Data ONTAP® 7.1
Storage Management Guide

Network Appliance, Inc.
495 East Java Drive
Sunnyvale, CA 94089 USA
Telephone: +1 (408) 822-6000
Fax: +1 (408) 822-4501
Support telephone: +1 (888) 4-NETAPP
Documentation comments: doccomments@netapp.com
Information Web: http://www.netapp.com

Part number 210-01085_B0
December 2005

Copyright and trademark information

Copyright information

Copyright © 1994–2005 Network Appliance, Inc. All rights reserved. Printed in the U.S.A.

No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Portions of this product are derived from the Berkeley Net2 release and the 4.4-Lite-2 release, which are copyrighted and publicly distributed by The Regents of the University of California.

Copyright © 1980–1995 The Regents of the University of California. All rights reserved.

Portions of this product are derived from NetBSD, which is copyrighted by Carnegie Mellon University.

Copyright © 1994, 1995 Carnegie Mellon University. All rights reserved. Author Chris G. Demetriou.

Permission to use, copy, modify, and distribute this software and its documentation is hereby granted, provided that both the copyright notice and its permission notice appear in all copies of the software, derivative works or modified versions, and any portions thereof, and that both notices appear in supporting documentation.

CARNEGIE MELLON ALLOWS FREE USE OF THIS SOFTWARE IN ITS “AS IS” CONDITION. CARNEGIE MELLON DISCLAIMS ANY LIABILITY OF ANY KIND FOR ANY DAMAGES WHATSOEVER RESULTING FROM THE USE OF THIS SOFTWARE.

Software derived from copyrighted material of The Regents of the University of California and Carnegie Mellon University is subject to the following license and disclaimer:

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notices, this list of conditions, and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notices, this list of conditions, and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. All advertising materials mentioning features or use of this software must display the following acknowledgment:

   This product includes software developed by the University of California, Berkeley and its contributors.
4. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER

IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software contains materials from third parties licensed to Network Appliance Inc. which is sublicensed, and not sold, and title to such material is not passed to the end user. All rights reserved by the licensors. You shall not sublicense or permit timesharing, rental, facility management or service bureau usage of the Software.

Portions developed by the Apache Software Foundation (http://www.apache.org/). Copyright © 1999 The Apache Software Foundation.

Portions Copyright © 1995–1998, Jean-loup Gailly and Mark Adler

Portions Copyright © 2001, Sitraka Inc.

Portions Copyright © 2001, iAnywhere Solutions

Portions Copyright © 2001, i-net software GmbH

Portions Copyright © 1995 University of Southern California. All rights reserved.

Redistribution and use in source and binary forms are permitted provided that the above copyright notice and this paragraph are duplicated in all such forms and that any documentation, advertising materials, and other materials related to such distribution and use acknowledge that the software was developed by the University of Southern California, Information Sciences Institute. The name of the University may not be used to endorse or promote products derived from this software without specific prior written permission.

Portions of this product are derived from version 2.4.11 of the libxml2 library, which is copyrighted by the World Wide Web Consortium.

Network Appliance modified the libxml2 software on December 6, 2001, to enable it to compile cleanly on Windows, Solaris, and Linux. The changes have been sent to the maintainers of libxml2. The unmodified libxml2 software can be downloaded from http://www.xmlsoft.org/.

Copyright © 1994–2002 World Wide Web Consortium, (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University). All Rights Reserved. http://www.w3.org/Consortium/Legal/

Software derived from copyrighted material of the World Wide Web Consortium is subject to the following license and disclaimer:

Permission to use, copy, modify, and distribute this software and its documentation, with or without modification, for any purpose and without fee or royalty is hereby granted, provided that you include the following on ALL copies of the software and documentation or portions thereof, including modifications, that you make:

The full text of this NOTICE in a location viewable to users of the redistributed or derivative work.

Any pre-existing intellectual property disclaimers, notices, or terms and conditions. If none exist, a short notice of the following form (hypertext is preferred, text is permitted) should be used within the body of any redistributed or derivative code: "Copyright © [$date-of-software] World Wide Web Consortium, (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University). All Rights Reserved. http://www.w3.org/Consortium/Legal/.

Notice of any changes or modifications to the W3C files, including the date changes were made.

THIS SOFTWARE AND DOCUMENTATION IS PROVIDED "AS IS," AND COPYRIGHT HOLDERS MAKE NO REPRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, WARRANTIES OF MERCHANTABILITY OR FITNESS

FOR ANY PARTICULAR PURPOSE OR THAT THE USE OF THE SOFTWARE OR DOCUMENTATION WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

COPYRIGHT HOLDERS WILL NOT BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF ANY USE OF THE SOFTWARE OR DOCUMENTATION.

The name and trademarks of copyright holders may NOT be used in advertising or publicity pertaining to the software without specific, written prior permission. Title to copyright in this software and any associated documentation will at all times remain with copyright holders.

Software derived from copyrighted material of Network Appliance, Inc. is subject to the following license and disclaimer:

Network Appliance reserves the right to change any products described herein at any time, and without notice. Network Appliance assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by Network Appliance. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of Network Appliance.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark information

NetApp, the Network Appliance logo, the bolt design, NetApp–the Network Appliance Company, DataFabric, Data ONTAP, FAServer, FilerView, MultiStore, NearStore, NetCache, SecureShare, SnapManager, SnapMirror, SnapMover, SnapRestore, SnapVault, SyncMirror, and WAFL are registered trademarks of Network Appliance, Inc. in the United States, and/or other countries. gFiler, Network Appliance, SnapCopy, Snapshot, and The Evolution of Storage are trademarks of Network Appliance, Inc. in the United States and/or other countries and registered trademarks in some other countries. ApplianceWatch, BareMetal, Camera-to-Viewer, ComplianceClock, ComplianceJournal, ContentDirector, ContentFabric, EdgeFiler, FlexClone, FlexVol, FPolicy, HyperSAN, InfoFabric, LockVault, Manage ONTAP, NOW, NOW NetApp on the Web, ONTAPI, RAID-DP, RoboCache, RoboFiler, SecureAdmin, Serving Data by Design, SharedStorage, Simulate ONTAP, Smart SAN, SnapCache, SnapDirector, SnapDrive, SnapFilter, SnapLock, SnapMigrator, SnapSuite, SnapValidator, SohoFiler, vFiler, VFM, Virtual File Manager, VPolicy, and Web Filer are trademarks of Network Appliance, Inc. in the United States and other countries. NetApp Availability Assurance and NetApp ProTech Expert are service marks of Network Appliance, Inc. in the United States. Spinnaker Networks, the Spinnaker Networks logo, SpinAccess, SpinCluster, SpinFS, SpinHA, SpinMove, and SpinServer are registered trademarks of Spinnaker Networks, LLC in the United States and/or other countries. SpinAV, SpinManager, SpinMirror, SpinRestore, SpinShot, and SpinStor are trademarks of Spinnaker Networks, LLC in the United States and/or other countries.

Apple is a registered trademark and QuickTime is a trademark of Apple Computer, Inc. in the United States and/or other countries. Microsoft is a registered trademark and Windows Media is a trademark of Microsoft Corporