is the name of the base Snapshot copy of the parent FlexVol volume. If no name is specified, Data ONTAP creates a base Snapshot copy with the name `clone_`*cl_name_prefix.id*, where *cl_name_prefix* is up to 16 characters of the name of the new FlexClone volume and *id* is a unique digit identifier (for example 1, 2, and so on).

   Note: The base Snapshot copy cannot be deleted as long as any clones based on that Snapshot copy exist.

Result

The FlexClone volume is created and, if NFS is in use, an entry is added to the `/etc/exports` file for every entry found for the parent volume.

The base Snapshot copy becomes a shared Snapshot copy between the FlexClone volume and its parent.

Example

To create a FlexClone volume named newclone from the parent FlexVol volume flexvol1, you would enter the following command:

```
vol clone create newclone -b flexvol1
```

Note: The Snapshot copy created by Data ONTAP is named clone_newclone.1.

After you finish

You can verify the status of the new FlexClone volume by using the `vol status -v` command.

Related concepts

Using FlexClone volumes to create efficient copies of your FlexVol volumes on page 213
How volume guarantees work with FlexVol volumes on page 264

Splitting a FlexClone volume from its parent

If you want the FlexClone volume to have its own disk space, rather than using that of its parent, you can split it from its parent.

Steps

1. Determine the approximate amount of free space required to split a FlexClone volume from its parent by entering the following command:

   ```
   vol clone split estimate clone_name
   ```

2. Verify that enough free space exists in the containing aggregate to support the split by entering the following command:

   ```
   df -A aggr_name
   ```

   The `avail` column tells you how much available space you have in your aggregate.

3. Enter the following command to split the volume:

   ```
   vol clone split start clone_name
   ```

   The clone-splitting operation begins. All existing Snapshot copies of the clone are deleted, and the creation of Snapshot copies of the clone is prevented for the duration of the split operation.

   Note: If an online data migration operation is in progress, this command might fail. In this case, wait and retry the command when the online data migration operation is complete.

   This operation could take some time to complete, depending on how much space is shared between the FlexClone volume and its parent.

If you take no further action, when all shared data has been copied, the clone will be split from its parent volume and become a regular FlexVol volume.

4. If you want to check the status of a clone-splitting operation, enter the following command:

   `vol clone split status` ***`clone_name`***

5. If you want to stop the progress of an ongoing clone-splitting operation, enter the following command:

   `vol clone split stop` ***`clone_name`***

   The clone-splitting operation halts; the original and FlexClone volumes remain clone partners, but they no longer share the disk space that was duplicated by the split.

6. You can display the status of the newly split FlexVol volume and verify the success of the clone-splitting operation by using the `vol status -v` command.

Related concepts

How splitting a FlexClone volume from its parent works on page 216

Determining the parent volume and base Snapshot copy for a FlexClone volume

You can determine the parent volume and base Snapshot copy for a FlexClone volume by using the `vol status` command.

Determining the space used by a FlexClone volume

You can determine the space used by a FlexClone volume based on its nominal size and the amount of space it shares with the parent FlexVol volume.

About this task

When a FlexClone volume is created, it shares all of its data with its parent volume. Therefore, although the nominal size of the FlexVol volume is the same as its parent's size, it uses very little free space from the aggregate. The free space used by a newly-created FlexClone volume is approximately 0.5% of its nominal size. This space is used to store the FlexClone volume's metadata.

New data written to either the parent or the FlexClone volume is not shared between the volumes. The increase in the amount of new data that gets written to the FlexClone volume leads to an increase in the space the FlexClone volume requires from its containing aggregate.

Steps

1. Determine the nominal size of the FlexClone volume by entering the following command:

   `df -m` ***`FlexClone_name`***

2. Determine the amount of space that is shared between the parent and FlexClone volume by entering the following command:

   `vol clone split estimate` ***`clone_name`***

3. Subtract the size of the shared space from the nominal size of the FlexClone volume to determine the amount of free space being used by the FlexClone volume.

Using FlexClone files and FlexClone LUNs to create efficient copies of files and LUNs

FlexClone files and FlexClone LUNs are writable, space-efficient clones of parent files and parent LUNs, and help in efficient utilization of the physical aggregate space.

FlexClone files and FlexClone LUNs utilize 0.4 percent of their size to store the metadata. Clones share the data blocks of their parent files and parent LUNs and occupy negligible storage space until clients write new data either to the parent file or LUN, or to the clone.

Clients can perform all file and LUN operations on both the parent and the clone entities.

Benefits of FlexClone files and FlexClone LUNs

The process of creating FlexClone files or FlexClone LUNs is highly space-efficient and time-efficient because the cloning operation does not involve physically copying any data.

You can instantaneously create space-efficient copies of your data by using FlexClone files and FlexClone LUNs in situations such as the following:

- When you need to deploy, upgrade, or redeploy thousands of standardized virtual desktops or servers.
- When you need a copy of a database for application development purposes.
- When you need to boot servers in a server farm.
  You can create FlexClone LUNs of the parent boot LUN, then use the FlexClone LUN to boot a server in a server farm.

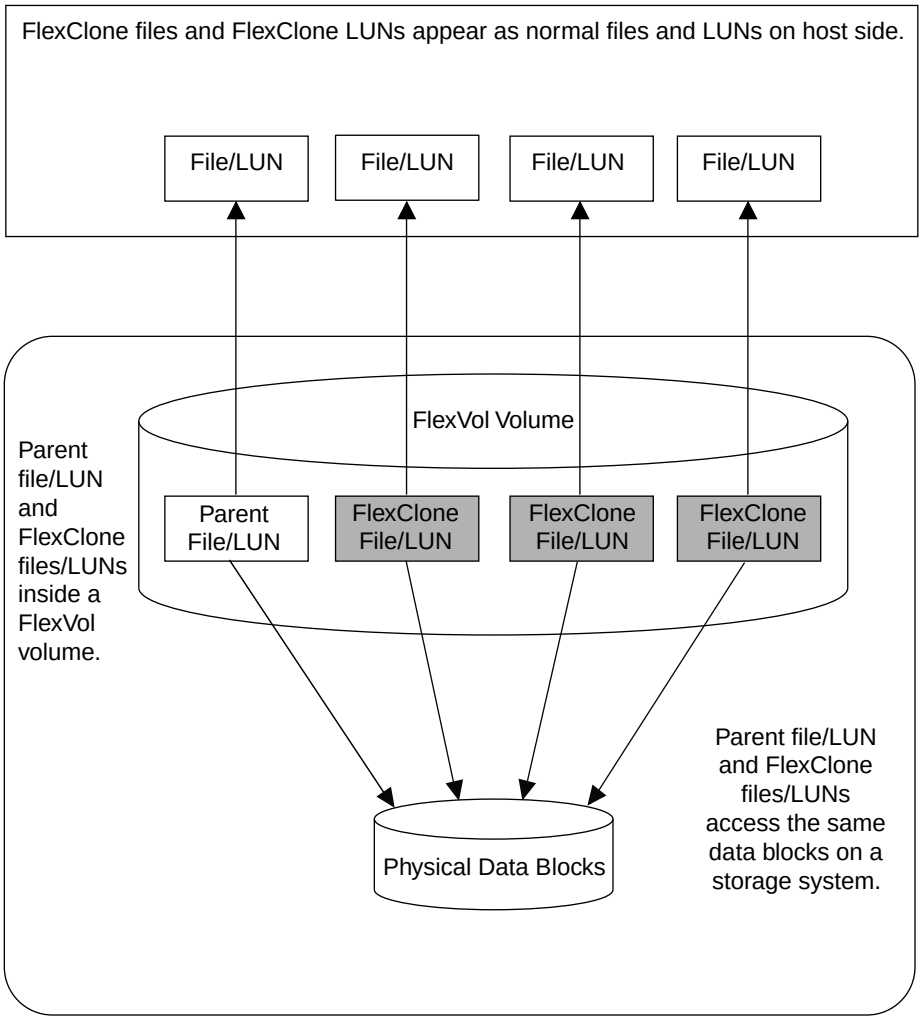
How FlexClone files and FlexClone LUNs work

FlexClone files and FlexClone LUNs share the same physical data blocks with their parent files and LUNs present in FlexVol or FlexClone volumes, and occupy negligible space in the form of metadata.

You can create a clone of a file that is present in a FlexVol volume in a NAS environment, and you can also clone a LUN in a SAN environment.

The cloned copies are highly space-efficient and time-efficient because the cloning operation does not copy physical blocks of data. When you write new data to a parent or clone, then the entity on which new data is written starts occupying extra storage space.

The following illustration shows the parent files or LUNs and FlexClone files or LUNs accessing the same data blocks on the storage system. On the host side, the parent files or LUNs and FlexClone files or LUNs appear as normal files and LUNs:



The cloning operation is instantaneous and has no impact on client access to the parent file or LUN. Clients that are accessing the parent file or LUN do not experience any disruption or outage. Clients can perform all operations on FlexClone files and FlexClone LUNs as they can on standard files and LUNs.

You can create a maximum of 32,767 FlexClone files or FlexClone LUNs from a parent file or LUN without creating a physical copy of the parent entity. If you try to create more than 32,767 clones, Data ONTAP automatically creates a new physical copy of the parent file or LUN.

Note: FlexClone LUNs are not the same as LUN clones. You can create a FlexClone LUN without a backing Snapshot copy of the parent LUN. However, to create a LUN clone, you must have a

backing Snapshot copy of the parent LUN. For more information about LUN clones, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

Considerations for working with FlexClone files and FlexClone LUNs

You should keep several considerations in mind when working with FlexClone files and FlexClone LUNs.

- You can create FlexClone files and LUNs only in the same FlexVol volume containing the parent files and LUNs.
- You can clone a complete file, sub-file, LUN, or sub-LUN.
  To clone a sub-file or sub-LUN, you should know the block range of the parent entity and clone entity.
- The `sis` attribute is added to a FlexVol volume when a FlexClone file or FlexClone LUN is created for the first time.
- When clients write new data either to a FlexClone file or FlexClone LUN, or the parent file or parent LUN, then the new data occupies additional storage space.
- If you delete the FlexClone files or LUNs, the parent files or LUNs are not affected.
  Deleting a parent file or LUN has no impact on the FlexClone files or FlexClone LUNs.
- If you create FlexClone files or LUNs from a Snapshot copy, you cannot create new Snapshot copies until the cloning process is complete.
- If a FlexVol volume containing FlexClone files and LUNs has the fractional reserve set to zero, you must follow extra configuration and management requirements to ensure that the applications accessing the FlexVol volume do not receive errors due to lack of space (ENOSPC).

Related concepts

Considerations for setting fractional reserve on page 268

Creating a FlexClone file or FlexClone LUN

You can create a FlexClone file or a FlexClone LUN of a parent file or LUN inside a FlexVol volume by using the `clone start` command. You can also perform a sub-file or sub-LUN cloning operation.

Before you begin

- FlexClone license is installed on your storage system.
- FlexClone license is installed in an HA pair.
- For creating a sub-file or sub-LUN clone, the block range of the parent entity and the clone entity should be known.

If providing multiple block ranges for cloning, ensure that the block ranges are not overlapping.

Step

1. To create a FlexClone file or FlexClone LUN or to perform a sub-file or sub-LUN cloning operation, choose one of actions from the following table.

If you want to create...	Then...
A FlexClone file or FlexClone LUN of a parent file or LUN inside a FlexVol volume.	Enter the following command: **`clone start src_path dest_path [-n]`** • *`src_path`*—Source path in the `/vol/volname/...` format • *`dest_path`*—Destination path in the `/vol/volname/...` format • `-n`— This option prevented creation of a temporary Snapshot copy of a FlexVol volume during the cloning operation in earlier Data ONTAP versions. This option is supported to maintain consistency with earlier Data ONTAP versions.
A sub-file or sub-LUN clone.	Enter the following command: **`clone start src_path [dest_path] [-n] -r src_fbn:dest_fbn:fbn_cnt ...`** • *`src_path`*—Source path in the `/vol/volname/...` format • *`dest_path`*—Destination path in the `/vol/volname/...` format • `-n`— This option prevented creation of a temporary Snapshot copy of a FlexVol volume during the cloning operation in earlier versions of Data ONTAP. This option is supported to maintain consistency with earlier Data ONTAP versions. • `-r`—Specifies block ranges. • *`src_fbn`*—Starting `fbn` of the source block range. For a LUN, the `fbn` is considered as LBA (Logical block address). • *`dest_fbn`*—Starting `fbn` of the destination block range. • *`fbn_cnt`*—Number of blocks to be cloned.
A FlexClone file or FlexClone LUN of a parent file or LUN inside a FlexVol volume from a Snapshot copy	Enter the following command: **`clone start src_path dest_path[-s <snapshot_name>]`** • *`src_path`*—Source path in the `/vol/volname/...` format • *`dest_path`*—Destination path in the `/vol/volname/...` format. • *`snapshot_name`*—Source Snapshot copy You should use the `-s` option only when you want to create a clone from a Snapshot copy. In addition, you should provide the destination path when you use the `-s` option. Otherwise, the command fails and an error message is displayed.

If you want to create...	Then...
A sub-file or sub-LUN clone from a Snapshot copy.	Enter the following command: **`clone start src_path [dest_path] [-s <snapshot_name>] -r src_fbn:dest_fbn:fbn_cnt ...`** • *src_path*—Source path in the `/vol/volname/...` format • *dest_path*—Destination path in the `/vol/volname/...` format. • *snapshot_name*—Source Snapshot copy You should use the `-s` option only when you want to create a clone from a Snapshot copy. In addition, you should provide the destination path when you use the `-s` option. Otherwise, the command fails and an error message is displayed. • `–r`—Specifies block ranges. • *src_fbn*—Starting `fbn` of the source block range. For a LUN, the `fbn` is considered as LBA (Logical block address). • *dest_fbn*—Starting `fbn` of the destination block range. • *fbn_cnt*—Number of blocks to be cloned.

Note: You cannot use the `clone start` command to create a FlexClone LUN from a LUN clone.

Viewing the space savings due to FlexClone files and FlexClone LUNs

You can view the percentage of disk space saved by block sharing within a volume containing FlexClone files and LUNs.

Step

1. To view the space saving achieved due to FlexClone files and FlexClone LUNs, enter the following command:

   `df -s volname`

   volname is the name of the FlexVol volume.

   Note: If you run the `df -s` command on a deduplication-enabled FlexVol volume, you can view the space saved by both deduplication and FlexClone files and LUNs.

For more information about the `df -s` command, see the df(1) man page.

Example

The following example shows the space saving on a FlexClone volume test1:

```
systemA> df -s test1

Filesystem          used   saved   %saved
/vol/test1/         4828    5744     54%
```

Features supported with FlexClone files and FlexClone LUNs

FlexClone files and FlexClone LUNs work with different Data ONTAP features such as deduplication, Snapshot copies, quotas, and volume SnapMirror.

The following features are supported with FlexClone files and FlexClone LUNs:

- Deduplication
- Snapshot copies
- Access control lists
- vFiler units
- Quotas
- FlexClone volumes
- NDMP
- Volume SnapMirror
- Qtree SnapMirror
- The `volume move` command
- The `volume copy` command
- Space reservation
- HA configuration

Note: Synchronous SnapMirror is not supported on FlexVol volumes with FlexClone files and FlexClone LUNs.

How deduplication works with FlexClone files and FlexClone LUNs

You can efficiently use the physical storage space of the data blocks by creating a FlexClone file or FlexClone LUN of the parent file and parent LUN in a deduplication-enabled volume.

The block-sharing mechanism used by FlexClone files and LUNs is also used by deduplication. You can maximize the space savings in a FlexVol volume by enabling deduplication on the volume and then cloning the deduplication-enabled volume.

Note: While executing the `sis revert_to` command on a deduplication-enabled volume, you cannot create FlexClone files and FlexClone LUNs of the parent files and parent LUNs residing in that volume.

How Snapshot copies work with FlexClone files and FlexClone LUNs

You can create FlexClone files and FlexClone LUNs from an existing Snapshot copy of the parent files and parent LUNs contained in a FlexVol volume. If the base Snapshot copy of a FlexVol volume is deleted, the dependent FlexClone files and FlexClone LUNs are not deleted.

You must consider the following while using a Snapshot copy for creating FlexClone files and FlexClone LUNs:

- If the Snapshot copy gets automatically deleted because of using the `snap autodelete` command, the dependent FlexClone files and FlexClone LUNs are not deleted. However, if the block-sharing on the FlexClone files and FlexClone LUNs is not complete, data might be lost.
- You cannot manually delete a Snapshot copy from which FlexClone files or FlexClone LUNs are being created until the block-sharing process between the parent and clone entities is complete. The Snapshot copy remains locked until the completion of the block-sharing process, which happens in the background. Therefore, when you try deleting a locked Snapshot copy, the system displays a message asking you to retry the operation after some time. In such a situation, if you want to manually delete the particular Snapshot copy, you must keep retrying the deletion operation so that the Snapshot copy gets deleted after the block sharing is complete.

For more information about the automatic and manual deletion of Snapshot copies, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

How access control lists work with FlexClone files and FlexClone LUNs

FlexClone files and FlexClone LUNs inherit the access control lists of their parent files and LUNs.

However, you cannot clone the FlexClone files associated with Windows NT streams.

For more information about access control lists, see the *Data ONTAP File Access and Protocols Management Guide for 7-Mode*.

How vFiler units work with FlexClone files and FlexClone LUNs

You can manage FlexClone files and FlexClone LUNs of parent files and parent LUNs in both the default and nondefault vFiler contexts.

If you are a default vFiler administrator, you can manage the FlexClone files and FlexClone LUNs of all the FlexClone volumes in both the default and the nondefault vFiler units. However, if you are a

nondefault vFiler administrator, you can only manage the FlexClone files and FlexClone LUNs of the FlexClone volumes in that nondefault vFiler unit.

While creating FlexClone files and FlexClone LUNs of FlexClone volumes in a nondefault vFiler unit, you must ensure that the source file path and the destination file path are in the same vFiler unit.

For more information about vFilers, see the *Data ONTAP MultiStore Management Guide for 7-Mode*.

How quotas work with FlexClone files and FlexClone LUNs

Quota limits are applied on the total logical size of the FlexClone files or FlexClone LUNs. Starting from Data ONTAP 8.1, cloning operations will not fail block sharing even if it could cause quotas to exceed.

When you create a FlexClone file or FlexClone LUN, quotas do not recognize any space savings. For example, if you create a FlexClone file of a parent file of 10 GB, you are only using 10 GB of physical space, but the quota utilization is recorded as 20 GB (10 GB for the parent and 10 GB for the FlexClone file).

If the creation of a FlexClone file or LUN would result in the group or user quota's being exceeded, the clone operation succeeds, provided the FlexVol volume has enough space to hold the metadata for the clone. However, the quota for that user or group is oversubscribed.

How FlexClone volumes work with FlexClone files and FlexClone LUNs

You can create a FlexClone volume of a FlexVol volume that has both a FlexClone file and FlexClone LUN and its parent file or LUN in it.

FlexClone files or FlexClone LUNs and their parent files or LUNs that are present in the FlexClone volume continue to share blocks the same way they do in the parent FlexVol volume. In fact, all the FlexClone entities and their parents share the same underlying physical data blocks, minimizing physical disk space usage.

If the FlexClone volume is split from its parent volume, then the FlexClone files or FlexClone LUNs and their parent files or LUNs stop sharing the blocks in the clone of the FlexClone volume. Thereafter they exist as independent files or LUNs. This means that the clone of the volume uses more space than before the splitting operation.

How NDMP works with FlexClone files and FlexClone LUNs

NDMP works at the logical level with FlexClone files and FlexClone LUNs. All FlexClone files or LUNs are backed up as separate files or LUNs.

When you use NDMP services to back up a qtree or a FlexVol volume that contains FlexClone files or FlexClone LUNs, block sharing between parent and clone entities is not preserved, and clone entities are backed up to tape as separate files or LUNs. The space saving is lost. Therefore, the tape onto which you are backing up should have sufficient space to store the expanded amount of data. When you restore, all the FlexClone files and FlexClone LUNs are restored as separate physical files and LUNs. You can enable deduplication on the volume to restore the block-sharing benefits.

Note: When FlexClone files and FlexClone LUNs are being created from an existing Snapshot copy of a FlexVol volume, you cannot back up the volume to tape until the block-sharing process, which happens in the background, is complete. If you use NDMP on the volume when the block-sharing process is in progress, the system displays a message asking you to retry the operation after some time. In such a situation, you must keep retrying the tape backup operation so that it succeeds after the block sharing is complete.

For more information about tape backup, see the *Data ONTAP Data Protection Tape Backup and Recovery Guide for 7-Mode*

How synchronous SnapMirror works with FlexClone files and FlexClone LUNs

You should not use a FlexVol volume that has FlexClone files and FlexClone LUNs as a source for synchronous SnapMirror.

Synchronous SnapMirror is not qualified on a FlexVol volume with FlexClone files or FlexClone LUNs.

For more information about synchronous SnapMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

How volume SnapMirror works with FlexClone files and FlexClone LUNs

Volume SnapMirror used with FlexClone files and FlexClone LUNs helps in maintaining space savings because the cloned entities are replicated only once.

If a FlexVol volume is a volume SnapMirror source and contains FlexClone files or FlexClone LUNs, volume SnapMirror transfers only the shared physical block and a small amount of metadata to the volume SnapMirror destination. The destination stores only one copy of the physical block,

and this block is shared between the parent and cloned entities. Therefore, the destination volume is an exact copy of the source volume and all the clone files or LUNs on the destination volume share the same physical block.

For more information about volume SnapMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

How qtree SnapMirror and SnapVault work with FlexClone files and FlexClone LUNs

Qtree SnapMirror and SnapVault mirror all the FlexClone files and LUNs to the destination volume as individual physical files and LUNs.

The destination FlexVol volume must have enough capacity to store the FlexClone files or LUNs, as separate files or LUNs.

Running deduplication on the destination volume after the qtree SnapMirror and SnapVault transfer is complete reduces the amount of used space on the destination FlexVol volume.

For more information about qtree SnapMirror and SnapVault, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

How volume move affects FlexClone files and FlexClone LUNs

During the cutover phase of a volume move operation, you cannot create FlexClone files or FlexClone LUNs of a FlexVol volume.

For more information about the volume move operation, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

How volume copy works with FlexClone files and FlexClone LUNs

You can perform a volume copy operation on a FlexVol volume that has FlexClone files and FlexClone LUNs in it.

After the volume copy operation is done, the FlexClone files and FlexClone LUNs and their parents on the destination FlexVol volume share the same data blocks as they did on the source FlexVol volume.

For more information about the `vol copy` command, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

How space reservation works with FlexClone files and FlexClone LUNs

FlexClone files and FlexClone LUNs inherit the space reservation attribute from the parent file and parent LUN.

FlexClone files and FlexClone LUNs inherit the space reservation settings of the parent file and the parent LUN. Therefore, if there is not enough space in the FlexVol volume to create a FlexClone LUN with the same space reservation as that of the parent, then the cloning operation fails.

Note: The space required according to space reservation attribute is separate for the parent LUN and the FlexClone LUN.

How an HA configuration works with FlexClone files and FlexClone LUNs

FlexClone file and FlexClone LUN operations are supported in an HA configuration.

In an HA pair, you cannot create FlexClone files or FlexClone LUNs on the partner while the takeover or giveback operation is in progress. All the pending block sharing operations on the partner are resumed after the takeover or giveback operation is complete.

Using deduplication and data compression to increase storage efficiency

You can run deduplication and data compression together or independently on a FlexVol volume to achieve optimal space savings. Deduplication eliminates the duplicate data blocks and data compression compresses the data blocks to reduce the amount of physical storage required.

How to set up efficiency operations

Depending on your storage environment setup, you can first estimate the space savings that can be achieved and then configure deduplication and data compression or only deduplication. You can run the efficiency operations on a volume by using schedules.

You can use the space savings estimation tool to estimate the savings you can achieve in an existing environment. The space savings estimation tool can evaluate a maximum of 2 TB of data. You can download the space savings estimation tool from *communities.netapp.com/docs/DOC-18699*.

Configuring deduplication

Deduplication is a Data ONTAP feature that reduces the amount of physical storage space required by eliminating duplicate data blocks within a FlexVol volume. You should not enable deduplication on the root volume.

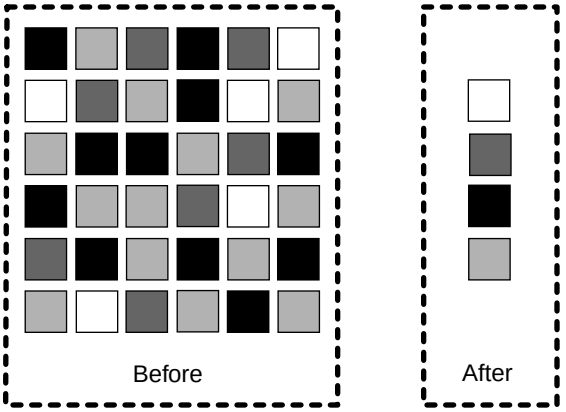
You can decide to deduplicate only the new data that is written to the volume after enabling deduplication or both the new data and the data existing in the volume prior to enabling deduplication.

How deduplication works

Deduplication operates at the block level within the entire FlexVol volume, eliminating duplicate data blocks, and storing only unique data blocks.

Each block of data has a digital signature that is compared with all other signatures in a data volume. If an exact block signature match exists, a byte-by-byte comparison is done for all the bytes in the block, and the duplicate block is discarded and its disk space is reclaimed.

Deduplication removes data redundancies, as shown in the following illustration:



Data ONTAP writes all data to a storage system in 4-KB blocks. When deduplication runs for the first time on a volume with existing data, it scans all the blocks in the volume and creates a digital fingerprint for each of the blocks. Each of the fingerprints is compared to all the other fingerprints within the volume. If two fingerprints are found to be identical, a byte-by-byte comparison is done for all data within the block. If the byte-by-byte comparison detects identical data, the pointer to the data block is updated, and the duplicate block is removed.

Note: When deduplication is run on a volume with existing data, it is best to configure deduplication to scan all the blocks in the volume for better space savings.

Deduplication runs on the active file system. Therefore, as additional data is written to the deduplicated volume, fingerprints are created for each new block and written to a change log file. For subsequent deduplication operations, the change log is sorted and merged with the fingerprint file, and the deduplication operation continues with fingerprint comparisons as previously described.

What deduplication metadata is

The deduplication metadata includes the fingerprint file and change logs. Fingerprints are the digital signatures for every 4-KB data block in a FlexVol volume.

The deduplication metadata contains two change log files. When deduplication is running, the fingerprints of the new data blocks from one change log file are merged into the fingerprint file, and the second change log file stores the fingerprints of the new data that is written to the volume during the deduplication operation. The roles of the change log files are reversed when the next deduplication operation is run.

In Data ONTAP 8.0.1, the deduplication metadata is located within the aggregate. Starting with Data ONTAP 8.1, two copies of deduplication metadata are maintained per volume. A copy of the deduplication metadata resides in the volume and another copy is in the aggregate. The deduplication metadata in the aggregate is used as the working copy for all the deduplication operations. An additional copy of the deduplication metadata resides in the volume.

When a volume is moved, the deduplication metadata is also transferred with the volume. If the volume ownership changes, the next time deduplication is run, then the deduplication metadata

which resides in the aggregate is created automatically by using the copy of deduplication metadata in the volume. This method is a faster operation than creating a new fingerprint file.

Starting with Data ONTAP 8.2, the fingerprints are stored for each physical block, this reduces the amount of space required to store the deduplication metadata.

Deduplication metadata can occupy up to 7 percent of the total physical data contained within the volume, as follows:

- In a volume, deduplication metadata can occupy up to 4 percent of the total amount of data contained within the volume.
- In an aggregate, deduplication metadata can occupy up to 3 percent of the total physical data contained within the volume.

You can use the `df -A aggrname` command to check the available space in an aggregate and the `df path` command to check the available space in a volume. For more information about these commands, see the man pages.

Example

A 2 TB aggregate has four volumes, each 400 GB in size, in the aggregate. You need three volumes to be deduplicated with varying savings percentage on each volume.

The space required in the different volumes for deduplication metadata is as follows:

- 2 GB [4% × (50% of 100 GB)] for a 100 GB of logical data with 50 percent savings
- 6 GB [4% × (75% of 200 GB)] for a 200 GB of logical data with 25 percent saving
- 3 GB [4% × (25% of 300 GB)] for a 300 GB of logical data with 75 percent savings

The aggregate needs a total of 8.25 GB [(3% × (50% of 100 GB)) + (3% × (75% of 200 GB)) + (3% × (25% of 300 GB)) = 1.5+4.5+2.25= 8.25 GB] of space available in the aggregate for deduplication metadata.

Guidelines for using deduplication

Deduplication runs as a system operation and consumes system resources when the deduplication operation is running on FlexVol volumes.

If the data does not change often in a volume, it is best to run deduplication less frequently. If you run multiple concurrent deduplication operations on a storage system, these operations lead to a higher consumption of system resources. It is best to begin with fewer concurrent deduplication operations. Increasing the number of concurrent deduplication operations gradually enables you to better understand the impact on the system.

Performance considerations for deduplication

Various factors affect the performance of deduplication. You should check the performance impact of deduplication in a test setup, including sizing considerations, before deploying deduplication in performance-sensitive or production environments.

The following factors can affect the performance of deduplication:

- The data access pattern (for example, sequential versus random access, the size, and pattern of the input and output)
- The amount of duplicate data, the amount of total data, and the average file size
- The nature of data layout in the volume
- The amount of changed data between deduplication operations
- The number of concurrent deduplication operations
- Hardware platform (system memory and CPU module)
- Load on the system
- Disk types (for example, ATA/FC, and RPM of the disk)

For more information about performance aspects of deduplication, see *Technical Report 3958: Data Compression and Deduplication Deployment and Implementation Guide for 7-Mode*.

Related information

Data Compression and Deduplication Deployment and Implementation Guide for 7-Mode: media.netapp.com/documents/tr-3958.pdf

Enabling deduplication on a volume

You can enable deduplication on a FlexVol volume to achieve storage efficiency by using the `sis on` command.

Before you begin

For a FlexVol volume, you must have verified that enough free space exists for deduplication metadata in the volumes and aggregates.

Step

1. Enable deduplication by entering the following command:
   `sis on` ***`path`***

   `path` is the complete path to the FlexVol volume.

Example

The following command enables deduplication on the volume VolA:

```
sis on /vol/VolA
```

Disabling deduplication on a volume

You can disable deduplication on a volume by using the `sis off` command.

About this task

If you have enabled data compression on the volume, running the `sis off` command disables data compression.

Steps

1. Stop any volume efficiency operation that is currently active on the volume by entering the following command:

   `sis stop` *`path`*

   `path` is the complete path to the FlexVol volume.

2. Disable the deduplication operation by entering the following command:

   `sis off` *`path`*

   `path` is the complete path to the FlexVol volume.

   This command stops all future deduplication operations. For more information about the sis command, see the sis(1) man page.

Example

The following command disables deduplication on the volume VolA:

```
sis off /vol/VolA
```

Configuring data compression

Data compression is a Data ONTAP feature that enables you to reduce the physical capacity that is required to store data on storage systems by compressing the data blocks within a FlexVol volume. You can use data compression only on volumes contained within 64-bit aggregates.

You can use data compression on primary, secondary, and tertiary storage tiers.

How data compression works

Data compression enables you to store more data in less space. Further, you can use data compression to reduce the time and bandwidth required to replicate data during volume SnapMirror transfers. Data compression can save space on regular files or LUNs.

However, storage system internal files, Windows NT streams, and volume metadata are not compressed.

Data compression works by compressing a small group of consecutive blocks known as a compression group. Data compression can be done in the following ways:

- Inline compression
  If inline compression is enabled on a volume, during subsequent data writes the compressible data is compressed and written to the volume. However, data which cannot be compressed or data bypassed by inline compression is written in the uncompressed format to the volume.
- Postprocess compression
  If postprocess compression is enabled on a volume, the new data writes to the volume which were not compressed initially (if inline compression is enabled), are rewritten as compressed data to the volume when postprocess compression is run. The postprocess compression operation runs as a low-priority background process.

If both inline and postprocess compression are enabled, then postprocess compression compresses only the blocks on which inline compression was not run. This includes blocks that were bypassed by inline compression such as small, partial compression group overwrites.

Note: You cannot enable data compression on the storage system root volumes or on the volumes that are contained within 32-bit aggregates.

How data compression detects incompressible data

Incompressible data detection allows you to check if a file is compressible and for large size files, you can check if a compression group within a file is compressible. Allowing incompressible data to be detected saves the system resources used by inline compression trying to compress incompressible files or compression groups.

For files with size less than 500 MB, inline compression checks if a compression group can be compressed. If incompressible data is detected within a compression group, then a flag is set for the file containing the compression group to indicate that the file is incompressible. During subsequent compression attempts, inline compression first checks if the incompressible data flag is set for the file. If the flag is set, then inline compression is not attempted on the file.

For files with size equal to or greater than 500 MB, inline compression performs a quick check on the first 4 KB block of each compression group to determine if it can be compressed. If the 4 KB block cannot be compressed, the compression group is left uncompressed. However, if compression of the 4 KB block is successful, then compression is attempted on the whole compression group.

Postprocess compression runs on all files irrespective of whether the file is compressible or not. If postprocess compression compresses at least one compression group in an incompressible file, then the incompressible data flag for that file is cleared. During the next compression attempt, inline compression can run on this file to achieve space savings.

For more information about enabling or disabling incompressible data detection and modifying the minimum file size to attempt quick check on a file, see the `sis config` command man page.

Enabling data compression on a volume

You can enable data compression only on FlexVol volumes that are contained within 64-bit aggregates. You can enable data compression to achieve space savings by using the `sis config` command.

Before you begin

You must have enabled deduplication on the volume.

About this task

You can enable both inline and postprocess compression or only postprocess compression on a volume. Before enabling inline compression, you must enable postprocess compression on the volume.

Step

1. Enable data compression by entering the following command:

   `sis config [-C {true|false}] | [-I {true|false}] | [-Q {true|false}] | [-z <file_size>]` ***`path`***

   `true` value for `-C` option enables postprocess compression.

   `true` value for `-I` option enables inline compression.

   `true` value for `-Q` option enables incompressible data detection.

   `file size` for `-z` option specifies the minimum file size to attempt quick check on a file.

   `path` is the complete path to the FlexVol volume.

Examples

The following command enables postprocess compression on the volume VolA:

```
sis config -C true /vol/VolA
```

The following command enables both postprocess and inline compression on the volume VolA:

```
sis config -C true -I true /vol/VolA
```

Disabling data compression on a volume

You can disable data compression on a FlexVol volume by using the `sis config` command.

About this task

If you want to disable postprocess compression, you must first disable inline compression on the volume.

Steps

1. Stop any volume efficiency operation that is currently active on the volume by entering the following command:

   `sis stop` ***`path`***

   `path` is the complete path to the FlexVol volume.

2. Disable data compression by entering the following command:

   `sis config [-C {true|false}] [-I {true|false}]` ***`path`***

   `false` value for `-C` disables background compression.

   `false` value for `-I` disables in-line compression.

   `path` is the complete path to the FlexVol volume.

Examples

The following command disables inline compression on the volume VolA:

```
sis config -I false /vol/VolA
```

The following command disables both postprocess and inline compression on the volume VolA:

```
sis config -C false -I false /vol/VolA
```

Managing volume efficiency operations using schedules

You can manage how the efficiency operations run on a volume by running efficiency operations manually, or using a schedule, or depending on the amount of new data written to the volume.

You can also control how the efficiency operations run based on the following conditions:

- Use checkpoints or not
- Run efficiency operations on existing data or only new data
- Not run efficiency operations at the end of each SnapVault transfer
- Stop efficiency operations if required

You can use the `sis config` command to view the schedule assigned to the volumes.

Modifying scheduling of efficiency operations

You can modify the scheduling of deduplication or data compression operation on a volume by using the `sis config -s` command.

Step

1. Modify the scheduling of deduplication or data compression operation on a volume by entering the following command:

   `sis config -s` *`schedule path`*

   `schedule` lists the days and hours of the day when the efficiency operations run. *`schedule`* can be one of the following types:

   - *`day_list`*[*@`hour_list`*]
     If *`hour_list`* is not specified, the efficiency operation runs at midnight on each scheduled day.
   - *`hour_list`*[*@`day_list`*]
     If *`day_list`* is not specified, the efficiency operation runs every day at the specified hours.
   - `-`
     If a hyphen (`-`) is specified, the scheduled efficiency operations are turned off on the specified volume.

     Note: When the efficiency operation schedule is turned off and additional data is written to the volume, the change log file size increases. This results in increase in deduplication metadata. You can schedule deduplication to be run periodically to avoid the increase in deduplication metadata.
   - `auto`
     The `auto` option can be used to run an efficiency operation on a volume depending on the amount of new data that is written to the volume after the previous efficiency operation.
   - `manual`
     The `manual` option can be used only on SnapVault destination volumes. This option ensures that deduplication or post-process compression is not triggered automatically after every update to the destination volumes.

   `path` is the complete path to the volume.

Examples

The following command modifies the scheduling of efficiency operation for VolA to run at 11 p.m., Monday through Friday:

`sis config -s mon-fri@23 /vol/VolA`

Running efficiency operations manually

You can run efficiency operations manually on a FlexVol volume by using the `sis start` command.

Before you begin

Depending on the efficiency operation you want to run manually, you must have enabled deduplication or both data compression and deduplication.

About this task

If only deduplication is enabled on a volume, then deduplication runs on the data. However, if deduplication and data compression are enabled on a volume, data compression is run initially followed by deduplication.

Deduplication is a background process that consumes system resources while it is running. If the data does not change often in a volume, it is best to run deduplication less frequently. Multiple concurrent deduplication operations running on a storage system lead to a higher consumption of system resources.

You can run a maximum of eight concurrent deduplication or data compression operations per node. If any more efficiency operations are scheduled, the operations are queued.

Step

1. Run an efficiency operation by entering the following command:

   `sis start [-s [-m][-o][-p]] [-f] [-d] [-q]` ***`path`***

   The `-s` option is used to scan the data that existed in the volume prior to enabling deduplication. You are prompted to confirm whether deduplication or data compression, if enabled, should be started on the volume.

   The `-m` option can be used to scan the existing data, generate deduplication metadata, and skip sharing. This option is used only with `-s` option.

   The `-o` option can be used to scan the existing data and skip shared block optimization. This option is used only with `-s` option.

   The `-p` enables the efficiency operation to use the previous checkpoint. This option is used only with `-s` option.

   The `-f` option can be used to start the efficiency operation without any prompts.

   The `-d` starts a new efficiency operation after deleting the previous checkpoint.

   The `-q` option queues the efficiency operation.

   `path` is the complete path to the volume.

Examples

The following command allows you to manually start only deduplication or data compression followed by deduplication on the volume VolA:

```
sis start /vol/VolA
```

The following command allows you to manually run the efficiency operations on the existing data in the volume VolA:

```
sis start -s /vol/VolA
```

Running efficiency operations depending on the amount of new data written

You can modify the efficiency operation schedule to run deduplication or data compression when the number of new blocks written to the volume after the previous efficiency operation (performed manually or scheduled) exceeds a specified threshold percentage.

About this task

If the `schedule` option is set to `auto`, the scheduled efficiency operation runs when the amount of new data exceeds the specified percentage. The default threshold value is 20 percent. This threshold value is the percentage of the total number of blocks already processed by the efficiency operation.

Step

1. Modify the threshold percentage value by entering the following command:

   `sis config -s auto@`*`num path`*

   `num` is a two-digit number to specify the percentage.

   `path` is the complete path to the volume.

Example

The following command modifies the threshold percentage value to 30 percent for the volume VolA:

```
sis config -s auto@30 /vol/VolA
```

Using checkpoints to resume efficiency operation

You can use the checkpoints to periodically log the execution process of an efficiency operation. When an efficiency operation is stopped for any reason (such as system halt, system disruption,

reboot, or because last efficiency operation failed or stopped) and checkpoint data exists, an efficiency operation can resume from the latest checkpoint file.

A checkpoint is created at the end of each stage or substage of the efficiency operation. When an efficiency operation scans the existing data in a volume, a checkpoint is created approximately every hour.

Resuming an efficiency operation using the checkpoint option

You can resume an efficiency operation by using the `sis start` command with the checkpoint option.

About this task

If only deduplication is enabled on the volume, deduplication runs on the data. However, if deduplication and data compression are enabled, data compression runs on the existing data followed by deduplication.

You can view the details of the checkpoint for a volume by using the `sis status` command.

By default the efficiency operations resume from checkpoints. However, if a checkpoint corresponding to a previous efficiency operation (the phase when the `sis start -s` command is run) is older than 24 hours, then the efficiency operation does not resume from the previous checkpoint automatically. In this case, the efficiency operation starts from the beginning. However, if you know that significant changes have not occurred in the volume since the last scan, you can force continuation from the previous checkpoint by using the `-p` option.

Step

1. Resume an efficiency operation by using the checkpoint option by entering the following command:

   `sis start -sp` *`path`*

   The `-sp` option enables you to resume an efficiency operation on the volume by using a checkpoint, even when the validity of the checkpoint has expired.

   `path` is the complete path to the volume.

Example

The following command enables you to resume an efficiency operation by using the checkpoint option on the volume VolA:

```
sis start -sp /vol/VolA
```

Running efficiency operations manually on existing data

You can manually run the efficiency operations on the data that exists in the FlexVol volume prior to enabling deduplication or data compression. Deduplication or data compression followed by deduplication can be run by using the `sis start -s` command.

About this task

If only deduplication is enabled on a volume, then deduplication runs on the data. However, If deduplication and data compression are enabled on a volume, data compression is run initially followed by deduplication.

When you run data compression on existing data, by default the data compression operation skips the data blocks that are shared by deduplication and the data blocks that are locked by Snapshot copies.

You can run a maximum of eight deduplication or data compression operations concurrently per node and the remaining operations are queued.

Step

1. Run deduplication or data compression followed by deduplication manually on the existing data by entering the following command:

   sis start -s *path*

   path is the complete path to the volume.

Example

The following command allows you to manually run deduplication or data compression followed by deduplication on the existing data in the volume VolA:

```
sis start -s /vol/VolA
```

Monitoring volume efficiency operations

You can monitor the progress of efficiency operations on a volume by viewing the status of the efficiency operations and the space savings achieved on the volume.

Viewing the status of efficiency operations on a FlexVol volume

You can view whether deduplication or data compression is enabled on a volume and check the status, state, and progress of the efficiency operations on a volume by using the `sis status` command.

Step

1. View the status of an efficiency operation on a volume by entering the following command:

   `sis status` ***`path`***

   `path` is the complete path to the volume.

   The following table describes status and progress messages that you might see after running the `sis status` command:

Message	Message type	Description
Idle	status and progress	No active efficiency operation is in progress.
Pending	status	The limit of maximum concurrent efficiency operations allowed for a storage system or a vFiler unit is reached. Any efficiency operation requested beyond this limit is queued.
Active	status	Efficiency operations are running.
`size` Scanned	progress	A scan of the entire volume is running, of which *`size`* is already scanned.
`size` Compressed	progress	Compression operations have compressed data, of which *`size`* is already compressed.
`size` Searched	progress	A search of duplicated data is running, of which *`size`* is already searched.
`size` (*`pct%`*) Done	progress	Efficiency operations have saved *`size`* amounts of data. *`pct%`* is the percentage saved of the total duplicated data that was discovered in the search stage.

Message	Message type	Description
size Verified	progress	A verification of the metadata of processed data blocks is running, of which *size* is already verified.
pct% merged	progress	Deduplication operations have merged *pct%* (percentage) of all the verified metadata of processed data blocks to an internal format that supports fast deduplication operations.

Example

The following command displays the status of an efficiency operation on volume VolA:

sis status /vol/VolA

If the efficiency operation is enabled on volume VolA and the operation is idle, then you can see the following in the system output:

```
node1> sis status /vol/VolA
Path              State      Status     Progress
/vol/VolA         Enabled    Active     23 MB Scanned, 20 MB Compressed
```

Viewing efficiency space savings on a FlexVol volume

You can view the amount of space savings achieved through deduplication and data compression on a volume by using the `df -S` command.

About this task

The space savings in Snapshot copies are not included when calculating the space savings achieved on a volume. Using deduplication does not affect volume quotas. Quotas are reported at the logical level, and remain unchanged.

Step

1. View the space savings achieved on a volume using deduplication and data compression by entering the following command:

   df -S path

   path is the complete path to the volume.

Example

The following command enables you to view the space savings achieved by using deduplication and data compression on volume VolA:

df -S /vol/VolA

```
node1> df -S /vol/VolA
Filesystem  used  total-saved  %total-saved deduplicated
%deduplicated compressed %compressed
/vol/vol1/   236         2028           90%          244
51%       1784         88%
```

Stopping volume efficiency operations

You can stop a deduplication or postprocess compression operation by using the `sis stop` command. This command automatically generates a checkpoint.

Step

1. You can stop an active deduplication or postprocess compression operation by entering the following command:

   sis stop *path*

   path is the complete path to the volume.

   If you specify the `-all` option, active and queued efficiency operations are aborted.

Examples

The following command stops the deduplication or postprocess compression operation that is currently active on volume VolA:

sis stop /vol/VolA

The following command aborts both active and queued deduplication or postprocess compression operations on volume VolA:

sis stop -all /vol/VolA

Information about removing space savings from a volume

You can choose to remove the space savings achieved by running efficiency operations on a volume. You must ensure that you contact technical support before removing or undoing the space savings on a volume.

For more information about removing space savings from a volume, see the technical report *TR-3958: Data Compression and Deduplication Deployment and Implementation Guide for 7-Mode*.

Related information

Data Compression and Deduplication Deployment and Implementation Guide for 7-Mode: media.netapp.com/documents/tr-3958.pdf

Deduplication interoperability with Data ONTAP features

When you use deduplication, you should be aware of the features supported by deduplication and how they work with deduplication.

The following features are supported by deduplication:

- Snapshot copies
- Volume SnapMirror
- Qtree SnapMirror
- SnapVault
- SMTape backup
- SnapRestore
- Stretch and fabric-attached MetroCluster configurations
-
- Volume copy
- Data compression
- FlexClone volumes
- HA pair
- vFiler units
- DataMotion for Volumes

You can enable extents on deduplicated volumes. You can perform read reallocation to improve the file layout and the sequential read performance of a deduplicated volume.

However, deduplicated volumes cannot be replicated using synchronous SnapMirror and semi-synchronous SnapMirror.

How fractional reserve works with deduplication

If you are using deduplication for a volume with a fractional reserve setting of 0, there are additional configuration requirements if you need to ensure that your applications never receive an ENOSPC (out of space) error. For more information, see the documentation on setting fractional reserve.

Related concepts

Considerations for setting fractional reserve on page 268

How Snapshot copies work with deduplication

You can run deduplication only on the active file system. However, this data can get locked in Snapshot copies created before you run deduplication, resulting in reduced space savings.

To avoid conflicts between deduplication and Snapshot copies, you should follow these guidelines:

- Run deduplication before creating new Snapshot copies.
- Remove unnecessary Snapshot copies stored in deduplicated volumes.
- Reduce the retention time of Snapshot copies stored 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 snap reserve is 0, you should turn off the schedule for automatic creation of Snapshot copies (which is the case in most LUN deployments).

How volume SnapMirror works with deduplication

You can use volume SnapMirror to replicate a deduplicated volume regardless of size of the volume and logical data in the volume.

When using volume SnapMirror with deduplication, you must consider the following information:

- You can enable deduplication on the source system, the destination system, or both systems.
- The shared blocks are transferred only once. Therefore, deduplication also reduces the use of network bandwidth.
- When the volume SnapMirror relationship is broken, the default deduplication schedule is applied at the destination storage system.

When configuring volume SnapMirror and deduplication, you should coordinate the deduplication schedule and the volume SnapMirror schedule. You should start volume SnapMirror transfers of a deduplicated volume after the deduplication operation is complete. This schedule prevents the sending of undeduplicated data and additional temporary metadata files over the network. If the temporary metadata files in the source volume are locked in Snapshot copies, these files consume extra space in the source and destination volumes.

How qtree SnapMirror works with deduplication

You can use deduplication for volumes that use qtree SnapMirror. Qtree SnapMirror does not automatically initiate a deduplication operation at the completion of every individual qtree SnapMirror transfer. You can set up a deduplication schedule independent of your qtree SnapMirror transfer schedule.

When using qtree SnapMirror with deduplication, you must consider the following information:

- You can enable deduplication on the source system, the destination system, or both systems.
- Even when deduplication is enabled on the source system, duplicate blocks are sent to the destination system. Therefore, no network bandwidth savings is achieved.

- To recognize space savings on the destination system, you should run deduplication on the destination after the qtree SnapMirror transfer is complete.
- Qtree SnapMirror recognizes deduplicated blocks as changed blocks. Therefore, when you run deduplication on an existing qtree SnapMirror source system for the first time, all the deduplicated blocks are transferred to the destination system. This process might result in a transfer several times larger than the regular transfers.

When using qtree SnapMirror with deduplication, you should ensure that qtree SnapMirror uses only the minimum number of Snapshot copies that it requires. To ensure this minimum, you should retain only the latest Snapshot copies.

How SnapVault works with deduplication

The deduplication feature is integrated with the SnapVault secondary license. This feature increases the efficiency of data backup and improves the use of secondary storage.

The behavior of deduplication with SnapVault is similar to the behavior of deduplication with qtree SnapMirror, with the following exceptions:

- Deduplication is also supported on the SnapVault destination volume.
- The deduplication schedule depends on the SnapVault update schedule on the destination system. However, the deduplication schedule on the source system does not depend on the SnapVault update schedule, and it can be configured independently on a volume. You can set manual schedules on a SnapVault destination volume.
- Every SnapVault update (baseline or incremental) starts a deduplication process on the destination system after the archival Snapshot copy is taken.
- A new Snapshot copy replaces the archival Snapshot copy after deduplication finishes running on the destination system. (The name of this new Snapshot copy is the same as that of the archival copy, but the Snapshot copy uses a new timestamp, which is the creation time.)
- You cannot configure the deduplication schedule on the destination system manually or run the `sis start` command. However, you can run the `sis start -s` command on the destination system.
- The SnapVault update does not depend on the deduplication operation. A subsequent incremental update is allowed to continue while the deduplication operation on the destination volume from the previous backup is still in progress. In this case, the deduplication operation continues; however, the archival Snapshot copy is not replaced after the deduplication operation is complete.
- The SnapVault update recognizes the deduplicated blocks as changed blocks. Thus, when deduplication is run on an existing SnapVault source for the first time, all saved space is transferred to the destination system. The size of the transfer might be several times larger than the regular transfers. Running deduplication on the source system periodically will help prevent this issue for future qtree SnapMirror transfers. You should run deduplication before the SnapVault baseline transfer.

  Note: You can run a maximum of eight concurrent deduplication operations on a system. This number includes the deduplication operations linked to SnapVault volumes and those that are not linked to SnapVault volumes.

How tape backup works with deduplication

When you backup to a tape through the SMTape engine, the data format of the source volume is preserved on the tape and all deduplication savings are retained on the tape. The deduplication savings are preserved on the volume to which the data is restored from the tape.

When you backup to a tape through NDMP, the deduplication savings are not preserved on the tape.

How SnapRestore works with deduplication

The metadata created during a deduplication operation is located both in the FlexVol volume and in the aggregate. Therefore, when you initiate a SnapRestore operation on a volume, the metadata is restored to the volume and the restored data retains the original space savings.

After a SnapRestore operation is completed, if deduplication is enabled on the volume, any new data written to the volume continues to be deduplicated.

How MetroCluster configurations work with deduplication

Deduplication is supported for stretch and fabric-attached MetroCluster configurations.

When deduplication is enabled for MetroCluster configurations, you must take into account the following information:

- Deduplication impacts the CPU resources because MetroCluster configurations write to two plexes.
  This results in extra disk write operations and additional system resources are required.
- In takeover mode deduplication operations continue to run as per the schedule.
- A node in takeover mode services the I/O modules targeted by the partner FlexVol volumes.

How works with deduplication

Deduplication is supported with the NetApp Management Console data protection capability, the NetApp Management Console provisioning capability, and Operations Manager in .

Deduplication and the NetApp Management Console data protection capability in

In releases earlier than , the NetApp Management Console data protection capability waits for an active deduplication operation to complete, before renaming the Snapshot copies. While the NetApp Management Console data protection capability waits, it does not allow clients to list the Snapshot copies or restore from them. Therefore, in releases prior to , the use of deduplication with the NetApp Management Console data protection capability is not optimal.

However, this limitation is removed in .

Deduplication and the NetApp Management Console provisioning capability in

With the NetApp Management Console provisioning capability in , you can enable the provisioning policies to support all three modes of deduplication, namely, on-demand deduplication, automated deduplication, and scheduled deduplication.

For more information about using deduplication with the NetApp Management Console provisioning capability and the NetApp Management Console data protection capability, see the *Provisioning Manager and Protection Manager Guide to Common Workflows for Administrators*.

Deduplication and Operations Manager in

You can perform deduplication operations from Operations Manager in .

You can configure deduplication on the system and generate reports or graphs summarizing space savings for file and LUN clones.

For more information about using deduplication with Operations Manager, see the *Operations Manager Administration Guide*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How volume copy works with deduplication

When deduplicated data is copied by using the `vol copy` command, the copy of the data at the destination inherits all the deduplication attributes and storage savings of the source data.

The metadata created during a deduplication operation (fingerprint files and change log files) is located inside the FlexVol volume and in the active copy within the aggregate. Therefore, when you run the volume copy operation on a FlexVol volume, the fingerprint files and change log files are copied to the destination volume. After a volume copy operation, if deduplication is enabled on the destination volume, any new data written to the volume continues to be deduplicated.

The first deduplication operation that is run on the copied FlexVol volume automatically reconstructs the deduplication metadata in the aggregate based on the copy that is available in the volume.

How deduplication works with data compression

When both data compression and deduplication are enabled on a FlexVol volume, the data is first compressed and then deduplicated. Depending on the type of data, the combined savings can yield higher savings than running deduplication alone.

How FlexClone volumes work with deduplication

Deduplication is supported on FlexClone volumes. The FlexClone volume of a deduplicated volume is a deduplicated volume. The cloned volume inherits the deduplication configuration of the parent volume (for example, deduplication schedules).

The metadata created during a deduplication operation (fingerprint files and change log files) is cloned. This metadata is located both in the FlexVol volume and in the aggregate.

If you run deduplication on a clone volume, the clone is deduplicated, but the parent volume remains non-deduplicated.

To run deduplication manually for all new data in the cloned volume, you should use the `sis start` command.

When a cloned volume is split from the parent volume, deduplication of all data in the clone that was part of the parent volume is removed after the volume-split operation. However, if deduplication is running on the clone volume, the data is deduplicated in the subsequent deduplication operation.

How HA pairs work with deduplication

Starting with Data ONTAP 8.1, deduplication operations can be run on volumes from either of the nodes during takeover in an HA pair. The maximum number of concurrent deduplication operations allowed on each node of an HA pair is eight.

If one of the nodes fails, the other node takes over the deduplication operations managed by the failed node. In takeover mode, the working node continues with the deduplication operations. The working node can start deduplication operations on volumes that belong to the failed node. When the working node is managing the deduplication operations on volumes that belong to both the nodes, the maximum number of concurrent deduplication operations is still eight.

How vFiler units work with deduplication

Deduplication commands are available in all the vfiler contexts. Deduplication support on vFiler units allows users to reduce redundant data blocks within vFiler units.

How to set the maximum deduplication sessions per vFiler unit

You can specify the number of concurrent deduplication sessions that can be run per vFiler unit by using the `option sis.max_vfiler_active_ops` command.

Note: The maximum number of concurrent deduplication operations per storage system is eight. However, on a 32-bit platform, the maximum number of concurrent deduplication operations per storage system is five. The `option sis.max_vfiler_active_ops` command first checks the deduplication sessions on the physical storage system, and then on the vFiler unit.

How DataMotion for Volumes works with deduplication

Deduplication savings on a FlexVol volume are retained when the volume is moved by using the DataMotion for Volumes (volume move) operation. If deduplication operations are running when a volume move operation is active, then these operations are stopped shortly before the final cutover is complete.

After the volume move is complete, the efficiency operations cannot be resumed from the previous checkpoint and the efficiency operations start from the beginning.

If you try to nondisruptively move a FlexVol volume that has deduplication operations running, then the deduplication operation is aborted.

For more information about DataMotion for Volumes, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

Data compression interoperability with Data ONTAP features

When you use data compression, you should be aware of the features supported by data compression and how they work with data compression.

The following features are supported by data compression:

- Snapshot copies
- Volume SnapMirror
- Qtree SnapMirror
- SnapVault
- SMTape backup
- SnapLock
- Volume-based SnapRestore
- Single file SnapRestore
- Stretch and fabric-attached MetroCluster configurations
- Volume copy
- Aggregate copy
- Deduplication
- FlexClone volumes
- FlexClone files
- HA pair
- Performance Acceleration Module or Flash cache cards
- vFiler units
- DataMotion for Volumes
- Flash Pool

Compressed volumes cannot be replicated using synchronous SnapMirror and semi-synchronous SnapMirror. Read reallocation and extents are not supported on compressed volumes.

How fractional reserve works with data compression

If you are using data compression for a volume with a fractional reserve setting of 0, there are additional configuration requirements if you need to ensure that your applications never receive an

ENOSPC (out of space) error. For more information, see the documentation on setting fractional reserve.

Related concepts

Considerations for setting fractional reserve on page 268

How Snapshot copies work with data compression

When you run data compression in the default mode after a Snapshot copy is created, the existing data that is locked by the Snapshot copy is compressed.

Snapshot copies lock blocks of data that cannot be freed until the Snapshot copy expires or is deleted. On any volume on which data compression is enabled, when a Snapshot copy of the data is created, any subsequent changes to the data temporarily requires additional disk space, until the Snapshot copy is deleted or expires.

How volume SnapMirror works with data compression

Because volume SnapMirror operates at the physical block level, when data compression is enabled on the source storage system, the data remains compressed when it is replicated to the destination storage system. This operation can significantly reduce the amount of required network bandwidth during replication.

When using volume SnapMirror with data compression, you must observe the following guidelines:

- For SnapMirror transfer to occur, the source and destination volumes must be contained in 64-bit aggregates and the destination storage system must be running the same or later version of Data ONTAP.

  For example, if the source storage system is running on Data ONTAP 8.0.1, then the destination storage system must be running on Data ONTAP 8.0.1 or later. If the source storage system is running on Data ONTAP 8.1, then the destination storage system must be running on Data ONTAP 8.1 or later.
- You can enable, run, and manage data compression only from the primary storage system.

  However, the FlexVol volume in the secondary storage system inherits all the data compression attributes and storage savings through the volume SnapMirror transfer.
- If you plan to compress existing data in the disk with the `-s` option on a FlexVol volume that has data blocks locked in Snapshot copies and has existing volume SnapMirror relationships, then this operation might result in a large transfer of data blocks.

  The data compression operation rewrites data as new compressed blocks and these blocks are transferred in the next incremental transfer.

For more information about volume SnapMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How qtree SnapMirror works with data compression

Because qtree SnapMirror operates at the logical level, when data compression is enabled on the source storage system, the data is uncompressed in memory before being replicated to the destination system. If data compression is enabled on the secondary storage system, then all transfers are compressed on the secondary storage system.

You can use data compression to compress any existing data on the secondary storage system.

When using qtree SnapMirror with data compression, consider the following guidelines:

- If you want to compress data on your source volume, you must enable data compression on the source volumes.
- If you want to compress data on the destination storage system, you must enable data compression on the destination volume.
- When data compression is enabled on the source storage system, blocks are sent as uncompressed data to the destination storage system.

  Therefore, no network bandwidth savings are achieved by data compression.
- You can enable data compression on the destination storage system even if it is not enabled on the source storage system.

  Therefore, space savings are achieved on the destination storage system.
- If you plan to compress existing data on the disk with the `-a` or `-b` option on a source FlexVol volume that has existing qtree SnapMirror relationships, then this might result in a large transfer of data blocks.

  When data compression is complete, all the newly compressed blocks are written as new blocks and all the newly compressed data blocks are transferred in the next incremental transfer.

For more information about qtree SnapMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How SnapVault works with data compression

SnapVault operates at the logical level and when data compression is enabled on the source storage system, the data is uncompressed in memory on the source storage system before it is backed up.

If inline compression is enabled on the destination storage system, all new writes to the destination storage system are compressed.

When using SnapVault with data compression, you must consider the following guidelines:

- When data compression is enabled on the source system, blocks are sent as uncompressed data to the destination storage system.

  Therefore, no network bandwidth savings are achieved by data compression.

- You can enable data compression on the destination storage system even if it is not enabled on the source storage system.
  After the SnapVault transfer is complete, the post-process compression runs automatically unless the schedule is set to `manual` on the destination system.

For more information about SnapVault, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How tape backup works with data compression

When you backup compressed data to a tape through the SMTape engine, the data format of the source volume is preserved on the tape and all compression savings are retained on the tape.

You can restore the compressed data from a SMTape to a destination volume and the saving will be retained. You must enable inline compression only if you want the new data that is written from the client to be compressed on the restored volume.

When you backup to a tape through NDMP, the data compression savings are not preserved on the tape.

How SnapLock works with data compression

You can use data compression on a SnapLock volume contained within 64-bit aggregates. However, inline compression is not supported for WORM append files in SnapLock volumes that are present on 64-bit storage systems with 4 or more processors.

For more information about SnapLock, see the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How volume-based SnapRestore works with data compression

When you initiate a volume-based SnapRestore operation on a FlexVol volume that contains compressed data, the compression setting is restored to that of the Snapshot copy and the restored data retains the original space savings of the Snapshot copy.

For more information about volume-based SnapRestore, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How single file SnapRestore works with data compression

When you initiate a single file SnapRestore operation, the data is restored from the Snapshot copy to the active file system and the original space savings are restored.

If incompressible data detection is enabled on the volume, then the incompressible data flag is restored to the active file system from the Snapshot copy.

For more information about single file SnapRestore, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How MetroCluster configurations work with data compression

Data compression is supported in both stretch and fabric-attached MetroCluster configurations.

When data compression is enabled for MetroCluster configurations, you must consider the following information:

- Data compression impacts the CPU resources because MetroCluster configurations write to two plexes.
  This results in extra disk write operations and additional system resources are required.
- In takeover mode data compression continues to run normally.

How volume copy works with data compression

When a volume with compressed data is copied to the destination system by using the `vol copy` command, the copy of the data at the destination system inherits all the compression attributes and storage savings of the original data.

For more information about volume copy, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How aggregate copy works with data compression

When an aggregate with compressed data is copied by using the `aggr copy` command, the copy of the data at the destination storage system inherits all the compression attributes and storage savings of the original data.

Data ONTAP 8.1 can read volumes that contain data compressed by Data ONTAP 8.0.1, but Data ONTAP 8.0.1 cannot read volumes that contain data compressed by Data ONTAP 8.1. If the source storage system is running Data ONTAP 8.1 and the destination storage system is running an earlier

version of Data ONTAP, the following operations take place at the destination storage system after the aggregate copy operation is complete and the aggregate is online:

- All FlexVol volumes are copied.
- All FlexVol volumes that contain compressed data go offline.
  Attempts to bring these volumes online fail.
- All FlexVol volumes that do not contain compressed data are online.

For more information about aggregate copy, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How deduplication works with data compression

When both data compression and deduplication are enabled on a FlexVol volume, the data is first compressed and then deduplicated. Depending on the type of data, the combined savings can yield higher savings than running deduplication alone.

How FlexClone volumes work with data compression

If you split a FlexClone volume from the parent volume, then the new volume inherits data compression attributes from the parent volume. The attributes inherited indicate whether deduplication, postprocess compression, and inline compression are enabled. The space savings achieved in the parent volume are inherited by the new volume.

If you create a FlexClone volume when the decompression operation is active on the parent volume, then the decompression operation does not run on the cloned volume.

How FlexClone files work with data compression

You can run data compression on a FlexVol volume that contains FlexClone files. Only fully cloned files can be created on FlexVol volumes that have data compression enabled. Partially cloned files cannot be created on FlexVol volumes that have data compression enabled.

If incompressible data detection is enabled on the volume, then the incompressible flag is inherited by the cloned file.

How HA pairs work with data compression

You can enable data compression in an HA pair. If one of the nodes fails, the other node takes over the operations of the failed node. In the takeover mode, the working node continues to perform the data compression operations.

For more information about HA pairs, see the *Data ONTAP High Availability and MetroCluster Configuration Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How Performance Acceleration Module and Flash cache cards work with data compression

Although data compression, and Performance Acceleration Module and Flash cache cards work in the same storage system, the read performance of the compressed data remains the same with or without PAM and Flash cache cards.

How vFiler units work with data compression

Data compression enables you to reduce the space required to store data in vFiler units by compressing data blocks within a FlexVol volume. Compressed data in volumes is also migrated during the online migration of vFiler units.

When you perform online migration of vFiler units, the bandwidth and time required for the migration are less. The FlexVol volumes on the destination vFiler unit inherit the data compression attributes of the source vFiler unit.

For more information about the compression commands that are not supported during an online migration of vFiler units, see the *Data ONTAP MultiStore Management Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How DataMotion for Volumes works with data compression

Data compression savings on a FlexVol volume are retained when the volume is moved using the DataMotion for Volumes (volume move) operation. For a volume move operation to be successful, the destination volumes must be contained within a 64-bit aggregate and both the source and destination volumes must run the same version of Data ONTAP.

For more information about DataMotion for Volumes, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

Related information

Documentation on the NetApp Support Site: support.netapp.com

How Flash Pools work with data compression

Data compression is supported on Flash Pools, but the compressed blocks are not read or write cached in the solid-state disk (SSD) of the Flash Pools. For a volume that has data compression enabled, only the blocks which not compressed are read or write cached.

How you use space management capabilities

To use the storage provided by FlexVol volumes as effectively as possible, you need to understand the space management capabilities that help you balance overall available storage against required user and application storage needs.

Data ONTAP enables space management using the following capabilities:

- Volume (space) guarantee
  The *volume guarantee*, also called *space guarantee* or just *guarantee*, determines how much space for the volume is preallocated from the volume's associated aggregate when the volume is created.
- Reservations
  Reservations, also called *space reservations*, *file reservations*, or *LUN reservations*, determine whether space for a particular file or LUN is preallocated from the volume.
- Fractional reserve
  Fractional reserve, also called *fractional overwrite reserve* or *LUN overwrite reserve*, enables you to control the size of the overwrite reserve for a FlexVol volume.
- Automatic free space preservation
  Automatic free space preservation can either increase the size of a volume or delete Snapshot copies to prevent a volume from running out of space—all without operator intervention.

These capabilities are used together to enable you to determine, on a volume-by-volume basis, whether to emphasize storage utilization, ease of management, or something in between.

How volume guarantees work with FlexVol volumes

Volume guarantees (sometimes called *space guarantees*) determine how space for a volume is allocated from its containing aggregate--whether the space is preallocated for the entire volume or for only the reserved files or LUNs in the volume, or whether space for user data is not preallocated.

The guarantee is an attribute of the volume.

You set the guarantee when you create a new volume; you can also change the guarantee for an existing volume by using the `vol options` command with the `guarantee` option. You can view the guarantee type and status by using the `vol status` command.

Volume guarantee types can be `volume` (the default type), `file`, or `none`

- A guarantee type of `volume` allocates space in the aggregate for the volume when you create the volume, regardless of whether that space is used for data yet.
  This approach to space management is called *thick provisioning*. The allocated space cannot be provided to or allocated for any other volume in that aggregate.
  When you use thick provisioning, all of the space specified for the volume is allocated from the aggregate at volume creation time. The volume cannot run out of space before the amount of data

it contains (including Snapshot copies) reaches the size of the volume. However, if your volumes are not very full, this comes at the cost of reduced storage utilization.

- A guarantee type of `file` allocates space for the volume in its containing aggregate so that any reserved LUN or file in the volume can be completely rewritten, even if its blocks are being retained on disk by a Snapshot copy.

  However, writes to any file in the volume that is not reserved could run out of space.

  Before configuring your volumes with a guarantee of `file`, you should refer to Technical Report 3965.
- A guarantee of `none` allocates space from the aggregate only as it is needed by the volume.

  This approach to space management is called *thin provisioning*. The amount of space consumed by volumes with this guarantee type grows as data is added instead of being determined by the initial volume size, which might leave space unused if the volume data does not grow to that size. The maximum size of a volume with a guarantee of `none` is not limited by the amount of free space in its aggregate.

  Writes to LUNs or files (including space-reserved files) contained by that volume could fail if the containing aggregate does not have enough available space to accommodate the write. If you configure your volumes with a volume guarantee of `none`, you should refer to Technical Report 3965 for information about how doing so can affect storage availability.

When space in the aggregate is allocated for the guarantee for an existing volume, that space is no longer considered free in the aggregate. Operations that consume free space in the aggregate, such as creation of aggregate Snapshot copies or creation of new volumes in the containing aggregate, can occur only if there is enough available free space in that aggregate; these operations are prevented from using space already allocated to another volume.

When the free space in an aggregate is exhausted, only writes to volumes or files in that aggregate with preallocated space are guaranteed to succeed.

Guarantees are honored only for online volumes. If you take a volume offline, any allocated but unused space for that volume becomes available for other volumes in that aggregate. When you try to bring that volume back online, if there is insufficient available space in the aggregate to fulfill its guarantee, it will remain offline. You must force the volume online, at which point the volume's guarantee will be disabled.

Related information

Technical Report: Thin Provisioning Deployment and Implementation Guide: media.netapp.com/documents/tr-3965.pdf

Technical Report: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment: media.netapp.com/documents/tr3483.pdf

Enabling guarantees for FlexVol volumes

When a volume's guarantee is disabled, the volume functions as though it has a guarantee of `none`. If you have volumes with disabled guarantees, you should address the situation by making more space available to those volumes as soon as possible.

Before you begin

The FlexVol volume must be online.

About this task

Enabled guarantees preallocate space in the aggregate. In volumes with disabled guarantees, operations that require space, such as writes and even deletions, might be disallowed. If a volume's guarantee becomes disabled, you should reenable the guarantee to be able to manually increase the volume size. Volumes with a disabled guarantee and the autogrow feature enabled can still automatically increase in size.

You can view the status of the volume's guarantee first or try to enable the guarantee. If enabling the guarantee fails, Data ONTAP provides the reason (typically insufficient space) and specifies the amount of free space needed in the aggregate. A guarantee type of `none` is never disabled, because there is no space allocated for this guarantee type.

You can set a volume's guarantee to `none`. In this case, there is no concept of enabled or disabled, because the guarantee requires no space.

Steps

1. Optional: View the status of the volume's guarantee by using the `vol status` command.

   If it is not disabled, it will not be listed in the `Options` column.

   If you want to see both the type of guarantee and also whether it is disabled, you can use the `vol options` command.

   In the command output, look for the guarantee type or whether the guarantee is disabled.

   #### Example

   The following `vol status` command displays the status of the guarantee for a volume called vol2. The `Options` column displays the guarantee because it is disabled.

   ```
   sys1> vol status vol2
           Volume State           Status                Options
             vol2 online          raid_dp, flex         guarantee=volume(disabled)
                                  64-bit
   ```

   The following `vol options` command displays the type of guarantee and whether it is disabled for a volume called vol2.

```
sys1> vol options vol2
nosnap=on, nosnapdir=off, minra=off, no_atime_update=off, nvfail=off,
ignore_inconsistent=off, snapmirrored=off, create_ucode=off,
convert_ucode=off, maxdirsize=41861, schedsnapname=ordinal,
fs_size_fixed=off, guarantee=volume(disabled), svo_enable=off,
svo_checksum=off, svo_allow_rman=off, svo_reject_errors=off,
no_i2p=off, fractional_reserve=100, extent=off,
try_first=volume_grow,
read_realloc=off, snapshot_clone_dependency=off,
dlog_hole_reserve=off,
nbu_archival_snap=off
```

The output displays the guarantee type and whether the guarantee is enabled or disabled for the specified volume.

2. Enable or reenable the guarantee.

   Use the `vol options` command with the `guarantee` option set to the existing guarantee type. If you specify a different guarantee than the one currently configured for this volume, Data ONTAP changes the guarantee to the one you specify and enables that guarantee.

   The guarantee is enabled, or you receive an error message that tells you how much space you need to create in the aggregate before the guarantee can be enabled.

3. If there is not enough space in the aggregate to enable the guarantee, you must create more space.

   Example

   The following is an example of an error message displayed when trying to enable a guarantee for a volume named testvol:

   ```
   sys1::> vol options testvol guarantee volume
   vol options: Request to enable guarantee for this volume failed because
   there is not enough space in the aggregate. Create 4.79MB of free space
   in the aggregate.
   ```

4. Try to enable the guarantee again, and view the command output to see whether the guarantee is now enabled.

   If the guarantee is still not enabled, you must try another method of creating more space.

How the guarantee affects FlexVol volume space requirements

The amount of space that a FlexVol volume requires from its aggregate varies depending on the volume's guarantee type. Understanding a volume's space requirement helps you predict how much space becomes available or is required when you change its guarantee or delete the volume.

A volume with a guarantee type of `none` requires space in the aggregate only for data that is already written to it.

A volume with a guarantee type of `volume` requires an amount of space in the aggregate equivalent to the volume's size, regardless of how much data (if any) is actually in the volume.

A volume with a guarantee type of `file` requires enough space in the aggregate to enable writes and overwrites to reserved files or LUNs, even if a block being overwritten is locked by a Snapshot copy or other block-sharing technology.

How file and LUN reservations work

When reservations are enabled for one or more files or LUNs, Data ONTAP reserves enough space in the volume so that writes to those files or LUNs do not fail because of a lack of disk space.

Reservations are an attribute of the file or LUN; they are persistent across storage system reboots, takeovers, and givebacks. Reservations are enabled for new LUNs by default, but you can create a file or LUN with reservations disabled or enabled. After you create a LUN, you change the reservation attribute by using the `lun set reservation` command. You change the reservation attribute for files by using the `file reservation` command.

When a volume contains one or more files or LUNs with reservations enabled, operations that require free space, such as the creation of Snapshot copies, are prevented from using the reserved space. If these operations do not have sufficient unreserved free space, they fail. However, writes to the files or LUNs with reservations enabled continue to succeed.

You can enable reservations for files and LUNs contained by volumes with volume guarantees of any value. However, if the volume has a guarantee of `none`, reservations do not provide protection against out-of-space errors.

Example

If you create a 100-GB space-reserved LUN in a 500-GB volume, that 100 GB of space is immediately allocated, leaving 400 GB remaining in the volume. In contrast, if space reservation is disabled on the LUN, all 500 GB in the volume remain available until writes are made to the LUN.

Considerations for setting fractional reserve

Fractional reserve, also called *LUN overwrite reserve*, enables you to control the size of the overwrite reserve for reserved LUNs and files in a volume. By using this volume attribute correctly you can maximize your storage utilization, but you should understand how it interacts with other technologies.

The fractional reserve setting is expressed as a percentage; the only valid values are 0 and 100 percent. You use the `vol options` command to set fractional reserve.

Setting fractional reserve to 0 increases your storage utilization. However, an application accessing data residing in the volume could experience a data outage if the volume is out of free space, even with the volume guarantee set to `volume`, when any of the following technologies and Data ONTAP features are in use:

- Deduplication
- Compression
- FlexClone files
- FlexClone LUNs
- Virtual environments

If you are using one or more of these technologies with no fractional reserve, and you need to prevent errors due to running out of space, you must use all of the following configuration settings for the volume:

- Volume guarantee of `volume`
- File or LUN reservations `enabled`
- Volume Snapshot copy automatic deletion enabled with a commitment level of `destroy`
- Autogrow feature `enabled`

In addition, you must monitor the free space in the associated aggregate. If the aggregate becomes full enough that the volume is prevented from growing, then data modification operations could fail even with all of the other configuration settings in place.

If you do not want to monitor aggregate free space, you can set the volume's fractional reserve setting to 100. This requires more free space up front, but guarantees that data modification operations will succeed even when the technologies listed above are in use.

The default value and allowed values for the fractional reserve setting depend on the guarantee of the volume:

Volume guarantee	Default fractional reserve	Allowed values
Volume	100	0, 100
None	0	0, 100
File	100	100

For more information about using fractional reserve, see the following Technical Reports:

- *TR-3965: Thin Provisioning Deployment and Implementation Guide*
- *TR-3483: Thin Provisioning in a NetApp SAN or IP SA Enterprise Environment*

Related information

Technical Report: Thin Provisioning Deployment and Implementation Guide: media.netapp.com/documents/tr-3965.pdf

Technical Report: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment: media.netapp.com/documents/tr3483.pdf

How Data ONTAP can automatically provide more space for full FlexVol volumes

Data ONTAP uses two methods for automatically providing more space for a FlexVol volume when that volume is nearly full: allowing the volume size to increase, and deleting Snapshot copies (with any associated storage object). If you enable both of these methods, you can specify which method Data ONTAP should try first.

Data ONTAP can automatically provide more free space for the volume by using one of the following methods:

- Increase the size of the volume when it is nearly full (known as the *autogrow* feature).
  This method is useful if the volume's containing aggregate has enough space to support a larger volume. You can configure Data ONTAP to increase the size in increments and set a maximum size for the volume. The increase is automatically triggered based on the amount of data being written to the volume in relation to the current amount of used space and any thresholds set.
- Delete Snapshot copies when the volume is nearly full.
  For example, you can configure Data ONTAP to automatically delete Snapshot copies that are not linked to Snapshot copies in cloned volumes or LUNs, or you can define which Snapshot copies you want Data ONTAP to delete first—your oldest or newest Snapshot copies. You can also determine when Data ONTAP should begin deleting Snapshot copies—for example, when the volume is nearly full or when the volume's Snapshot reserve is nearly full.
  For more information about deleting Snapshot copies automatically, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

If you enable both of these methods, you can specify which method Data ONTAP tries first when a volume is nearly full. If the first method does not provide sufficient additional space to the volume, Data ONTAP tries the other method next. By default, Data ONTAP tries to increase the size of the volume first.

Selecting the first method to increase space for full FlexVol volumes

If you enable both the volume autosize capability and automatic Snapshot deletion for the same volume, you can also specify which method Data ONTAP tries first when that volume needs additional free space. How you configure the volume depends on whether you would prefer that the volume continue to grow or that Snapshot copies are deleted.

About this task

If the first method that Data ONTAP tries to use for providing additional free space does not result in addressing the free space needs of the volume, then Data ONTAP tries the other method.

In most cases, the default configuration (growing the volume first) is preferable, because when a Snapshot copy is deleted, it cannot be restored. However, in certain situations, you might want to

avoid growing the size of a volume when possible. In this case, you can configure Data ONTAP to delete Snapshot copies before increasing the size of the volume.

Step

1. Select the first method that Data ONTAP should use to provide free space to a volume by using the `vol options` command with the `try-first` option.

   To specify increasing the size of the volume first (the default), use `volume_grow`. To specify deleting Snapshot copies first, use `snap_delete`.

Example

The following command configures Data ONTAP to delete Snapshot copies before increasing the volume size for the volume vol0001:

```
vol options vol0001 try_first snap_delete
```

How a FlexVol volume can automatically change its size

A volume can be configured to grow and shrink automatically in response to space usage requirements. Automatic growing occurs when used space exceeds an autogrow threshold. Automatic shrinking occurs when used space drops below an autoshrink threshold.

The autosizing feature consists of two possible functionalities:

- The autogrow functionality grows a volume's size automatically (`grow` option).

  Automatic growing can provide additional space to a volume when it is about to run out of space, as long as there is space available in the associated aggregate for the volume to grow. When the volume's free space percentage is below the specified threshold, it will continue to grow by the specified increment until either the free space percentage arrives at the threshold or the associated aggregate runs out of space.
- The autoshrink functionality shrinks a volume's size automatically (`grow_shrink` option).

  The autoshrink functionality is only used in combination with autogrow to meet changing space demands and is not available alone. Automatic shrinking helps to more accurately size a volume and prevents a volume from being larger than it needs to be at any given point. The volume shrinks and returns space to the aggregate if the guarantee type is `volume`.

Because the size of the Snapshot reserve is a percentage of the size of the volume, Snapshot spill can start to occur or increase as a result of a volume shrinking.

Configuring a FlexVol volume to automatically change its size

You can configure a volume to grow automatically or grow and shrink automatically (known as *autosizing*) in response to space usage requirements. Automatic growing helps prevent a volume

from running out of space or forcing you to delete files manually. Automatic shrinking prevents a volume from being larger than needed.

Before you begin

The FlexVol volume must be online.

About this task

The autosize capability is off by default, except for data protection mirrors, which have the `grow_shrink` option enabled by default.

You cannot configure a root FlexVol volume for the `grow_shrink` autosize mode.

Step

1. Enter the applicable command to grow the volume size automatically or to grow and shrink the volume size automatically:
   - **`vol autosize <vol_name> grow`**
   - **`vol autosize <vol_name> grow_shrink`**

   You can specify the following parameters related to growing the volume automatically:
   - `-m` sets the maximum size to which a volume can grow.

     The default is 120 percent of the volume size. If you resize the volume manually, this value is reset to 120 percent of the current volume size. A volume does not grow automatically if its current size is greater than or equal to the value of this option.

     If you attempt to set this parameter greater than the platform-dependent maximum volume size, it is silently reset to the maximum volume size.
   - `-i` specifies the amount by which the volume size increases each time the volume grows automatically.

     You can specify the increment amount either as a fixed size (in bytes) or as a percentage. The percentage is converted to a fixed size that is based on the volume size when the command is issued. The default is the lesser value of either 1 GB or 5 percent of the volume size at the time the volume was created. When increasing the size of a volume, Data ONTAP uses the specified increment as a guide; the actual size increase can be larger or smaller.
   - `-grow-threshold-percent` specifies the used space threshold above which growing should start.

     When the volume's used space exceeds this threshold, the volume grows automatically unless it has reached the maximum size specified for automatic growth. The default depends on the size of the volume.

   You can specify the following parameters related to shrinking the volume automatically (in addition to the parameters related to growing):
   - `-minimum size` specifies the smallest size to which the volume can shrink, enabling you to maintain a percentage of free space.

The default minimum size is the initial volume size. Manually resizing the volume or using an invalid minimum size value when you enable the autosizing feature resets this value to the current volume size.

- `-shrink-threshold percent` specifies the volume's used space percent threshold below which shrinking should start.

  When the amount of used space drops below this threshold, the volume shrinks automatically unless it has reached the specified minimum size. For example, if used space is 50 percent and the threshold is 51 percent, automatic shrinking begins. The default is 50 percent.

Example

The command in this example enables the autosizing feature on a volume called thevol. No other parameters were specified, so the defaults are used for growing and shrinking.

```
sys1> vol autosize thevol grow_shrink
vol autosize: Flexible volume 'thevol' autosize settings UPDATED.
sys1> Fri Nov 30 02:42:02 GMT [sys1:wafl.spacemgmnt.policyChg:info]: The space
management policy for volume thevol has changed: autosize state enabled.
```

Requirements for enabling both autoshrink and automatic Snapshot copy deletion

The autoshrink functionality can be used with automatic Snapshot copy deletion if certain configuration requirements are met.

If you want to enable both the autoshrink functionality and automatic Snapshot copy deletion, your configuration must meet the following requirements:

- Data ONTAP must be configured to attempt to increase volume size before trying to delete Snapshot copies (the `try_first` option must be set to `volume_grow`).
- The trigger for automatic Snapshot copy deletion must be volume fullness (the `trigger` parameter must be set to `volume`).

How the autoshrink functionality interacts with Snapshot copy deletion

Because the autoshrink functionality shrinks the size of a FlexVol volume, it can also affect when volume Snapshot copies are automatically deleted.

The autoshrink functionality interacts with automatic volume Snapshot copy deletion in the following ways:

- If both the `grow_shrink` autosize mode and automatic Snapshot copy deletion are enabled, when a volume size shrinks it can trigger an automatic Snapshot copy deletion.

  This is because the Snapshot reserve is based on a percentage of the volume size (5 percent by default), and that percentage is now based on a smaller volume size. This can cause Snapshot copies to spill out of the reserve and be deleted automatically.
- If the `grow_shrink` autosize mode is enabled and you manually delete a Snapshot copy, it might trigger an automatic volume shrinkage.

Considerations for using thin provisioning with FlexVol volumes

Using thin provisioning, you can configure your volumes so that they appear to provide more storage than they actually have available, provided that the storage that is actually being used does not exceed the available storage.

To use thin provisioning with FlexVol volumes, you create the volume with a guarantee of none. With a guarantee of none, the volume size is not limited by the aggregate size. In fact, each volume could, if required, be larger than the containing aggregate. The storage provided by the aggregate is used up only as data is written to the LUN or file.

If the volumes associated with an aggregate show more storage as available than the physical resources available to that aggregate, the aggregate is *overcommitted*. When an aggregate is overcommitted, it is possible for writes to LUNs or files in volumes contained by that aggregate to fail if there is not sufficient free space available to accommodate the write.

If you have overcommitted your aggregate, you must monitor your available space and add storage to the aggregate as needed to avoid write errors due to insufficient space.

Aggregates can provide storage to FlexVol volumes associated with more than one Vserver. When sharing aggregates for thin-provisioned volumes in a multi-tenancy environment, be aware that one tenant's aggregate space availability can be adversely affected by the growth of another tenant's volumes.

For more information about thin provisioning, see the following technical reports:

- *TR 3965: NetApp Thin Provisioning Deployment and Implementation Guide*
- *TR 3483: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment*

How to determine space usage in a volume or aggregate

Enabling a feature in Data ONTAP might consume space that you are not aware of or more space than you expected. Data ONTAP helps you determine how space is being consumed by providing three perspectives from which to view space: the volume, a volume's footprint within the aggregate, and the aggregate.

A volume can run out of space due to space consumption or insufficient space within the volume, aggregate, or a combination of both. By seeing a feature-oriented breakdown of space usage from different perspectives, you can assess which features you might want to adjust or turn off, or take other action (such as increase the size of the aggregate or volume).

You can view space usage details from any of these perspectives:

- The volume's space usage

This perspective provides details about space usage within the volume, including usage by Snapshot copies. The volume's active file system consists of user data, file system metadata, and inodes. Data ONTAP features that you enable might increase the amount of metadata, and in the case of Snapshot copies, can sometimes spill into the user data portion of the active file system.
You see a volume's space usage by using the `vol status -S` command.

- The volume's footprint within the aggregate
  This perspective provides details about the amount of space each volume is using in the containing aggregate, including the volume's metadata.
  You see a volume's footprint with the aggregate by using the `vol status -F` command.
- The aggregate's space usage
  This perspective includes totals of the volume footprints of all of the volumes contained in the aggregate, space reserved for aggregate Snapshot copies, and other aggregate metadata.
  You can see the aggregate's space usage by using the `aggr status -S` command.

Certain features, such as tape backup and deduplication, use space for metadata both from the volume and directly from the aggregate. These features show different space usage between the volume and volume footprint perspectives.

How to determine space usage in an aggregate

You can view space usage by all FlexVol volumes in one or more aggregates with the `aggr status -S` command. This helps you see which volumes are consuming the most space in their containing aggregates so that you can take actions to free more space.

The used space in an aggregate is directly affected by the space used in the FlexVol volumes it contains. Measures that you take to increase space in a volume also affect space in the aggregate.

When the aggregate is offline no values are displayed. Only non-zero values are displayed in the command output. However, you can use the `-v` parameter to display all possible feature rows regardless of whether they are enabled and using any space. A value of `-` indicates that there is no data available to display.

The following rows are included in the `aggr status -S` command output:

- Volume Footprints
  The total of all volume footprints within the aggregate. It includes all of the space that is used or reserved by all data and metadata of all volumes in the containing aggregate. It is also the amount of space that is freed if all volumes in the containing aggregate are destroyed.
- Aggregate Metadata
  The total file system metadata required by the aggregate, such as allocation bitmaps and inode files.
- Snapshot Reserve
  The amount of space reserved for aggregate Snapshot copies, based on volume size. It is considered used space and is not available to volume or aggregate data or metadata. The aggregate's Snapshot reserve is set to 0 percent by default.
- Total Used
  The sum of all space used or reserved in the aggregate by volumes, metadata, or Snapshot copies.

There is never a row for Snapshot spill.

The following example shows the `aggr status -S` command output:

```
sys1> aggr status -S
Aggregate : aggr0

      Feature                                 Used      Used%
      --------------------------------    ----------     ------
      Volume Footprints                      6.05GB        95%
      Aggregate Metadata                      688KB         0%

      Total Used                             6.05GB        95%
```

How you can determine and control a volume's space usage in the aggregate

You can determine which FlexVol volumes are using the most space in the aggregate and specifically which features within the volume. The `vol status -F` command provides information about a volume's footprint, or its space usage within the containing aggregate.

The `vol status -F` command shows details about the space usage of each volume in an aggregate, including offline volumes. This command does not directly correspond to the output of the `df` command, but instead bridges the gap between the output of `vol status -S` and `aggr status -S` commands. All percentages are calculated as a percent of aggregate size.

Only non-zero values are displayed in the command output. However, you can use the `-v` parameter to display all possible feature rows regardless of whether they are enabled and using any space. A value of `-` indicates that there is no data available to display.

The following example shows the `vol status -F` command output for a volume called testvol:

```
sys1> vol status -F
Volume : testvol

      Feature                                 Used      Used%
      --------------------------------    ----------     ------
      Volume Data Footprint                  2.64GB        42%
      Volume Guarantee                       3.32GB        52%
      Flexible Volume Metadata               34.2MB         1%
      Delayed Frees                          54.8MB         1%

      Total                                  6.05GB        95%
```

The following table explains some of the key rows of the output of the `vol status -F` command and what you can do to try to decrease space usage by that feature:

Row/feature name	**Description/contents of row**	**Some ways to decrease**
Volume Data Footprint	The total amount of space used in the containing aggregate by a volume's data in the active file system and the space used by the volume's Snapshot copies. This row does not include reserved space, so if volumes have reserved files, the volume's total used space in the `vol status -S` command output can exceed the value in this row.	• Deleting data from the volume. • Deleting Snapshot copies from the volume.
Volume Guarantee	The amount of space reserved by the volume in the aggregate for future writes. The amount of space reserved depends on the guarantee type of the volume.	Changing the type of guarantee for the volume to `none`. This row will go to 0. If you configure your volumes with a volume guarantee of `none`, you should refer to Technical Report 3965 or 3483 for information about how doing so can affect storage availability.
Flexible Volume Metadata	The total amount of space used in the aggregate by the volume's metadata files.	No direct method to control.
Delayed Frees	Blocks that Data ONTAP used for performance and cannot be immediately freed. When Data ONTAP frees blocks in a FlexVol volume, this space is not always immediately shown as free in the aggregate because operations to free the space in the aggregate are batched for increased performance. Blocks that are declared free in the FlexVol volume but that are not yet free in the aggregate are called “delayed free blocks” until the associated delayed free blocks are processed. For SnapMirror destinations, this row has a value of `0` and is not displayed.	No direct method to control.
Total	The total amount of space that the volume uses in the aggregate. It is the sum of all of the rows.	Any of the methods used to decrease space used by a volume.

Related information

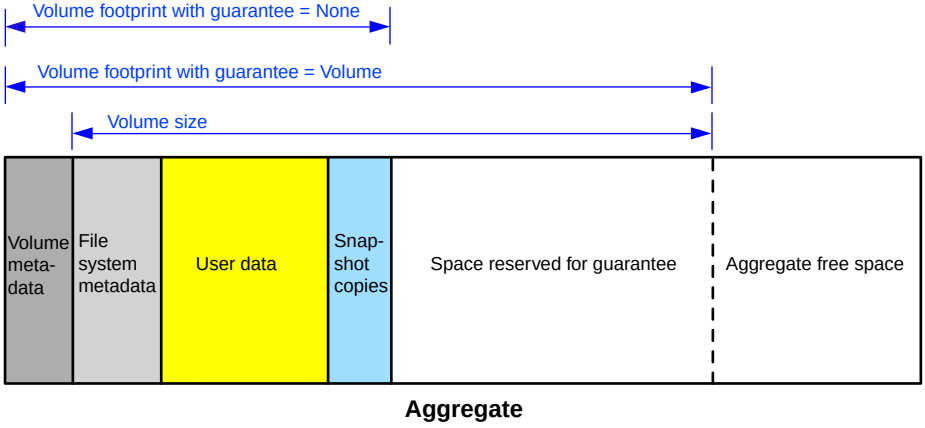
Technical Report: Thin Provisioning Deployment and Implementation Guide: media.netapp.com/documents/tr-3965.pdf

Technical Report: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment: media.netapp.com/documents/tr3483.pdf

What the volume footprint is

A volume footprint is the amount of space a volume is using within the aggregate. Understanding what is included in the volume footprint helps you interpret the output of the space usage commands.

The volume footprint consists of the space used by user data and metadata, including metadata that resides in the aggregate rather than within the volume itself. For this reason it can be larger than the volume size, as shown in the following diagram:



How you can determine and control space usage in a volume

You can view details about space usage in a volume to understand which Data ONTAP features are consuming space and what you can do to decrease that used space.

The `vol status -S` command displays the space used by each of the file system components as well as other features. The `vol status -S` command includes offline volumes. For example, you might want to understand why the `df` command output shows that a large amount of space is still used even though you deleted all of your data in a volume. In this case, output for the volume space usage command might show that it is due to Snapshot copies, inodes, or other metadata that does not shrink.

Only non-zero values are displayed in the command output. However, you can use the `-v` parameter to display all possible feature rows regardless of whether they are enabled and using any space. A value of `-` indicates that there is no data available to display.

The following tables explain some of the common rows in the `vol status -S` command output and what you can do to try to decrease space usage by that feature:

The output for this command consists of the following main categories:

- User data
- Volume metadata
- Snapshot copy information
- Total used space

For information about how to reduce space consumed by other features (such as deduplication), see the respective Data ONTAP guides.

The available space in a volume with a guarantee type of `None` is limited by the available space in the aggregate. Checking the available space in the aggregate might show you that the aggregate is nearly full from aggregate Snapshot copies.

User data

The following output row relates to user data:

Row/feature name	Description	Some ways to decrease space usage
User Data	Everything related to user data, including the data written to the volume, including indirect blocks and directory blocks associated with user inodes, and the space reserved in the volume.	• Deleting user data • Turning off file or LUN reservations Note that turning off file or LUN reservations disables Data ONTAP's ability to guarantee writes to those files or LUNs. This can result in out of space errors being returned. Turning off reservations should be a temporary measure, and reservations should be reenabled as soon as you have provided more free space to the volume.

Volume metadata

The following output rows relate to volume metadata:

Row/feature name	Description	Some ways to decrease space usage
Deduplication / Deduplication Percent	The amount of space used by deduplication metadata files.	Comparison of the space savings you are getting from deduplication with the size of the metadata required. If the metadata requirement is larger than the savings, you can disable deduplication on the volume.
Temporary Deduplication / Temporary Deduplication Percent	The amount of space used by temporary deduplication metadata files.	No direct method to control. The temporary metadata usage decreases after deduplication scanners finish running.
Filesystem Metadata / Filesystem Metadata Percent	Internal tracking for the file system required by Data ONTAP.	No direct method to control.
SnapMirror Metadata / SnapMirror Metadata Percent	The amount of space in use by SnapMirror metadata files. This row relates only to logical replication. During transfers, some additional space is used temporarily.	No direct method to control. You can allow the transfer to finish so the additional space used temporarily is freed.
Tape Backup Metadata / Tape Backup Metadata Percent	The amount of space in use by tape backup metadata files in the volume.	The amount of space consumed by tape backup metadata is cleared when the next baseline (Level 0) backup is successfully run. You can initiate a baseline backup or let it run at the next scheduled time.
Quota Metadata / Quota Metadata Percent	The amount of space used by quota metadata files.	Turning off quotas.
Inodes / Inodes Percent	This row is proportional to the maximum number of files ever created in the volume.	No direct method to control current usage. You can reduce the maximum amount of space that can be used for inode allocations by lowering the maximum public inode setting (maxfiles). However, space that has already been allocated for inodes is never returned to the volume, so if those inodes have already been used, this action has no effect.

Snapshot copy information

The following output rows relate to Snapshot copies:

Row/feature name	Description	Some ways to decrease space usage
Snapshot Reserve	Based on the current volume size. The Snapshot reserve is not available to the active file system and is counted as used space, even if there are no Snapshot copies in the reserve. This row is the same as the total space used for the `.snapshot` row in the `df` command.	You can use the `snap reserve` command with the `percent` option to lower the space allowed for Snapshot copies in the volume.
Snapshot Spill	The amount of space used by Snapshot copies that exceeds the Snapshot reserve size, and spills over into the active file system. This space is not available for writes to the active file system until Snapshot copies are deleted. A non-zero value in this row indicates that your Snapshot reserve has not been sized correctly for your current configuration. Volume clones, SnapMirror, and regularly scheduled Snapshot copies can cause Snapshot copy spill.	• Increasing the size of the Snapshot reserve. • Deleting volume Snapshot copies, either manually or by enabling the Snapshot autodelete capability. • Changing the SnapMirror schedule.

Total used space

The following row relates to total used space in the volume:

Row/ feature name	Description	Some ways to decrease space usage
Total	The total amount of used space in the volume, including the amount of space allotted for the entire Snapshot reserve and space for the active file system. Snapshot space is treated as used space, so this row can be higher than the `df` command's output. In the `df` command, this row is equivalent to the volume's used space in the `used` column plus the Snapshot total (in the `total` column) for the Snapshot used space (`.snapshot`) row. When Snapshot spill exists, the `vol status -S` command only accounts for the used space once. However, the `df` command shows that space as used for both the active file system and for the `.snapshot` row.	Any of the methods for individual output rows.

Example output

The following is an example of command output for a FlexVol volume called testvol.

```
sys1> vol status -S
Volume : testvol

      Feature                                 Used      Used%
      --------------------------------    ----------     ------
      User Data                             2.40GB        40%
      Filesystem Metadata                   1.07MB         0%
      Inodes                                2.63MB         0%
      Snapshot Reserve                       308MB         5%

      Total                                 2.71GB        45%
```

How Snapshot copies and Snapshot reserve use space in a volume

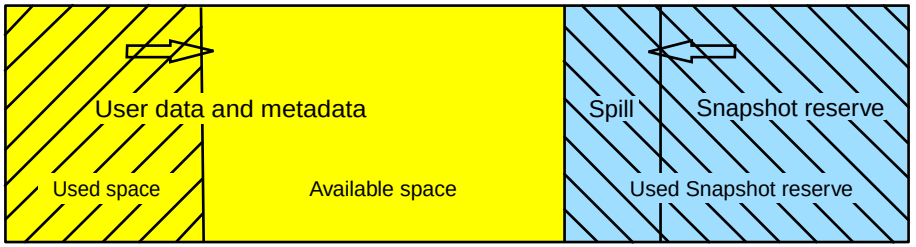
Understanding the Snapshot reserve area of a FlexVol volume and what Snapshot spill is can help you correctly size the Snapshot reserve and decide whether to enable the Snapshot autodelete capability.

When Snapshot copies use more space than the Snapshot reserve, they spill over and use space in the active file system. The *Snapshot reserve* area of a volume is the space reserved exclusively for Snapshot copies. It is not available to the user data or metadata area of the volume. The size of the Snapshot reserve is a specified percentage of the current volume size, and does not depend on the number of Snapshot copies or how much space they consume.

If all of the space allotted for the Snapshot reserve is used but the active file system (user data and metadata) is not full, Snapshot copies can use more space than the Snapshot reserve and spill into the active file system. This extra space is called *Snapshot spill*.

The following illustration shows a FlexVol volume with no Snapshot spill occurring. The two blocks on the left show the volume's used and available space for user data and metadata. The two blocks on the right show the used and unused portions of the Snapshot reserve. When you modify the size of the Snapshot reserve, it is the blocks on the right that change.

The following illustration shows a FlexVol volume with Snapshot spill occurring. The Snapshot reserve area is full and Snapshot copies spilling over into a Spill area that is part of the user data and metadata area's available space. The size of the Snapshot reserve remains the same.



For more information about Snapshot copies, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*

When to use the df command and the space usage commands

You use the `df` command when you want concise information about used and available space in volumes or aggregates. If you want a detailed breakdown of space usage by feature in a volume, aggregate, or a volume's footprint within an aggregate, you use the space usage commands.

The `df` command is useful if you want a quick view of how much space each volume has available or used.

You use the `df` command (or `volume show` and `aggregate show` commands) to see total space, available space, and used space. If you need more information about how or why space is being used in your volume or aggregate, you use the `show-space` and `show-footprint` commands (the space usage commands) for the volume or aggregate.

The space usage commands, on the other hand, provide much more detail about used space and what Data ONTAP capability is causing the space to be used. For example, they could be used to help you understand why the `df` command output shows used space even though there is no data in a volume.

Used space is dynamic, even for a system that is not being accessed by clients. For this reason, you should not try to compare the output of two different space commands, or even the same command invoked twice, too closely.

Methods to create space in a FlexVol volume

There are multiple ways to create space in a FlexVol volume. Understanding what these methods are and their respective benefits and drawbacks helps you decide which method is best for your requirements.

Some common ways to create space in a volume are as follows:

- Increase the size of the volume.

  You can do this manually, or automatically by enabling the autogrow functionality.
- Reduce the size of the Snapshot reserve if the `df` command shows that the Snapshot reserve is not 100 percent full.

  This makes space available to the active file system.
- Make more space in the aggregate.

  This results directly or indirectly in more space being made for the volume. For example, more space in the aggregate can allow a volume to increase in size automatically with the autogrow capability.
- Enable storage efficiency technologies, such as deduplication and compression.
- Delete volume Snapshot copies if the Snapshot reserve is 100 percent full and Snapshot copies are spilling into the active file system.

  You can delete Snapshot copies manually, or automatically by enabling the Snapshot autodelete capability for the volume.
- Delete FlexClone LUNs and LUN clones.
- (Temporarily) change the fractional reserve to 0 percent if your volume contains reserved files or LUNs and the fractional reserve is 100 percent.

  You should only use this as a temporary measure to create space. When the fractional reserve is set to 0 percent, overwrites might fail, and in certain deployments write failures might not be acceptable.
- Delete files.

  If the volume is 100 percent full, it might not be possible to delete a file if it participates in any block sharing, such as volume Snapshot copies or deduplication, and you cannot recover the space. In addition, modifying a directory to delete a file might require additional space, so deleting the file can actually consume space. Under these conditions, you can do one of the following:

  - You can use the `rm` command, available at the advanced privilege level, to delete files even in full volumes with Snapshot copies.

- You can use any of the other methods listed to create more space in the volume and aggregate so that there is enough space available for file deletions.

Methods to create space in an aggregate

If an aggregate runs out of free space, various problems can result that range from loss of data to disabling a volume's guarantee. There are multiple ways to make more space in an aggregate.

All of the methods have various consequences. Prior to taking any action, you should read the relevant section in the documentation.

The following are some common ways to make space in an aggregate, in order of least to most consequences:

- Add disks to the aggregate.
- Move some volumes to another aggregate with available space.
- Shrink the size of volumes whose guarantee type is `volume` in the aggregate.

  You can do this manually or with the `autoshrink` option of the autosize capability.
- Change volume guarantee types to `none` on volumes that are using large amounts of space (large volume-guaranteed volumes or file-guaranteed volumes with large reserved files) so that the volumes take up less space in the aggregate.

  A volume with a guarantee type of `none` has a smaller footprint in the aggregate than volumes with other guarantee types. The `Volume Guarantee` row of the `vol status -F` command output shows whether a volume is reserving a large amount of space in the aggregate due to its guarantee.

  If you configure your volumes with a volume guarantee of `none`, you should refer to Technical Report 3965 for information about how doing so can affect storage availability.
- Delete unneeded volume Snapshot copies if the volume's guarantee type is `none`.
- Delete unneeded volumes.
- Enable space-saving features, such as deduplication or compression.
- (Temporarily) disable features that are using a large amount of metadata (visible with the `vol status -F` command).

Related information

Technical Report: Thin Provisioning Deployment and Implementation Guide: media.netapp.com/documents/tr-3965.pdf

Technical Report: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment: media.netapp.com/documents/tr3483.pdf

About qtrees

Qtrees enable you to partition your volumes into smaller segments that you can manage individually. You can set a qtree's size or security style, back it up, and restore it.

When to use qtrees

Qtrees enable you to partition your data without incurring the overhead associated with a volume. You might create qtrees to organize your data, or to manage one or more of the following factors: quotas, backup strategy, security style, and CIFS oplocks setting.

The following list describes examples of qtree usage strategies:

- Quotas
  You can limit the size of the data used by a particular project, by placing all of that project's files into a qtree and applying a tree quota to the qtree.
- Backups
  You can use qtrees to keep your backups more modular, to add flexibility to backup schedules, or to limit the size of each backup to one tape.
- Security style
  If you have a project that needs to use NTFS-style security, because the members of the project use Windows files and applications, you can group the data for that project in a qtree and set its security style to NTFS, without requiring that other projects also use the same security style.
- CIFS oplocks settings
  If you have a project using a database that requires CIFS oplocks to be off, you can set CIFS oplocks to `off` for that project's qtree, while allowing other projects to retain CIFS oplocks.

How qtrees compare with FlexVol volumes

In general, qtrees are similar to FlexVol volumes. However, the two technologies have some key differences. Understanding these differences helps you choose between them when you design your storage architecture.

The following table compares qtrees and FlexVol volumes.

Functionality	Qtree	FlexVol volume
Enables organizing user data	Yes	Yes
Enables grouping users with similar needs	Yes	Yes
Accepts a security style	Yes	Yes

Functionality	Qtree	FlexVol volume
Accepts oplocks configuration	Yes	Yes
Can be backed up and restored as a unit using SnapMirror	Yes	Yes
Can be backed up and restored as a unit using SnapVault	Yes	No
Can be resized	Yes (using quota limits)	Yes
Supports Snapshot copies	No (qtree data can be extracted from volume Snapshot copies)	Yes
Supports quotas	Yes	Yes
Can be cloned	No (except as part of a FlexVol volume)	Yes

Related references

Storage limits on page 342

Qtree name restrictions

Qtree names can be no more than 64 characters in length. In addition, using some special characters in qtree names, such as commas and spaces, can cause problems with other Data ONTAP capabilities, and should be avoided.

The following characters should be avoided in qtree names:

- Space
  Spaces in qtree names can prevent SnapMirror updates from working correctly.
- Comma
  Commas in qtree names can prevent quotas from working correctly for that qtree, unless the name is enclosed in double quotation marks.

Managing qtrees

You can create, delete, and rename qtrees. In addition, you can display their status and access statistics. You can also convert directories at the root of a volume into qtrees. You do many of these operations using your UNIX or Windows client.

About this task

Note: Many qtree commands cannot be performed while a volume move operation is in progress. If you are prevented from completing a qtree command for this reason, wait until the volume move is complete and then retry the command.

Creating a qtree

You create qtrees using the `qtree create` command. You can also specify a UNIX-style permission for the new qtree.

Steps

1. Enter the following command:

   `qtree create` ***`path`*** **`[-m`** ***`mode`*** **`]`**

   `mode` is a UNIX-style octal number that specifies the permissions for the new qtree. If you do not specify a mode, the qtree is created with the permissions specified by the `wafl.default_qtree_mode` option.

   For more information about the format of the mode number, see your UNIX documentation.

   Note: If you are using this qtree in an NTFS-only environment, you can set the appropriate ACLs after creation using Windows tools.

   `path` is the path name of the qtree, with the following notes:

   - If you want to create the qtree in a volume other than the root volume, include the volume in the name.
   - If the path name does not begin with a slash (/), the qtree is created in the root volume.
   - Qtree names can be up to 64 characters long. The entire path can be up to 1,024 characters long.

2. If you want to change the default security style or the default CIFS oplocks setting of the new qtree, you can change it now by using the `qtree security` or `qtree oplocks` commands.

Examples

The following command creates the news qtree in the users volume, giving the owner and the owner's group permission to read, write and execute the qtree:

```
qtree create /vol/users/news -m 770
```

The following command creates the news qtree in the root volume:

```
qtree create news
```

Related concepts

Qtree name restrictions on page 287
About qtrees on page 286

Displaying qtree status

To find the security style, oplocks attribute, and SnapMirror status for all volumes and qtrees on the storage system or for a specified volume, you use the `qtree status` command.

Step

1. Enter the following command:

   ```
   qtree status [-i] [-v] [vol_name]
   ```

   The `-i` option includes the qtree ID number in the display.

   The `-v` option includes the owning vFiler unit, if the MultiStore license is enabled.

Displaying qtree access statistics

You display statistics on user accesses to files in qtrees on your system using the `qtree stats` command. This can help you determine which qtrees are incurring the most traffic. Determining traffic patterns helps with qtree-based load balancing.

About this task

The `qtree stats` command displays the number of NFS and CIFS accesses to the designated qtrees since the counters were last reset. The qtree stats counters are reset when one of the following actions occurs:

- The system is booted.
- The volume containing the qtree is brought online.
- The counters are explicitly reset using the `qtree stats -z` command.

Step

1. Enter the following command:

   qtree stats [-z] [*vol_name*]

   The `-z` option clears the counter for the designated qtree, or clears all counters if no qtree is specified.

   `vol_name` optionally specifies a volume. Statistics for all qtrees in that volume are displayed. If no volume is specified, statistics for all qtrees on the storage system are displayed.

Example output

```
system> qtree stats vol1
Volume        Tree         NFS ops       CIFS ops
--------      --------     -------       --------
vol1          proj1           1232             23
vol1          proj2             55            312
```

Converting a directory to a qtree

If you have a directory at the root of a FlexVol volume that you want to convert to a qtree, you must migrate the data contained in the directory to a new qtree with the same name, using your client application.

About this task

The steps you take to convert a directory to a qtree depend on what client you use. The following process outlines the general tasks you need to complete:

Steps

1. Rename the directory to be made into a qtree.
2. Create a new qtree with the original directory name.
3. Use the client application to move the contents of the directory into the new qtree.
4. Delete the now-empty directory.

   Note: You cannot delete a directory if it is associated with an existing CIFS share.

Converting a directory to a qtree using a Windows client

To convert a directory to a qtree using a Windows client, you rename the directory, create a qtree on the storage system, and move the directory's contents to the qtree.

About this task

You must use Windows Explorer for this procedure. You cannot use the Windows command-line interface or the DOS prompt environment.

Steps

1. Open Windows Explorer.
2. Click the folder representation of the directory you want to change.

   Note: The directory must reside at the root of its containing volume.
3. From the File menu, select Rename to give this directory a different name.
4. On the storage system, use the `qtree create` command to create a new qtree with the original name of the directory.
5. In Windows Explorer, open the renamed directory folder and select the files inside it.
6. Drag these files into the folder representation of the new qtree.

   Note: The more subfolders contained in the folder that you are moving, the longer the move operation takes.
7. From the File menu, select Delete to delete the renamed, now-empty directory folder.

Converting a directory to a qtree using a UNIX client

To convert a directory to a qtree in UNIX, you rename the directory, create a qtree on the storage system, and move the directory's contents to the qtree.

Steps

1. Open a UNIX client window.
2. Use the `mv` command to rename the directory.

   ### Example

   ```
   client: mv /n/joel/vol1/dir1 /n/joel/vol1/olddir
   ```
3. From the storage system, use the `qtree create` command to create a qtree with the original name.

Example

```
system1: qtree create /n/joel/vol1/dir1
```

4. From the client, use the `mv` command to move the contents of the old directory into the qtree.

   Note: The more subdirectories contained in a directory that you are moving, the longer the move operation will take.

   Example

   ```
   client: mv /n/joel/vol1/olddir/* /n/joel/vol1/dir1
   ```

5. Use the `rmdir` command to delete the old, now-empty directory.

   Example

   ```
   client: rmdir /n/joel/vol1/olddir
   ```

After you finish

Depending on how your UNIX client implements the `mv` command, file ownership and permissions might not be preserved. If this occurs, update file owners and permissions to their previous values.

Deleting a qtree

You can delete a qtree using Windows Explorer or a UNIX client, if the qtree permissions allow.

Before you begin

The following conditions must be true:

- The volume that contains the qtree you want to delete must be mounted (for NFS) or mapped (for CIFS).
- The qtree you are deleting must not be directly mounted and must not have a CIFS share directly associated with it.
- The qtree permissions must allow you to modify the qtree.

Steps

1. Find the qtree you want to delete.

   Note: The qtree appears as a normal directory at the root of the volume.

2. Delete the qtree using the method appropriate for your client.

Example

The following command on a UNIX host deletes a qtree that contains files and subdirectories:

```
rm -Rf directory
```

Note: On a Windows host, you must use Windows Explorer to delete a qtree.

Related concepts

How deleting a qtree affects tree quotas on page 311
About qtrees on page 286

Renaming a qtree

You can rename a qtree using Windows Explorer or a UNIX client, if the qtree permissions allow.

Before you begin

The following conditions must be true:

- The volume that contains the qtree you want to rename must be mounted (for NFS) or mapped (for CIFS).
- The qtree you are renaming must not be directly mounted and must not have a CIFS share directly associated with it.
- The qtree permissions must allow you to modify the qtree.

Steps

1. Find the qtree you want to rename.

   Note: The qtree appears as a normal directory at the root of the volume.

2. Rename the qtree using the method appropriate for your client.

   #### Example

   The following command on a UNIX host renames a qtree:

   ```
   mv old_name new_name
   ```

   Note: On a Windows host, you must use Windows Explorer to rename a qtree.

After you finish

If you have quotas on the renamed qtree, update the quotas file to use the new qtree name.

Related concepts

How renaming a qtree affects quotas on page 311

About qtrees on page 286

About quotas

Quotas provide a way to restrict or track the disk space and number of files used by a user, group, or qtree. You specify quotas using the `/etc/quotas` file. Quotas are applied to a specific volume or qtree.

Why you use quotas

You can use quotas to limit resource usage, to provide notification when resource usage reaches specific levels, or to track resource usage.

You specify a quota for the following reasons:

- To limit the amount of disk space or the number of files that can be used by a user or group, or that can be contained by a qtree
- To track the amount of disk space or the number of files used by a user, group, or qtree, without imposing a limit
- To warn users when their disk usage or file usage is high

Overview of the quota process

Quotas can be soft or hard. Soft quotas cause Data ONTAP to send a notification when specified thresholds are exceeded, and hard quotas prevent a write operation from succeeding when specified thresholds are exceeded.

When Data ONTAP receives a request to write to a volume, it checks to see whether quotas are activated for that volume. If so, Data ONTAP determines whether any quota for that volume (and, if the write is to a qtree, for that qtree) would be exceeded by performing the write operation. If any hard quota is exceeded, the write operation fails, and a quota notification is sent. If any soft quota is exceeded, the write operation succeeds, and a quota notification is sent.

Understanding quota notifications

Quota notifications are messages sent to the console and the `/etc/messages` file. You can also configure SNMP traps to be triggered when a quota is exceeded.

Notifications are sent in response to the following events:

- A hard quota is reached; in other words, an attempt is made to exceed it
- A soft quota is exceeded
- A soft quota is no longer exceeded

Thresholds are slightly different from other soft quotas. Thresholds trigger notifications only when they are exceeded, not when they are no longer exceeded.

Hard-quota notifications are configurable using the `quota logmsg` command. You can turn them off completely, and you can change their frequency, for example, to prevent sending of redundant messages.

Soft-quota notifications are not configurable because they are unlikely to generate redundant messages and their sole purpose is notification.

SNMP traps can be used to arrange alternative methods of notification, such as email. You can find details on SNMP traps in the `/etc/mib/netapp.mib` file.

Note: Notifications contain qtree ID numbers rather than qtree names. You can correlate qtree names to ID numbers by using the `qtree status -i` command.

Quota targets and types

Quotas have a type: they can be either user, group, or tree. Quota targets specify the user, group, or qtree for which the quota limits are applied.

The following table lists the kinds of quota targets, what types of quotas each quota target is associated with, and how each quota target is represented.

Quota target	Quota type	How target is represented	Notes
user	user quota	UNIX user name UNIX UID A file or directory whose UID matches the user Windows user name in pre-Windows 2000 format Windows SID A file or directory with an ACL owned by the user's SID	User quotas can be applied for a specific volume or qtree.
group	group quota	UNIX group name UNIX GID A file or directory whose GID matches the group	Group quotas can be applied for a specific volume or qtree. **Note:** Data ONTAP does not apply group quotas based on Windows IDs.

Quota target	Quota type	How target is represented	Notes
qtree	tree quota	path name to the qtree For example, `vol/vol1/qtree2`	Tree quotas are applied to a particular volume and do not affect qtrees in other volumes.
*	user quota group quota tree quota	The asterisk character (*)	A quota target of * denotes a *default quota*. For default quotas, the quota type is determined by the value of the type field.

Special kinds of quotas

You use default, explicit, derived and tracking quotas to manage disk usage in the most efficient manner.

How default quotas work

You can use default quotas to apply a quota to all instances of a given quota type. For example, a default user quota affects all users on the system for the specified volume or qtree. . In addition, default quotas enable you to modify your quotas easily.

You can use default quotas to automatically apply a limit to a large set of quota targets without having to create separate quotas for each target. For example, if you want to limit most users to 10 GB of disk space, you can specify a default user quota of 10 GB of disk space instead of creating a quota for each user. If you have specific users for whom you want to apply a different limit, you can create explicit quotas for those users. (Explicit quotas—quotas with a specific target or list of targets—override default quotas.)

In addition, default quotas enable you to use resizing rather than reinitialization when you want quota changes to take effect. For example, if you add an explicit user quota to a volume that already has a default user quota, you can activate the new quota by resizing.

Default quotas can be applied to all three types of quota target (users, groups, and qtrees).

Default quotas do not necessarily have specified limits; a default quota can be a tracking quota.

Default user quota example

The following quotas file uses a default user quota to apply a 50-MB limit on each user for vol1:

```
#Quota target type           disk  files  thold  sdisk  sfile
#-----------  ----           ----  -----  -----  -----  -----
*             user@/vol/vol1  50M
```

If any user on the system enters a command that would cause that user's data to take up more than 50 MB in vol1 (for example, writing to a file from an editor), the command fails.

How you use explicit quotas

You can use explicit quotas to specify a quota for a specific quota target, or to override a default quota for a specific target.

An explicit quota specifies a limit for a particular user, group, or qtree. An explicit quota replaces any default quota that is in place for the same target.

When you add an explicit user quota for a user that has a derived user quota, you must use the same user mapping setting as the default user quota. Otherwise, when you resize quotas, the explicit user quota is rejected because it is considered a new quota.

Explicit quotas only affect default quotas at the same level (volume or qtree). For example, an explicit user quota for a qtree does not affect the default user quota for the volume that contains that qtree. However, the explicit user quota for the qtree overrides (replaces the limits defined by) the default user quota for that qtree.

Examples of explicit quotas

The following quotas file contains a default user quota that limits all users in vol1 to 50 MB of space. However, one user, jsmith, is allowed 80 MB of space, because of the explicit quota (shown in bold):

```
#Quota target type            disk  files  thold  sdisk  sfile
#-----------  ----            ----  -----  -----  -----  -----
*             user@/vol/vol1  50M
jsmith        user@/vol/vol1  80M
```

The following quotas entry restricts the specified user, represented by four IDs, to 500MB of disk space and 10,240 files in the vol1 volume:

```
jsmith,corp\jsmith,engineering\”john smith”,
S-1-5-32-544   user@/vol/vol1         500M      10K
```

The following quotas entry restricts the eng1 group to 150 MB of disk space and an unlimited number of files in the /vol/vol2/proj1 qtree:

```
eng1        group@/vol/vol2/proj1  150M
```

The following quotas entry restricts the proj1 qtree in the vol2 volume to 750 MB of disk space and 76,800 files:

```
/vol/vol2/proj1   tree                750M      75K
```

How derived quotas work

A quota enforced as a result of a default quota, rather than an explicit quota (a quota with a specific target), is referred to as a *derived quota*.

The number and location of the derived quotas depends on the quota type:

- A default tree quota on a volume creates derived tree quotas for every qtree on the volume.
- A default user or group quota creates a derived user or group quota for every user or group that owns a file at the same level (volume or qtree).
- A default user or group quota on a volume creates a default user or group quota on every qtree that also has a tree quota.

The settings—including limits and user mapping—of derived quotas are the same as the settings of the corresponding default quotas. For example, a default tree quota with a 20-GB disk limit on a volume creates derived tree quotas with 20-GB disk limits on the qtrees in the volume. If a default quota is a tracking quota (with no limits), the derived quotas are also tracking quotas.

To see derived quotas, you can generate a quota report. In the report, a derived user or group quota is indicated by a Quota Specifier that is either blank or an asterisk (*). A derived tree quota, however, has a Quota Specifier; to identify a derived tree quota, you must look for a default tree quota on the volume with the same limits.

Explicit quotas interact with derived quotas in the following ways:

- Derived quotas are not created if an explicit quota already exists for the same target.
- If a derived quota exists when you create an explicit quota for a target, you can activate the explicit quota by resizing rather than having to perform a full quota initialization.

How you use tracking quotas

Tracking quotas generate reports of disk and file usage and do not limit resource usage. When tracking quotas are used, modifying quota values is less disruptive, because you can resize quotas rather than turning them off and back on.

To create a tracking quota, you specify a dash ("-") for the disk and files values. This tells Data ONTAP to monitor disk and files usage for that target at that level (volume or qtree), without imposing any limits.

You can also specify a *default tracking quota*, which applies to all instances of the target. Default tracking quotas enable you to track usage for all instances of a quota type (for example, all qtrees or all users). In addition, they enable you use resizing rather than reinitialization when you want quota changes to take effect.

Examples

The following quotas file shows tracking quotas in place for a specific user, group, and qtree:

```
#Quota target   type           disk files thold sdisk sfile
#-----------    ----           ---- ----- ----- ----- -----
kjones          user@/vol/vol1    -  -
eng1            group@/vol/vol1   -  -
proj1           tree@/vol/vol1    -  -
```

The following quotas file contains the three possible default tracking quotas (users, groups, and qtrees):

```
#Quota target   type           disk  files thold sdisk sfile
#-----------    ----           ----  ----- ----- ----- -----
*               user@/vol/vol1    -   -
*               group@/vol/vol1   -   -
*               tree@/vol/vol1    -   -
```

How quotas are applied

Understanding how quotas are applied enables you to configure quotas and set the expected limits.

Whenever an attempt is made to create a file or write data to a file in a volume that has quotas enabled, the quota limits are checked before the operation proceeds. If the operation exceeds either the disk limit or the files limit, the operation is prevented.

Quota limits are checked in the following order:

1. The tree quota for that qtree (This check is not relevant if the file is being created or written to qtree0.)
2. The user quota for the user that owns the file on the volume
3. The group quota for the group that owns the file on the volume
4. The user quota for the user that owns the file on the qtree (This check is not relevant if the file is being created or written to qtree0.)
5. The group quota for the group that owns the file on the qtree (This check is not relevant if the file is being created or written to qtree0.)

The quota with the smallest limit might not be the one that is exceeded first. For example, if a user quota for volume vol1 is 100 GB, and the user quota for qtree q2 contained in volume vol1 is 20 GB, the volume limit could be reached first if that user has already written more than 80 GB of data in volume vol1 (but outside of qtree q2).

Related concepts

How quotas work with users and groups on page 301
How you use explicit quotas on page 298
How default quotas work on page 297
Quota targets and types on page 296

How quotas work with users and groups

When you specify a user or group as the target of a quota, the limits imposed by that quota are applied to that user or group. However, some special groups and users are handled differently. There are different ways to specify IDs for users, depending on your environment.

Related concepts

How default quotas work on page 297
How you use tracking quotas on page 299

How you specify UNIX users for quotas

You can specify a UNIX user for a quota using one of three formats: the user name, the UID, or a file or directory owned by the user.

To specify a UNIX user for a quota, you can use one of the following formats:

- The user name as defined in the `/etc/passwd` file or the NIS password map, for example, jsmith.

  Note: You cannot use a UNIX user name to specify a quota if that name includes a backslash (\) or an @ sign. This is because Data ONTAP treats names containing these characters as Windows names.

- The UID, such as 20.
- The path of a file or directory owned by that user, so that the file's UID matches the user.

  Note: If you specify a file or directory name, you must select a file or directory that will last as long as the user account remains on the system.

  Specifying a file or directory name for the UID does not cause Data ONTAP to apply a quota to that file or directory.

How you specify Windows users for quotas

You can specify a Windows user for a quota using one of three formats: the Windows name in pre-Windows 2000 format, the SID, or a file or directory owned by the SID of the user.

To specify a Windows user for a quota, you can use one of the following formats:

- The Windows name in pre-Windows 2000 format.
- The security ID (SID), as displayed by Windows in text form, such as S-1-5-32-544.
- The name of a file or directory that has an ACL owned by that user's SID.

  Note: If you specify a file or directory name, you must select a file or directory that will last as long as the user account remains on the system.

  For Data ONTAP to obtain the SID from the ACL, the ACL must be valid.

  If the file or directory exists in a UNIX-style qtree, or if the storage system uses UNIX mode for user authentication, Data ONTAP applies the user quota to the user whose *UID*, not SID, matches that of the file or directory.

  Specifying a file or directory name to identify a user for a quota does not cause Data ONTAP to apply a quota to that file or directory.

How you specify a user name in pre-Windows 2000 format

The pre-Windows 2000 format, for example `engineering\john_smith`, is used by the quotas file for specifying Windows users.

Keep in mind the following rules when creating pre-Windows 2000 format user names:

- The user name must not exceed 20 characters.
- The NetBIOS form of the domain name must be used.

How you specify a Windows domain using the QUOTA_TARGET_DOMAIN directive

Using the QUOTA_TARGET_DOMAIN directive in the quotas file enables you to specify the domain name only once for a group of Windows users.

The QUOTA_TARGET_DOMAIN directive takes an optional argument. This string, followed by a backslash (\), is prefixed to the name specified in the quota entry. Data ONTAP stops adding the domain name when it reaches the end of the quotas file or another QUOTA_TARGET_DOMAIN directive.

Example

The following example illustrates the use of the QUOTA_TARGET_DOMAIN directive:

```
QUOTA_TARGET_DOMAIN corp
roberts    user@/vol/vol2         900M      30K
smith      user@/vol/vol2         900M      30K
QUOTA_TARGET_DOMAIN engineering
daly       user@/vol/vol2         900M      30K
thomas     user@/vol/vol2         900M      30K
QUOTA_TARGET_DOMAIN
stevens    user@/vol/vol2         900M      30K
```

The string corp\ is added as a prefix to the user names of the first two entries. The string engineering\ is added as a prefix to the user names of the third and fourth entries. The last

entry is unaffected by the QUOTA_TARGET_DOMAIN entry because the entry contains no argument.

The following entries produce the same effects:

```
corp\roberts       user@/vol/vol2         900M      30K
corp\smith         user@/vol/vol2         900M      30K
engineering\daly   user@/vol/vol2         900M      30K
engineering\thomas user@/vol/vol2         900M      30K
stevens            user@/vol/vol2         900M      30K
```

How default user and group quotas create derived quotas

When you create default user or group quotas, corresponding derived user or group quotas are automatically created for every user or group that owns files at the same level.

Derived user and group quotas are created in the following ways:

- A default user quota on a volume creates derived user quotas for every user that owns a file anywhere on the volume.
- A default user quota on a qtree creates derived user quotas for every user that owns a file in the qtree.
- A default group quota on a volume creates derived group quotas for every group that owns a file anywhere on the volume.
- A default group quota on a qtree creates derived group quotas for every group that owns a file in the qtree.

If a user or group does not own files at the level of a default user or group quota, derived quotas are not created for the user or group. For example, if a default user quota is created for qtree proj1 and the user jsmith owns files on a different qtree, no derived user quota is created for jsmith.

The derived quotas have the same settings as the default quotas, including limits and user mapping. For example, if a default user quota has a 50-MB disk limit and has user mapping turned on, any resulting derived quotas also have a 50-MB disk limit and user mapping turned on.

However, no limits exist in derived quotas for three special users and groups. If the following users and groups own files at the level of a default user or group quota, a derived quota is created with the same user-mapping setting as the default user or group quota, but it is only a tracking quota (with no limits):

- UNIX root user (UID 0)
- UNIX root group (GID 0)
- Windows BUILTIN\Administrators group

  Since quotas for Windows groups are tracked as user quotas, a derived quota for this group is a user quota that is derived from a default user quota, not a default group quota.

Example of derived user quotas

If you have volume where three users—root, jsmith, and bob—own files, and you create a default user quota on the volume, Data ONTAP automatically creates three derived user quotas. Therefore, after you reinitialize quotas on the volume, four new quotas appear in the quota report:

```
filer1> quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
user       *      vol1        -           0     51200       0       - *
user       root   vol1        -           5         -       1       -
user       jsmith vol1        -          30     51200      10       - *
user       bob    vol1        -          40     51200      15       - *
```

The first new line is the default user quota that you created, which is identifiable by the asterisk (*) as the ID. The other new lines are the derived user quotas. The derived quotas for jsmith and bob have the same 50-MB disk limit as the default quota. The derived quota for the root user is a tracking quota without limits.

How quotas are applied to the root user

The root user (UID=0) on UNIX clients is subject to tree quotas, but not user quotas or group quotas. This allows the root user to take actions on behalf of other users that would otherwise be prevented by a quota.

When root carries out a file or directory ownership change or other operation (such as the UNIX `chown` command) on behalf of a user with less privileges, Data ONTAP checks the quotas based on the new owner but does not report errors or stop the operation, even if the hard quota restrictions of the new owner are exceeded. This can be useful when an administrative action, such as recovering lost data, results in temporarily exceeding quotas.

Note: After the ownership transfer is carried out, however, a client system will report a disk space error if the user attempts to allocate more disk space while the quota is still exceeded.

How quotas work with special Windows groups

Quotas are applied to the Everyone group and the BUILTIN\Administrators group differently than to other Windows groups.

The following list describes what happens if the quota target is a special Windows group ID:

- If the quota target is the Everyone group, a file whose ACL shows that the owner is Everyone is counted under the SID for Everyone.
- If the quota target is BUILTIN\Administrators, the entry is considered a user quota, for tracking only.

  You cannot impose restrictions on BUILTIN\Administrators.

If a member of BUILTIN\Administrators creates a file, the file is owned by BUILTIN \Administrators and is counted under the SID for BUILTIN\Administrators, not the user's personal SID.

Note: Data ONTAP does not support group quotas based on Windows group IDs. If you specify a Windows group ID as the quota target, the quota is considered to be a user quota.

How quotas are applied to users with multiple IDs

A user can be represented by multiple IDs. You can set up a single user quota for such a user by specifying a list of IDs as the quota target. A file owned by any of these IDs is subject to the restriction of the user quota.

Suppose a user has the UNIX UID 20 and the Windows IDs corp\john_smith and engineering\jsmith. For this user, you can specify a quota where the quota target is a list of the UID and Windows IDs. When this user writes to the storage system, the specified quota applies, regardless of whether the write originates from UID 20, corp\john_smith, or engineering\jsmith.

Note: Separate quota file entries are considered separate targets, even if the IDs belong to the same user.

For example, for the same user you can specify one quota that limits UID 20 to 1 GB of disk space and another quota that limits corp\john_smith to 2 GB of disk space, even though both IDs represent the same user. Data ONTAP applies quotas to UID 20 and corp\john_smith separately.

In this case, no limits are applied to engineering\jsmith, even though limits are applied to the other IDs used by the same user.

How Data ONTAP determines user IDs in a mixed environment

If you have users accessing your Data ONTAP storage from both Windows and UNIX clients, then both Windows and UNIX security are used to determine file ownership. Several factors determine whether Data ONTAP uses a UNIX or Windows ID when applying user quotas.

If the security style of the qtree or volume that contains the file is only NTFS or only UNIX, then the security style determines the type of ID used when applying user quotas. For qtrees with the mixed security style, the type of ID used is determined by whether the file has an ACL.

The following table summarizes what type of ID is used:

Security Style	**ACL**	**No ACL**
UNIX	UNIX ID	UNIX ID
Mixed	Windows ID	UNIX ID
NTFS	Windows ID	Windows ID

Note: If a file is in a volume or qtree with a particular security style, but is owned by a user of the other type, and no mapping to the determined type exists, then Data ONTAP uses the default user ID for the determined type as defined in the following options:

- `wafl.default_nt_user`
- `wafl.default_unix_user`

For example, suppose the winfile file is in a qtree with the UNIX security style, and it is owned by Windows user corp\bob. If there is no mapping between corp\bob and a UNIX user id in the quotas file, the winfile file is charged against the user defined by the `wafl.default_nt_user` option.

How quotas with multiple users work

When you put multiple users in the same quota target, the quota limits defined by that quota are not applied to each individual user; in this case, the quota limits are *shared* among all users listed in the quota target.

Note: If you combine separate user quotas into one multi-user quota, you can activate the change by resizing quotas. However, if you want to remove users from a quota target with multiple users, or add users to a target that already has multiple users, you must reinitialize quotas before the change takes effect.

Example of more than one user in a quotas file entry

In the following example, there are two users listed in the quota entry:

```
#Quota       target type    disk files thold sdisk sfile
#----------  ----           ---- ----- ----- ----- -----
jsmith,chen  user@/vol/vol1 80M
```

The two users can use up to 80 MB of space combined. If one uses 75 MB, then the other one can use only 5 MB.

How you link UNIX and Windows names for quotas

In a mixed environment, users can log in as either Windows users or UNIX users. You can configure quotas to recognize that a user's UNIX id and Windows ID represent the same user.

How you map names using the same quotas file entry

You can map Windows to UNIX names by putting them together in the same entry in the quotas file. However, this requires a quotas file entry for every user.

Example

The following quotas file entry links the Windows ID corp\jroberts to the UNIX ID roberts for quotas:

```
roberts,corp\jroberts user@/vol/vol2 900M 30K
```

How you map names using the QUOTA_PERFORM_USER_MAPPING directive

If you have configured the system's `/etc/usermap.cfg` file with a one-to-one correspondence between UNIX names and Windows names, the QUOTA_PERFORM_USER_MAPPING directive in the quotas file automatically links the names. You do not have to add a separate entry for each user.

When you use this directive, Data ONTAP consults the `usermap.cfg` file to map the user names. When a UNIX and Windows name are mapped together, they are treated as the same person for determining quota usage.

For more information about the `usermap.cfg` file, see the *Data ONTAP File Access and Protocols Management Guide for 7-Mode*.

Note: If you use the user mapping feature, make sure that CIFS is either completely configured or not licensed. Leaving CIFS installed but not configured while using user mapping could impact storage system performance.

Note: This directive requires a one-to-one correspondence between Windows names and UNIX names. If a name maps to more than one name in the `usermap.cfg` file, there are duplicate entries in the quotas file and you might see unpredictable results.

Note: If you are using this directive, when you make changes to the `usermap.cfg` file, you must turn quotas off and back on before your changes will take effect.

Example

The following example illustrates the use of the QUOTA_PERFORM_USER_MAPPING directive:

```
QUOTA_PERFORM_USER_MAPPING ON
roberts        user@/vol/vol2         900M      30K
corp\stevens   user@/vol/vol2         900M      30K
QUOTA_PERFORM_USER_MAPPING OFF
```

If the `usermap.cfg` file maps roberts to corp\jroberts, the first quota entry applies to the user whose UNIX name is roberts and whose Windows name is corp\jroberts. A file owned by either user name is subject to the restriction of this quota entry.

If the `usermap.cfg` file maps corp\stevens to cws, the second quota entry applies to the user whose Windows name is corp\stevens and whose UNIX name is cws. A file owned by either user name is subject to the restriction of this quota entry.

The effect of this example could also be achieved with multiple user names in a single quotas file entry, as in the following example:

```
roberts,corp\jroberts        user@/vol/vol2         900M      30K
corp\stevens,cws             user@/vol/vol2         900M      30K
```

About using wildcard entries in the usermap.cfg file

The use of wildcard entries in the `/etc/usermap.cfg` file causes ambiguity because all trusted domains are searched in an unspecified order for a match. To prevent this problem, you should specify the order in which Data ONTAP searches domains by using the `cifs.search_domains` option.

Unexpected results might occur if your `usermap.cfg` file contains the following entry:

```
*\*     *
```

If you use the QUOTA_PERFORM_USER_MAPPING directive in your quotas file with this wildcard entry in the `usermap.cfg` file, Data ONTAP tries to find users in one of the trusted domains. However, because Data ONTAP searches domains in an unspecified order, the results of this search can be unpredictable.

To address this issue, you can specify the order that Data ONTAP searches domain by using the `cifs.search_domains` option.

How quotas work with qtrees

You can create quotas with a qtree as their target; these quotas are called *tree quotas*. You can also create user and group quotas for a specific qtree. In addition, quotas for a volume are sometimes inherited by the qtrees contained by that volume.

How tree quotas work

You can create a quota with a qtree as its target to limit how large the target qtree can become. These quotas are also called *tree quotas*.

When you apply a quota to a qtree, the result is similar to a disk partition, except that you can change the qtree's maximum size at any time by changing the quota. When applying a tree quota, Data ONTAP limits the disk space and number of files in the qtree, regardless of their owners. No users, including root and members of the BUILTIN\Administrators group, can write to the qtree if the write operation causes the tree quota to be exceeded.

Note: The size of the quota does not guarantee any specific amount of available space. The size of the quota can be larger than the amount of free space available to the qtree. You can use the `df` command to determine the true amount of available space in the qtree.

How user and group quotas work with qtrees

Tree quotas limit the overall size of the qtree. To prevent individual users or groups from consuming the entire qtree, you specify a user or group quota for that qtree.

Example user quota in a qtree

Suppose you have the following quotas file:

```
#Quota target type            disk files thold sdisk sfile
#-----------  ----            ---- ----- ----- ----- -----
*             user@/vol/vol1  50M   -     45M
jsmith        user@/vol/vol1  80M   -     75M
```

It comes to your attention that a certain user, kjones, is taking up too much space in a critical qtree, qt1, which resides in vol2. You can restrict this user's space by adding the following line to the quotas file:

```
kjones        user@/vol/vol2/qt1  20M  -   15M
```

How default tree quotas on a volume create derived tree quotas

When you create a default tree quota on a volume, corresponding derived tree quotas are automatically created for every qtree in that volume.

These derived tree quotas have the same limits as the default tree quota. If no additional quotas exist, the limits have the following effects:

- Users can use as much space in a qtree as they are allotted for the entire volume (provided they did not exceed the limit for the volume by using space in the root or another qtree).
- Each of the qtrees can grow to consume the entire volume.

The existence of a default tree quota on a volume continues to affect all new qtrees that are added to the volume. Each time a new qtree is created, a derived tree quota is also created.

Like all derived quotas, derived tree quotas display the following behaviors:

- Are created only if the target does not already have an explicit quota.
- Appear in quota reports.

Example of derived tree quotas

You have a volume with three qtrees (proj1, proj2, and proj3) and the only tree quota is an explicit quota on the proj1 qtree limiting its disk size to 10 GB. If you create a default tree quota on the volume and reinitialize quotas on the volume, the quota report now contains four tree quotas:

```
filer1>quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- --------- -------- --------- --------- ------- ------- ---------------
tree      1       vol1     proj1       0    10485760       1       - /vol/vol1/proj1
tree      *       vol1     -           0    20971520       0       -  *
tree      2       vol1     proj2       0    20971520       1       - /vol/vol1/proj2
tree      3       vol1     proj3       0    20971520       1       - /vol/vol1/proj3
...
```

The first line shows the original explicit quota on the proj1 qtree. This quota remains unchanged.

The second line shows the new default tree quota on the volume. The asterisk (*) Quota Specifier indicates it is a default quota. This quota is a result of the quota rule that you created.

The last two lines show new derived tree quotas for the proj2 and proj3 qtrees. Data ONTAP automatically created these quotas as a result of the default tree quota on the volume. These derived tree quotas have the same 20-GB disk limit as the default tree quota on the volume. Data ONTAP did not create a derived tree quota for the proj1 qtree because the proj1 qtree already had an explicit quota.

How default user quotas on a volume affect quotas for the qtrees in that volume

If a default user quota is defined for a volume, a default user quota is automatically created for every qtree contained by that volume for which an explicit or derived tree quota exists.

If a default user quota on the qtree already exists, it remains unaffected when the default user quota on the volume is created.

The automatically created default user quotas on the qtrees have the same limits as the default user quota you create for the volume.

An explicit user quota for a qtree overrides (replaces the limits applied by) the automatically created default user quota, the same way as it overrides a default user quota on that qtree that was created by an administrator.

How qtree changes affect quotas

When you delete, rename, or change the security style of a qtree, the quotas applied by Data ONTAP might change, depending on the current quotas being applied.

How deleting a qtree affects tree quotas

When you delete a qtree, all quotas applicable to that qtree, whether they are explicit or derived, are no longer applied by Data ONTAP.

If you create a new qtree with the same name as the one you deleted, the quotas previously applied to the deleted qtree are not applied automatically to the new qtree until you reinitialize quotas. If a default tree quota exists, Data ONTAP creates new derived quotas for the new qtree.

If you don't create a new qtree with the same name as the one you deleted, you can delete the quotas that applied to that qtree to avoid getting errors when you reinitialize quotas.

How renaming a qtree affects quotas

When you rename a qtree, its ID does not change. As a result, all quotas applicable to the qtree continue to be applicable, without reinitializing quotas. However, before you reinitialize quotas, you must update the quota with the new qtree name to ensure that the quota continues to be applied for that qtree.

How changing the security style of a qtree affects user quotas

You can apply Access Control Lists (ACLs) on qtrees by using NTFS or mixed security styles, but not by using the UNIX security style. Therefore, changing the security style of a qtree might affect how quotas are calculated. You should always reinitialize quotas after you change the security style of a qtree.

If you change the security style of a qtree from NTFS or mixed to UNIX, any ACLs on files in that qtree are ignored and the file usage is charged against the UNIX user IDs.

If you change the security style of a qtree from UNIX to either mixed or NTFS, the previously hidden ACLs become visible. In addition, any ACLs that were ignored become effective again, and the NFS user information is ignored. If no ACL existed before, the NFS information continues to be used in the quota calculation.

Note: To make sure that quota usages for both UNIX and Windows users are properly calculated after you change the security style of a qtree, you must reinitialize quotas for the volume containing that qtree.

Example

The following example shows how a change in the security style of a qtree results in a different user being charged for the usage of a file in the particular qtree.

Suppose NTFS security is in effect on qtree A, and an ACL gives Windows user corp\joe ownership of a 5-MB file. User corp\joe is charged with 5 MB of disk space usage for qtree A.

Now you change the security style of qtree A from NTFS to UNIX. After quotas are reinitialized, Windows user corp\joe is no longer charged for this file; instead, the UNIX user corresponding to the UID of the file is charged for the file. The UID could be a UNIX user mapped to corp\joe or the root user.

Related concepts

How Data ONTAP determines user IDs in a mixed environment on page 305

Differences among hard, soft, and threshold quotas

Hard quotas prevent operations while soft quotas trigger notifications.

Hard quotas impose a hard limit on system resources; any operation that would result in exceeding the limit fails. The following settings create hard quotas:

- Disk field
- Files field

Soft quotas send a warning message when resource usage reaches a certain level, but do not affect data access operations, so you can take appropriate action before the quota is exceeded. The following settings create soft quotas:

- Threshold field
- Soft Disk field
- Soft Files field

Threshold and Soft Disk quotas enable administrators to receive more than one notification about a quota. Typically, administrators set the Threshold to a value that is only slightly smaller than the Disk limit, so that the threshold provides a "final warning" before writes start to fail.

Related concepts

Understanding quota notifications on page 295

How the quotas file works

The quotas file, found in the /etc directory, contains one or more entries specifying limit or tracking quotas for qtrees, groups, and users. The file can contain default (general) and specific entries.

The syntax of quota entries

The syntax of a quota entry in the quotas file is *quota_target type*[@/vol/*dir/qtree_path*] *disk* [*files*] [*threshold*] [*soft_disk*] [*soft_files*]. Fields are separated by space characters or tabs.

How the Quota Target field works

The Quota Target field specifies the name of the qtree, group, or user to which this quota is being applied. An asterisk (*) in this field denotes a default quota, which is applied to all members of the type specified in this entry that do not have an explicit quota.

If you create multiple explicit quotas with the same target, only the first quota with that target is accepted and applied. The others are rejected and do not take effect.

Related concepts

Quota targets and types on page 296

How the Type field works

The Type field specifies the type of entity (qtree, group, or user) to which this quota is being applied. If the type is user or group, this field can optionally restrict the quota to a specific volume, directory, or qtree.

The Type field specifies the quota type, which can be one of the following types:

- User or group quotas, which specify the amount of disk space and the number of files that particular users and groups can own.
- Tree quotas, which specify the amount of disk space and the number of files that particular qtrees can contain.

The following table summarizes the possible values for the Type field, along with examples.

Quota type	Value in the Type field	Sample Type field
User quota in a volume (explicit or default)	`user@/vol/volume`	`user@/vol/vol1`
User quota in a qtree (explicit or default)	`user@/vol/volume/qtree`	`user@/vol/vol0/home`

Quota type	Value in the Type field	Sample Type field
Group quota in a volume (explicit or default)	`group@/vol/volume`	`group@/vol/vol1`
Group quota in a qtree (explicit or default)	`group@/vol/volume/qtree`	`group@/vol/vol0/home`
Explicit tree quota	`tree`	`tree`
Default tree quota	`tree@/vol/volume`	`tree@/vol/vol0`

How the Disk field works

The Disk field specifies the maximum amount of disk space that the quota target can use. The value in this field represents a hard limit that cannot be exceeded.

The following list describes the rules for specifying a value in this field:

- You cannot leave the Disk field blank.
  The value that follows the Type field is always assigned to the Disk field; thus, for example, Data ONTAP regards the following two quotas file entries as equivalent:

```
#Quota Target         type           disk      files
 /export              tree           75K
 /export              tree                     75K
```

- K means 1,024 bytes, M means 2 to the 20th power or 1024 * 1024 bytes, and G means 2 to the 30th power or 1024 * 1024 * 1024 bytes.

  Note: The Disk field is not case-sensitive. Therefore, you can use K, k, M, m, G, or g.

- The maximum value you can enter in the Disk field is one of the following values (approximately equivalent to 1,023 PB):
  - 1,073,741,823G
  - 1,099,511,627,775M
  - 1,125,899,906,842,620K

  Note: If you omit the K, M, or G, Data ONTAP assumes a default value of K. The value cannot be specified in decimal notation.

- The value in the Disk field should be a multiple of 4 KB.
  If it is not, the Disk field can appear incorrect in quota reports. This happens because the Disk field is always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.
- Your quota limit can be larger than the amount of disk space available in the volume.
  In this case, a warning message is printed to the console when quotas are initialized.

- To apply a tracking quota (which tracks disk usage without imposing a limit), type a hyphen (-).

How the Files field works

The Files field specifies the maximum number of files that the quota target can own. This field is optional. The value in this field represents a hard limit that cannot be exceeded.

The following list describes the rules for specifying a value in this field:

- K means 1,024 files, M means 2 to the 20th power or 1024 * 1024 files, and G means 2 to the 30th power or 1024 * 1024 * 1024 files.
  You can omit the K, M, or G. For example, if you type 100, it means that the maximum number of files is 100.

  Note: The Files field is not case-sensitive. Therefore, you can use K, k, M, m, G, or g.
- The maximum value you can enter in the Files field is `4G` or one of the following values:
  - 4,294,967,295
  - 4,194,304K
  - 4,096M
- To apply a tracking quota (which tracks file usage without imposing a limit), type a hyphen (-).

  Note: If the quota target is root, or if you specify 0 as the UID or GID, you *must* type a hyphen.
- A blank in the Files field means there is no restriction on the number of files that the quota target can use.

  Note: If you leave the Files field blank, you cannot specify values for the Threshold, Soft Disk, or Soft Files fields.
- The Files field must be on the same line as the Disk field.
  Otherwise, the Files field is ignored.

How the Threshold field works

The Threshold field specifies the disk space threshold. If a write causes the quota target to exceed the threshold, the write still succeeds, but a warning message is logged to the storage system console and an SNMP trap is generated. This field is optional.

The following list describes the rules for specifying a value in this field:

- K means 1,024 bytes, M means 2 to the 20th power or 1024 * 1024 bytes, and G means 2 to the 30th power or 1024 * 1024 * 1024 bytes.

  Note: The Threshold field is not case-sensitive. Therefore, you can use K, k, M, m, G, or g.
- The maximum value you can enter in the Threshold field is one of the following values (roughly equivalent to 1,023 PB):
  - 1,073,741,823G
  - 1,099,511,627,775M

  - 1,125,899,906,842,620K

    Note: If you omit the K, M, or G, Data ONTAP assumes a default value of K. The value cannot be specified in decimal notation.

- The value in the Threshold field, if any, should be a multiple of 4 KB.
  If it is not, the Threshold field can appear incorrect in quota reports. This happens because the Threshold field is always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.
- The Threshold field must be on the same line as the Disk field.
  Otherwise, the Threshold field is ignored.
- If you do not want to specify a threshold for the quota target, enter a hyphen (-) in this field or leave it blank.

How the Soft Disk field works

The Soft Disk field specifies the amount of disk space that the quota target can use before a warning is issued. If the quota target exceeds the soft limit, a warning message is logged to the storage system console and an SNMP trap is generated. This field is optional, and works the same way as the Threshold field.

The following list describes the rules for specifying a value in this field:

- K means 1,024 bytes, M means 2 to the 20th power or 1024 * 1024 bytes, and G means 2 to the 30th power or 1024 * 1024 * 1024 bytes.

  Note: The Soft Disk field is not case-sensitive. Therefore, you can use K, k, M, m, G, or g.
- The maximum value you can enter in the Soft Disk field is one of the following values (roughly equivalent to 1,023 PB):
  - 1,073,741,823G
  - 1,099,511,627,775M
  - 1,125,899,906,842,620K

  Note: If you omit the K, M, or G, Data ONTAP assumes a default value of K. The value cannot be specified in decimal notation.
- The value in the Threshold field, if any, should be a multiple of 4 KB.
  If it is not, the Soft Disk field can appear incorrect in quota reports. This happens because the Soft Disk field is always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.
- The Soft Disk field must be on the same line as the Disk field.
  Otherwise, the Soft Disk field is ignored.
- If you do not want to specify a soft disk limit for the quota target, enter a hyphen (-) in this field or leave it blank.

How the Soft Files field works

The Soft Files field specifies the number of files that the quota target can use before a warning is issued. If the quota target exceeds the soft limit, a warning message is logged to the storage system console and an SNMP trap is generated. This is an optional field.

The following list describes the rules for specifying a value in the Soft Files field:

- K means 1,024 files, M means 2 to the 20th power or 1024 * 1024 files, and G means 2 to the 30th power or 1024 * 1024 * 1024 files.

  You can omit the K, M, or G. For example, if you type 100, it means that the soft limit on the number of files is 100.

  Note: The Soft Files field is not case-sensitive. Therefore, you can use K, k, M, m, G, or g.
- The maximum value you can enter in the Soft Files field is 4G or one of the following values:
  - 4,294,967,295
  - 4,194,304K
  - 4,096M
- A blank in the Soft Files field means there is no soft quota on the number of files that the quota target can use.
- The Soft Files field must be on the same line as the Disk field.

  Otherwise, the Soft Files field is ignored.

How Data ONTAP reads the quotas file

There are a few simple rules to follow to ensure that Data ONTAP can read your quotas file properly.

An entry in the quotas file can extend to multiple lines. However, the Files, Threshold, Soft Disk, and Soft Files fields must be on the same line as the Disk field; otherwise, they are ignored.

If you do not want to specify a value for a field in the middle of an entry, you can use a dash (-).

Any text after a pound sign (#) is considered a comment.

Entries in the quotas file can be in any order. After Data ONTAP receives a write request, it grants access only if the request meets the requirements specified by all quotas entries.

If you create multiple explicit quotas file entries with the same target, only the first quota with that target is accepted and applied. The others are rejected and do not take effect.

What character encodings are supported by the quotas file

The quotas file supports two types of character encoding: Unicode and root volume UNIX encoding (the language specified for the root volume using the `vol lang` command).

You can edit the quotas file from either a PC or a UNIX workstation. Data ONTAP can detect whether a file was edited and saved by a Unicode-capable editor, such as Notepad. If so, Data

ONTAP considers all entries in the file to be in Unicode. Otherwise, Data ONTAP considers the entries to be in the root volume UNIX encoding.

Standard Generalized Markup Language (SGML) entities are allowed only in the root volume UNIX encoding.

Note: If you want to include non-ASCII characters in your quotas file, you must use Unicode or SGML.

Sample quotas file

A short example quotas file, together with explanations, can help you to understand the different types of quota entries and how they affect your quotas.

The following sample quotas file contains both default and explicit quotas:

```
#Quota Target type                disk  files  thold  sdisk  sfile
#------------ ----                ----  -----  -----  -----  -----
*            user@/vol/vol1       50M   15K
*            group@/vol/vol1      750M  85K
*            tree@/vol/vol1       100M  75K
jdoe         user@/vol/vol1/proj1 100M  75K
msmith       user@/vol/vol1       75M   75K
msmith       user@/vol/vol1/proj1 75M   75K
```

This quotas file has the following effects:

- Any user not otherwise mentioned in this file can use 50 MB of disk space and 15,360 files in the vol1 volume.
- Any group not otherwise mentioned in this file can use 750 MB of disk space and 87,040 files in the vol1 volume.
- Any qtree in the vol1 volume not otherwise mentioned in this file can use 100 MB of disk space and 76,800 files.
- If a qtree is created in the vol1 volume (for example, a qtree named /vol/vol1/proj2), Data ONTAP enforces a derived default user quota and a derived default group quota that have the same effect as the following quota entries:

  ```
  *               user@/vol/vol1/proj2     50M      15K
  *               group@/vol/vol1/proj2   750M      85K
  ```

- If a qtree is created in the vol1 volume (for example, a qtree named /vol/vol1/proj2), Data ONTAP tracks the disk space and number of files owned by UID 0 and GID 0 in the /vol/vol1/proj2 qtree. This is due to the following quotas file entry:

  ```
  *               tree@/vol/vol1          100M      75K
  ```

- A user named msmith can use 75 MB of disk space and 76,800 files in the vol1 volume because an explicit quota for this user exists in the /etc/quotas file, overriding the default limit of 50 MB of disk space and 15,360 files.

- By giving jdoe and msmith 100 MB and 75 MB explicit quotas for the proj1 qtree, which has a tree quota of 100MB, that qtree becomes oversubscribed. This means that the qtree could run out of space before the user quotas are exhausted.

  Note: Quota oversubscription is supported; however, a warning is printed alerting you to the oversubscription.

How quotas are activated

New quotas and changes to quotas do not take effect until they are activated. Knowing how quota activation works can help you manage your quotas less disruptively.

You can activate quotas at the volume level.

Your quotas file does not need to be free of all errors to activate quotas. Invalid entries are reported and skipped. If the quotas file contains any valid entries, the quotas are activated.

Quotas are activated either by *initializing* (turning them on) or by *resizing*. Turning off quotas and turning them on again is called reinitializing.

The length of the activation process and its impact on quota enforcement depends on the type of activation:

- The initialization process involves two parts: a `quota on` command and a quota scan of the volume's entire file system. The scan begins after the `quota on` command completes successfully. The quota scan can take some time; the more files that the volume has, the longer it takes. Until the scan is finished, quota activation is not complete and quotas are not enforced.
- The resize process involves only a `quota resize` command. Because it does not involve a quota scan, resizing takes less time than a quota initialization. During a resize process, quotas are enforced.

By default, the `quota on` and `quota resize` commands run in the background, which permits you to use other commands at the same time.

Errors and warnings from the activation process are sent to the system log messages. Using the `-w` option with the `quota on` command displays the error messages as part of the command output; this is useful if you are reinitializing from a script. (The `quota resize` command does not have a `-w` option.)

Quota activation persists across halts and reboots. You should not activate quotas in the `/etc/rc` file. The process of quota activation does not affect the availability of the storage system data.

Related concepts

When you can use resizing on page 320
When a full quota reinitialization is required on page 321

When you can use resizing

Because quota resizing is faster than quota initialization, you should use resizing whenever possible. However, resizing only works for certain types of quota changes.

You can resize quotas when making the following types of changes to the quotas file:

- Changing an existing quota.
  For example, changing the limits of an existing quota.
- Adding a quota for a quota target for which a default quota or a default tracking quota exists.
- Deleting a quota for which a default quota or default tracking quota entry is specified.
- Combining separate user quotas into one multi-user quota.

Attention: After you have made extensive quotas changes, you should perform a full reinitialization to ensure that all of the changes take effect.

Note: If you attempt to resize and not all of your quota changes can be incorporated by using a resize operation, Data ONTAP issues a warning.

You can determine from the quota report whether your storage system is tracking disk usage for a particular user, group, or qtree. If you see a quota in the quota report, it means that the storage system is tracking the disk space and the number of files owned by the quota target.

Example quotas changes that can be made effective by resizing

Some quotas file changes can be made effective by resizing. Consider the following quotas:

```
#Quota Target type             disk  files thold sdisk sfile
#------------ ----             ----  ----- ----- ----- -----
*             user@/vol/vol2    50M   15K
*             group@/vol/vol2  750M   85K
*             tree@/vol/vol2      -     -
jdoe          user@/vol/vol2/  100M   75K
kbuck         user@/vol/vol2/  100M   75K
```

Suppose you make the following changes:

- Increase the number of files for the default user target.
- Add a new user quota for a new user, boris, that needs more disk limit than the default user quota.
- Delete the kbuck user's explicit quota entry; the new user now needs only the default quota limits.

These changes result in the following quotas:

```
#Quota Target type             disk  files thold sdisk sfile
#------------ ----             ----  ----- ----- ----- -----
```

```
*              user@/vol/vol2    50M   25K
*              group@/vol/vol2  750M   85K
*              tree@/vol/vol2      -     -
jdoe           user@/vol/vol2/  100M   75K
boris          user@/vol/vol2/  100M   75K
```

Resizing activates all of these changes; a full quota reinitialization is not necessary.

Related concepts

How quota reports work on page 322

When a full quota reinitialization is required

Although resizing quotas is faster, you must do a full quota reinitialization if you make certain or extensive changes to your quotas.

A full quota reinitialization is necessary in the following circumstances:

- You create a quota for a target that has not previously had a quota.
- You change user mapping in the `usermap.cfg` file and you use the QUOTA_PERFORM_USER_MAPPING entry in the quotas file.
- You change the security style of a qtree from UNIX to either mixed or NTFS.
- You change the security style of a qtree from mixed or NTFS to UNIX.
- You remove users from a quota target with multiple users, or add users to a target that already has multiple users.
- You make extensive changes to your quotas.

Example quotas changes that require initialization

Suppose you have a volume that contains three qtrees and the only quotas in the volume are three tree quotas. You decided to make the following changes:

- Add a new qtree and create a new tree quota for it.
- Add a default user quota for the volume.

Both of these changes require a full quota initialization. Resizing would not make the quotas effective.

Related concepts

How you map names using the QUOTA_PERFORM_USER_MAPPING directive on page 307

How quotas work with vFiler units

When you create vFiler units, or move resources between vFiler units, quotas for the containing volume are deactivated.

After you create vFiler units or reassign resources between vFiler units, you should ensure that quotas are on.

Note: If having quotas briefly deactivated is disruptive to any applications, you should disable those applications before assigning resources to vFiler units.

How quota reports work

Quota reports enable you to see what quotas Data ONTAP is applying. You can change the format of the quota report and how user IDs are displayed using the options for the `quota report` command.

What fields quota reports contain

Some quota report fields are always displayed; others depend on what options you use for the `quota report` command.

The following table lists the headings that can appear in quota reports, with a description and the option required to display that heading if needed.

Quota report heading	Description
Type	Quota type: user, group, or tree.
ID	User ID, UNIX group name, qtree name. If the quota is a default quota, the value in this field is an asterisk.
Volume	Volume to which the quota is applied.
Tree	Qtree to which the quota is applied.
K-Bytes Used	Current amount of disk space used by the quota target. If the quota is a default quota, the value in this field is 0.
Limit	Maximum amount of disk space that can be used by the quota target (the value in the Disk field of the quotas file).

Quota report heading	Description
S-Limit	Maximum amount of disk space that can be used by the quota target before a warning is issued (the value in the Soft Disk field of the quotas file). This column is displayed only when you use the `-s` option for the `quota report` command.
T-hold	Disk space threshold (the value in the Threshold field of the quotas file). This column is displayed only when you use the `-t` option for the `quota report` command.
Files Used	Current number of files used by the quota target. If the quota is a default quota, the value in this field is 0.
Limit	Maximum number of files allowed for the quota target (the value in the File field of the quotas file).
S-Limit	Maximum number of files that can be used by the quota target before a warning is issued (the value in the Soft Files field of the quotas file). This column is displayed only when you use the `-s` option for the `quota report` command.
VFiler	Displays the name of the vFiler unit for this quota entry. This column is displayed only when you use the `-v` option for the `quota report` command. This option is available only on storage systems that have MultiStore licensed.
Quota Specifier	For an explicit quota, this field shows how the quota target is specified in the quotas file. For a derived quota, the field is blank.

How quota report options affect quota reports

What options you use for the `quota report` command affect how the report is formatted and how user IDs are displayed.

The following table lists the options for the `quota report` command with their results on the quota report:

Option	Result
none	Generates the default quota report. The ID field displays one of the IDs using the following formats: • For a Windows name, the first seven characters of the user name with a preceding backslash are displayed. The domain name is omitted. • For a SID, the last eight characters are displayed. The Quota Specifier field displays an ID that matches the one in the ID field, using the same format as the /etc/quotas file entry.
`-q`	Displays the quota target's UNIX UID, GID or Windows SID in the following formats: • UNIX UIDs and GIDs are displayed as numbers. • Windows SIDs are displayed as text. **Note:** Data ONTAP does not perform a lookup of the name associated with the target ID.
`-s`	The soft limit (S-limit) columns are included.
`-t`	The threshold (T-hold) column is included.
`-v`	The vFiler column is included.
`-u`	Displays multiple IDs for your quota targets. The ID field displays all the IDs listed in the quota target of a user quota in the following format: • On the first line, the format is the same as the default format. • Each additional name in the quota target is displayed, in its entirety, on a separate line. The Quota Specifier field displays the list of IDs specified in the quota target. **Note:** You cannot combine the `-u` and `-x` options.
`-x`	Displays all the quota target's IDs on the first line of that quota target's entry, as a comma separated list. **Note:** You cannot combine the `-u` and `-x` options. The threshold column is included.

How the ID field is displayed in quota reports

Usually, the ID field of the quota report displays a user name instead of a UID or SID. However, there are some exceptions to this rule.

The ID field does *not* display a user name in the following circumstances:

- For a quota with a UNIX user as the target, the ID field shows the UID instead of a name if either of the following conditions applies:
  - No user name for the UID is found in the password database.
  - You specifically request the UID by including the `-q` option for the `quota reports` command.
- For a quota with a Windows user as the target, the ID field shows the SID instead of a name if either of the following conditions applies:
  - The SID is specified as a quota target and the SID no longer corresponds to a user name.
  - Data ONTAP cannot find an entry for the SID in the SID-to-name map cache and cannot connect to the domain controller to ascertain the user name for the SID when it generates the quota report.

How you can use the quota report to see what quotas are in effect

Because of the various ways that quotas interact, more quotas are in effect than just the ones you have explicitly created. To see what quotas are in effect, you can view the quota report.

The following examples show quota reports for different types of quotas applied on a volume vol1, and a qtree q1 contained in that volume.

Example with no user quotas specified for the qtree

In this example, there is one qtree, q1, which is contained by the volume vol1. The administrator has created three quotas:

- A default tree quota limit on vol1 of 400 MB
- A default user quota limit on vol1 of 100 MB
- An explicit user quota limit on vol1 of 200 MB for the user jsmith

The quotas file for these quotas looks similar to the following excerpt:

```
#Quota target type            disk files  thold sdisk  sfile
#------------  ----            ---  -----  ----  -----  -----
*             tree@/vol/vol1  400M
*             user@/vol/vol1  100M
jsmith        user@/vol/vol1  200M
```

The quota report for these quotas looks similar to the following excerpt:

```
sys1> quota report
                                  K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
tree         *     vol1        -         0    409600       0       - *
user         *     vol1        -         0    102400       0       - *
user    jsmith     vol1        -       112    204800       7       - jsmith
tree         1     vol1       q1         0    409600       6       - /vol/vol1/q1
user         *     vol1       q1         0    102400       0       -
user    jsmith     vol1       q1         0    102400       5       -
user      root     vol1       q1         0         -       1       -
user      root     vol1        -         0         -       8       -
```

The first three lines of the quota report display the three quotas specified by the administrator. Since two of these quotas are default quotas, Data ONTAP automatically creates derived quotas.

The fourth line displays the tree quota that is derived from the default tree quota for every qtree in vol1 (in this example, only q1).

The fifth line displays the default user quota that is created for the qtree as a result of the existence of the default user quota on the volume and the qtree quota.

The sixth line displays the derived user quota that is created for jsmith on the qtree because there is a default user quota for the qtree (line 5) and the user jsmith owns files on that qtree. Note that the limit applied to the user jsmith in the qtree q1 is not determined by the explicit user quota limit (200 MB). This is because the explicit user quota limit is on the volume, so it does not affect limits for the qtree. Instead, the derived user quota limit for the qtree is determined by the default user quota for the qtree (100 MB).

The last two lines display more user quotas that are derived from the default user quotas on the volume and on the qtree. A derived user quota was created for the root user on both the volume and the qtree because the root user owned files on both the volume and the qtree. Since the root user gets special treatment in terms of quotas, its derived quotas are tracking quotas only.

Example with user quotas specified for the qtree

This example is similar to the previous one, except that the administrator has added two quotas on the qtree.

There is still one volume, vol1, and one qtree, q1. The administrator has created the following quotas:

- A default tree quota limit on vol1 of 400 MB
- A default user quota limit on vol1 of 100 MB
- An explicit user quota limit on vol1 for the user jsmith of 200 MB
- A default user quota limit on qtree q1 of 50 MB
- An explicit user quota limit on qtree q1 for the user jsmith of 75 MB

The quotas file for these quotas looks like this:

```
#Quota target type              disk files  thold sdisk  sfile
#------------  ----              ---  -----  ----  -----  -----
*              tree@/vol/vol1    400M
*              user@/vol/vol1    100M
jsmith         user@/vol/vol1    200M
*              user@/vol/vol1/q1 50M
jsmith         user@/vol/vol1/q1 75M
```

The quota report for these quotas looks like this:

```
sys1> quota report
                                    K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
tree         *     vol1        -         0    409600       0       - *
user         *     vol1        -         0    102400       0       - *
user    jsmith     vol1        -       112    204800       7       - jsmith
user         *     vol1       q1         0     51200       0       - *
user    jsmith     vol1       q1         0     76800       5       - jsmith
tree         1     vol1       q1         0    409600       6       - /vol/vol1/q1
user      root     vol1        -         0         -       2       -
user      root     vol1       q1         0         -       1       -
```

The first five lines of the quota report display the five quotas created by the administrator. Since some of these quotas are default quotas, Data ONTAP automatically creates derived quotas.

The sixth line displays the tree quota that is derived from the default tree quota for every qtree in vol1 (in this example, only q1).

The last two lines display the user quotas that are derived from the default user quotas on the volume and on the qtree. A derived user quota was created for the root user on both the volume and the qtree because the root user owned files on both the volume and the qtree. Since the root user gets special treatment in terms of quotas, its derived quotas are tracking quotas only.

No other default quotas or derived quotas were created for the following reasons:

- A derived user quota was not created for the jsmith user even though the user owns files on both the volume and the qtree because the user already has explicit quotas at both levels.
- No derived user quotas were created for other users because no other users own files on either the volume or the qtree.
- The default user quota on the volume did not create a default user quota on the qtree because the qtree already had a default user quota.

Related concepts

How default user quotas on a volume affect quotas for the qtrees in that volume on page 310
How you use explicit quotas on page 298
How default quotas work on page 297

How derived quotas work on page 299

Related tasks

Using the quota report to determine which quotas limit writes to a specific file on page 341

Difference in space usage displayed by a quota report and a UNIX client

The value of used disk space that is displayed in a quota report for a volume or qtree can be different from the value displayed by a UNIX client for the same volume or qtree. The difference in usage values is because of the difference in methods followed by the quota report and the UNIX client for calculating the data blocks in the volume or qtree.

For example, if a volume contains a file that has empty data blocks (to which data is not written), the quota report for the volume does not count the empty data blocks while reporting the space usage. However, when the volume is mounted on a UNIX client and the file is shown as the output of the `ls` command, the empty data blocks are also included in the space usage. Therefore, the `ls` command displays a higher file size when compared to the space usage displayed by the quota report.

Similarly, the space usage values shown in a quota report can also differ from the values shown as a result of UNIX commands such as `df` and `du`.

How a quota report accounts for disk space and file usage

The number of files used and the amount of disk space specified in a quota report for a volume or a qtree depend on the count of the used data blocks corresponding to every inode in the volume or the qtree.

The block count includes both direct and indirect blocks used for regular and stream files. The blocks used for directories, Access Control Lists (ACLs), stream directories, and metafiles do not get accounted for in the quota report. In case of UNIX sparse files, empty data blocks are not included in the quota report.

Related concepts

How the ls command accounts for space usage on page 329
How the df command accounts for file size on page 330
How the du command accounts for space usage on page 330

How the ls command accounts for space usage

When you use the `ls` command to view the contents of a volume mounted on a UNIX client, the file sizes displayed in the output could be lesser or more than the space usage displayed in the quota report for the volume depending on the type of data blocks for the file.

The output of the `ls` command displays only the size of a file and does not include indirect blocks used by the file. Any empty blocks of the file also get included in the output of the command.

Therefore, if a file does not have empty blocks, the size displayed by the `ls` command might be less than the disk usage specified by a quota report because of the inclusion of indirect blocks in the quota report. Conversely, if the file has empty blocks, then the size displayed by the `ls` command might be more than the disk usage specified by the quota report.

The output of the `ls` command displays only the size of a file and does not include indirect blocks used by the file. Any empty blocks of the file also get included in the output of the command.

Example of the difference between space usage accounted by the ls command and a quota report

The following quota report shows a limit of 10 MB for a qtree q1:

```
system1>quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
tree    user1      vol1     q1    10485760  10485760       1       - /vol/vol1/q1
...
```

A file present in the same qtree can have a size exceeding the quota limit when viewed from a UNIX client by using the `ls` command, as shown in the following example:

```
[user1@lin-sys1 q1]$ ls -lh
-rwxr-xr-x  1 user1 nfsuser  27M Apr 09  2013 file1
```

Related concepts

How a quota report accounts for disk space and file usage on page 328
How the df command accounts for file size on page 330
How the du command accounts for space usage on page 330

How the df command accounts for file size

When you run the `df` command from the mount point of a qtree for which a quota rule is configured, the output of the command shows the same space usage as the value specified by the quota report.

If quotas are enabled for the volume that contains the qtree, the space usage reported by the `df` command excludes blocks used by directories, ACLs, stream directories, and metafiles. Therefore, the reported space usage exactly matches the value specified by the quota report.

However, if the qtree does not have a quota rule configured or if quotas are not enabled for the volume, then the reported space usage includes the blocks consumed by directories, ACLs, stream directories and metafiles for the entire volume, including other qtrees within the volume. In such a situation, the space usage reported by the `df` command is more than the value specified by the quota report.

Example of space usage accounted by the df command and a quota report

The following quota report shows a limit of 10 MB for a qtree q1:

```
system1>quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
tree    user1      vol1     q1    10485760  10485760       1       - /vol/vol1/q1
...
```

In the following example, the space usage as the output of the `df` command shows the same limit of 10 MB (in terms of 1K blocks) because quota rules are configured for the qtree:

```
[user1@lin-sys1 q1]$ df -k
192.0.2.245:/vol/vol1/q1
                      10240  10240 0  100% /q1
```

Related concepts

How a quota report accounts for disk space and file usage on page 328
How the ls command accounts for space usage on page 329
How the du command accounts for space usage on page 330

How the du command accounts for space usage

When you run the `du` command to check the disk space usage for a qtree or volume mounted on a UNIX client, the usage value might be higher than the value displayed by a quota report for the qtree or volume.

The output of the `du` command contains the combined space usage of all the files through the directory tree beginning at the level of the directory where the command is issued. Because the usage

value displayed by the du command also includes the data blocks for directories, it is higher than the value displayed by a quota report.

Example of the difference between space usage accounted by the du command and a quota report

The following quota report shows a limit of 10 MB for a qtree q1:

```
system1>quota report
                                  K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
tree    user1      vol1     q1    10485760  10485760       1       - /vol/vol1/q1
...
```

In the following example, the disk space usage as the output of the du command shows a higher value that exceeds the quota limit:

```
[user1@lin-sys1 q1]$ du -sh
11M     q1
```

Related concepts

How a quota report accounts for disk space and file usage on page 328
How the ls command accounts for space usage on page 329
How the df command accounts for file size on page 330

Progressive quota examples

Following through a series of progressive examples can help you to understand how to create your quotas file and read your quota reports.

For the following examples, assume that you have a storage system that has one volume, vol1.

Example 1: default quota

You decide to impose a hard limit of 50 MB for each user in vol1, using the following quotas file:

```
#Quota target type            disk files  thold sdisk  sfile
#------------  ----            ---  -----  ----  -----  -----
*             user@/vol/vol1  50M
```

If any user on the system enters a command that would use more than 50 MB in vol1, the command fails (for example, writing to a file from an editor).

Example 2: default quota override

Suppose that you have received a complaint from an important user, saying that she needs more space in vol1. To give this user more space, you update your quotas file as follows (her username is jsmith):

```
#Quota target type             disk files  thold sdisk  sfile
#------------  ----             ---  -----  ----  -----  -----
*              user@/vol/vol1   50M
jsmith         user@/vol/vol1   80M
```

Now, jsmith can use up to 80 MB of space on vol1, even though all other users are still limited to 50 MB.

The quota report looks like this:

```
filer1> quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
user       *      vol1        -         0     51200       0       -    *
user      jsmith vol1        -     63275     81920      37       -    jsmith
user      root   vol1        -         0         -       1       -
```

Note that an extra quota is shown, for the root user. Default user quotas do not apply to root, so the root user has no space limit on vol1, as shown in the quota report by the dash (“-”) in the Limit column for the root user.

Example 3: thresholds

This example sets up a threshold for all users at 45 MB, except for jsmith, who will get a threshold at 75 MB. To set up a user-specific threshold, we change the quotas file to read as follows:

```
#Quota target   type            disk    files   thold   sdisk
sfile
#------------    ----            ----    -----   -----   -----
-----
*               user@/vol/vol1  50M     -       45M
jsmith          user@/vol/vol1  80M     -       75M
```

Note that it was necessary to add a dash (-) in the Files field as a placeholder because the Threshold field comes after the Files field in the quotas file.

Now the quota report looks like this:

```
filer1> quota report -t
                          K-Bytes                    Files
Type    ID    Volume  Tree  Used    Limit     T-hold   Used    Limit   Quota Specifier
------ ----- -------- ----- ------ --------  -------- ------- -------- ----------------
user    *      vol1    -     0       51200    46080     0        -     *
```

```
user    jsmith vol1    -     63280   81920    76800     47       -     jsmith
user    root   vol1    -     0       -        -         51       -
```

Note that the `-t` flag is used to display threshold limits.

Example 4: quotas on qtrees

Suppose that you decide you need to partition some space for two projects. You create two qtrees, named proj1 and proj2, to accommodate those projects within vol1. Creating qtrees does not cause any change for your quotas, because the quotas file only applies quotas to the volume so far. Users can use as much space in a qtree as they are allotted for the entire volume (provided they did not exceed the limit for the volume by using space in the root or another qtree). In addition, each of the qtrees can grow to consume the entire volume.

You decide that you want to make sure that neither qtree grows to more than 20 GB. Your quotas file now looks like this:

```
#Quota target   type             disk files thold sdisk   sfile
#-----------    ----             ---- ----- ----- -----   -----
*               user@/vol/vol1   50M   -     45M
jsmith          user@/vol/vol1   80M   -     75M
*               tree@/vol/vol1   20G
```

Note that the correct type is *tree*, not qtree.

Now your quota report looks like this:

```
filer1> quota report -t
                           K-Bytes                       Files
Type   ID    Volume    Tree  Used      Limit     T-hold   Used    Limit  Quota
Specifier
----- ------ -------- ------ --------- --------- -------- ------- ----- -------------
user   *      vol1      -           0     51200    46080     0       -  *
user   jsmith vol1      -       63280     81920    76800    55       -  jsmith
tree   *      vol1      -           0   20971520        -     0       -  *
tree   1      vol1      proj1       0   20971520        -     1       - /vol/vol1/proj1
user   *      vol1      proj1       0      51200    46080     0       -
user   root   vol1      proj1       0          -        -     1       -
tree   2      vol1      proj2       0   20971520        -     1       - /vol/vol1/proj2
user   *      vol1      proj2       0      51200    46080     0       -
user   root   vol1      proj2       0          -        -     1       -
user   root   vol1      -           0          -        -     3       -
```

Several new lines have appeared. The first new line is exactly what you added to the quotas file:

```
tree * vol1 - 0 20971520 - 0 - *
```

The next line shows what is called a *derived quota*. You did not add this quota directly. It is derived from the default tree quota that you just added. This new line means that a quota of 20 GB is being applied to the proj1 qtree:

```
tree 1 vol1 proj1 0 20971520 - 1 - /vol/vol1/proj1
```

The next line shows another derived quota. This quota is derived from the default user quota you added in an earlier example. Default user quotas on a volume are automatically inherited

for all qtrees contained by that volume, if quotas are enabled for qtrees. When you added the first qtree quota, you enabled quotas on qtrees, so this derived quota was created:
`user * vol1 proj1 0 51200 46080 0 -`
The rest of the new lines are for the root user and for the other qtree.

Example 5: user quota on a qtree

You decide to limit users to less space in the proj1 qtree than they get in the volume as a whole. You want to keep them from using any more than 10 MB in the proj1 qtree. To do so, you update the quotas file as follows:

```
#Quota target   type                disk    files   thold   sdisk   sfile
#------------   ----                ----    -----   -----   -----   -----
*               user@/vol/vol1      50M      -       45M
jsmith             user@/vol/vol1      80m      -       75M
*               tree@/vol/vol1      20G
*               user@/vol/vol1/proj1 10M
```

Now a quota report looks like this:

```
filer1> quota report
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
user          *  vol1        -         0     51200       0       -     *
user    jsmith  vol1        -         0     81920      57       -     jsmith
tree          *  vol1        -         0  20971520       0       -     *
user          *  vol1    proj1         0     10240       0       -     *
tree          1  vol1    proj1         0  20971520       1       -     /vol/vol1/proj1
tree          2  vol1    proj2         0  20971520       1       -     /vol/vol1/proj2
user          *  vol1    proj2         0     51200       0       -
user      root  vol1    proj2         0         -       1       -
user      root  vol1        -         0         -       3       -
user      root  vol1    proj1         0         -       1       -
```

The new report entry that appears as a result of the line you added is this one:
`user * vol1 proj1 0 10240 0 - *`
However, now your phone is ringing. It’s jsmith again, complaining that her quota has been decreased. You ask where she is trying to put data, and she says “in proj1.” She is being prevented from writing more data to the proj1 qtree because the quota you created to override the default user quota (to give her more space) was on the volume. But now that you have added a default user quota on the proj1 qtree, that quota is being applied and limiting all users' space in that qtree, including jsmith. You must add a new line to the quotas file overriding the qtree default quota to give her more space in the proj1 qtree:

```
jsmith user@/vol/vol1/proj1 80M
```

This adds the following line to your quota report:

```
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
user   jsmith    vol1     proj1   57864     81920       57    -       jsmith
```

Related concepts

How default quotas work on page 297
How derived quotas work on page 299
How you use explicit quotas on page 298
How the quotas file works on page 313
How quota reports work on page 322
About quotas on page 295

Managing quotas

You create, delete, and modify quotas as your users and their storage requirements and limitations change. You can also manage how quota messages are logged, and view quota reports, which help you understand what quotas Data ONTAP is applying.

Activating quotas

You activate quotas to turn quotas on and read the quotas file. You activate quotas using the `quota on` command, for one volume at a time.

Before you begin

If the quotas file contains user quotas that use Windows IDs as targets, CIFS must be running when you activate quotas.

Step

1. Enter the following command:

   `quota on [-w] vol_name`

   The `-w` option causes the command to return only after the entire quotas file has been scanned (synchronous mode). This is useful when activating quotas from a script.

   #### Example

   The following example activates quotas on a volume named vol2:

   `quota on vol2`

   Quota reinitialization is started for the specified volume. Quota reinitialization can take some time, during which storage system data is available, but quotas are not enforced for the specified volume.

Result

When quota initialization is complete, quotas are on for the specified volume. This procedure does not modify or initialize quotas for any other volume.

After you finish

If a quota initialization is still running when the storage system is upgraded, Data ONTAP terminates the quota initialization, which must be manually restarted from the beginning. For this reason, you should allow any running quota initialization to complete before upgrading your storage system.

Related concepts

How quotas are activated on page 319
About quotas on page 295

Reinitializing quotas

You reinitialize quotas by using the `quota off` command followed by the `quota on` command. This causes Data ONTAP to reread the quotas file. Reinitializing quotas takes time. In some cases resizing is more efficient.

Before you begin

If the quotas file contains user quotas that use Windows IDs as targets, CIFS must be running when you reinitialize quotas.

About this task

Depending on how many quotas you have and the size of the file system, quota reinitialization can take some time. During quota reinitialization, data access is not affected. However, quotas are not enforced until reinitialization completes.

Steps

1. If quotas are already activated for the volume on which you want to reinitialize quotas, enter the following command:

   `quota off` *`vol_name`*

   Quotas are turned off for the specified volume.

2. Enter the following command:

   `quota on [-w]` *`vol_name`*

   The `-w` option causes the command to return only after the entire quotas file has been scanned (synchronous mode). This is useful when activating quotas from a script.

   Quota reinitialization is started for the specified volume. Quota reinitialization can take some time, during which storage system data is available, but quotas are not enforced for the specified volume.

Result

When quota initialization is complete, quotas are back on for the specified volume.

Note: Quotas are not affected for any volume other than the volume specified in the `quota on` command.

Related concepts

How quotas are activated on page 319
About quotas on page 295

Deactivating quotas

You use the `quota off` command to deactivate quotas for a specific volume.

About this task

If a quota initialization is almost complete, the `quota off` command can fail. If this happens, retry the command after a minute or two.

Canceling quota initialization

If you started a quota initialization and you now want to cancel it, you can use the `quota off` command.

About this task

If a quota initialization is almost complete, the `quota off` command can fail. If this happens, the `quota on` command should finish shortly.

Resizing quotas

You use the `quota resize` command to cause Data ONTAP to reread the quotas file for the specified volume. Resizing only works for certain types of changes to the quotas file. For other changes, you need to reinitialize quotas.

Related concepts

When you can use resizing on page 320
About quotas on page 295

Deleting quotas

You can remove quota restrictions for a quota target in two ways: by changing the quotas file entry so that there is no restriction on resource use for that quota target, or by deleting the quotas file entry for that quota target.

Deleting a quota by removing resource restrictions

You can remove a quota for a specific target by removing the resource restrictions for that target. This is equivalent to changing that quota entry to a tracking quota.

Steps

1. Open the quotas file with the editor of your choice and edit the quotas file entry for the specified target so that the quota entry becomes a tracking quota.

   #### Example

   Suppose your quotas file contained the following entry for the jdoe user:

   ```
   jdoe            user@/vol/vol2/       100M    75K
   ```

   To remove the restrictions for jdoe, you edit the entry as follows:

   ```
   jdoe            user@/vol/vol2/         -       -
   ```

2. Save and close the quotas file.

   The quotas file is updated but the change is not yet effective.

After you finish

Run the `quota resize` command to cause Data ONTAP to reread the quotas file; this will cause your change to become effective.

Deleting a quota by removing the quotas file entry

You can remove a quota for a specific target by removing the quotas file entry for that target. Depending on what other quotas you have set up, you then need to resize or reinitialize quotas.

Steps

1. Open the quotas file with the editor of your choice and remove the entry for the quota you want to delete.

   Note: If the change is temporary, you can disable the quota by prepending the pound sign (#) to the line. This causes Data ONTAP to treat the line as a comment.

2. Save and close the quotas file.

   The quotas file is updated but the change is not yet effective.

After you finish

If you have a default quota or default tracking quota in place for the quota type you modified, you can use the `quota resize` command to cause Data ONTAP to reread the quotas file. Otherwise, reinitialize quotas using the `quota off` and `quota on` commands for the volume for which you modified the quota.

Managing quota message logging

You turn quota message logging on or off, for a single volume or for all volumes, using the `quota logmsg` command. You can also specify a time interval during which quota messages are not logged. This interval defaults to 60 minutes.

About this task

For more information about the `quota logmsg` command, see the na_quota(1) man page.

Displaying a quota report

You display a quota report using the `quota report` command. You can display a quota report for all quotas or for a specific file, directory, qtree or volume by specifying a pathname.

Step

1. To display a quota report, enter the following command:

   `quota report [path]`

   You can display a quota report for all quotas or for a specific file, directory, qtree or volume by specifying a path.

   You can control the format and fields displayed using the `quota report` command options. For more information on the available options, see the na_quota(1) man page.

Related concepts

How quota reports work on page 322
About quotas on page 295

Using the quota report to determine which quotas limit writes to a specific file

You can use the `quota report` command with a specific file path to determine which quota limits affect write operations to a file. This can help you understand which quota is preventing a write operation.

Step

1. Use the `quota report` command with the `filepath` parameter.

Example of showing quotas affecting a specific file

The following example shows the command and output to determine what quotas are in effect for writes to the file f4.txt, which resides in the qtree q1 in the volume vol1:

```
sys1> quota report /vol/vol1/q1/f4.txt
                                K-Bytes             Files
Type       ID    Volume    Tree  Used      Limit     Used    Limit   Quota Specifier
----- -------- -------- -------- --------- --------- ------- ------- ---------------
user    jsmith     vol1        -       112    204800       7       - jsmith
user    jsmith     vol1       q1         0     76800       5       - jsmith
tree         1     vol1       q1         0    409600       6       - /vol/vol1/q1
```

Related concepts

How you can use the quota report to see what quotas are in effect on page 325

Storage limits

There are limits for storage objects that you should consider when planning and managing your storage architecture.

Limits are listed in the following sections:

Volume limits

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Array LUNs Minimum size for root volume	N/A	Model-dependent	N/A	See the *Hardware Universe*.
Files Maximum size	16 TB	16 TB	16 TB	
Files Maximum per volume	Volume size dependent, up to 2 billion	Volume size dependent, up to 2 billion	Volume size dependent, up to 2 billion	2 billion = 2 × 10 to the 9th power.
FlexCache volumes Maximum per system	100	100	N/A	
FlexClone volumes Hierarchical clone depth	499	499	499	The maximum depth of a nested hierarchy of FlexClone volumes that can be created from a single FlexVol volume.

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
FlexVol volumes Maximum per system	Model-dependent	Model-dependent	200	See the *Hardware Universe*.
FlexVol volumes Minimum size	20 MB	20 MB	20 MB	
FlexVol volumes (32-bit) Maximum size	16 TB	16 TB	16 TB	
FlexVol volumes (64-bit) Maximum size	Model-dependent	Model-dependent	16 TB	See the *Hardware Universe*.
FlexVol root volumes Minimum size	Model-dependent	Model-dependent	Model-dependent	See the *Hardware Universe*.
LUNs Maximum per controller or HA pair	• FAS2220: 1,024 • FAS2240: 1,024 • All other models: 2,048	• FAS2220: 1,024 • FAS2240: 1,024 • All other models: 2,048	1,024	
LUNs Maximum per volume	• FAS2220: 1,024 • FAS2240: 1,024 • All other models: 2,048	• FAS2220: 1,024 • FAS2240: 1,024 • All other models: 2,048	512	
LUNs Maximum size	16 TB	16 TB	16 TB	

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Qtrees Maximum per volume	4,995	4,995	4,995	
Snapshot copies Maximum per FlexVol volume	255	255	255	The use of certain Data ONTAP capabilities could reduce this limit. For more information, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.
Traditional volumes Maximum size	16 TB	16 TB	Model-dependent	See the *Hardware Universe*.
Traditional volumes and aggregates Maximum per system	100	100	100	In an HA configuration, this limit applies to each node individually, so the overall limit for the pair is doubled.

Aggregate limits

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Aggregates and traditional volumes (combined) Maximum per system	100	100	100	In an HA configuration, this limit applies to each node individually, so the overall limit for the pair is doubled.
Aggregates (32-bit) Maximum size	16 TB	16 TB	16 TB	

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Aggregates (64-bit) Maximum size	Model-dependent	Model-dependent	16 TB	See the *Hardware Universe*.
Aggregates Minimum size	RAID-DP: 3 disks RAID4: 2 disks	Model-dependent	N/A	See the *Hardware Universe* for the minimum aggregate size for RAID0.
Aggregates (mirrored) Maximum suggested per system	64	64	N/A	You can create more than 64 mirrored aggregates on a storage system, but doing so could cause plex synchronization problems after certain types of failures.
RAID groups Maximum per aggregate	150	150	150	
Traditional volumes Maximum size	16 TB	16 TB	16 TB	

RAID group limits

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Maximum per system	400	400	400	
Maximum per aggregate	150	150	150	

RAID group sizes

RAID type	Default size	Maximum size	Minimum size
RAID-DP	SATA/BSAS/FSAS/ MSATA/ATA: 14 FC/SAS: 16 SSD: 23	SATA/BSAS/FSAS/ MSATA/ATA: 20 FC/SAS: 28 SSD: 28	3
RAID4	SATA/BSAS/FSAS/ MSATA/ATA: 7 FC/SAS/SSD: 8	SATA/BSAS/FSAS/ MSATA/ATA: 7 FC/SAS/SSD: 14	2
RAID0	8	26	1

FlexClone file and FlexClone LUN limits

Limit	Native storage	Storage arrays	Virtual storage (Data ONTAP-v)	Notes
Maximum per file or LUN	32,767	32,767	32,767	If you try to create more than 32,767 clones, Data ONTAP automatically creates a new physical copy of the parent file or LUN. This limit might be lower for FlexVol volumes that use deduplication.
Maximum total shared data per FlexVol volume	640 TB	640 TB	640 TB	

Related concepts

How Data ONTAP reports drive types on page 18
About RAID group size on page 111
How you use aggregates to provide storage to your volumes on page 126
Using volumes on page 163
Differences between 64-bit and 32-bit FlexVol volumes on page 164

Copyright information

Trademark information

How to send your comments

You can help us to improve the quality of our documentation by sending us your feedback.

Your feedback is important in helping us to provide the most accurate and high-quality information. If you have suggestions for improving this document, send us your comments by email to *doccomments@netapp.com*. To help us direct your comments to the correct division, include in the subject line the product name, version, and operating system.

You can also contact us in the following ways:

- NetApp, Inc., 495 East Java Drive, Sunnyvale, CA 94089 U.S.
- Telephone: +1 (408) 822-6000
- Fax: +1 (408) 822-4501
- Support telephone: +1 (888) 463-8277

Index

B

C

D

E

F

M

N

O

P

Q

R

S

T

U

V

W

Z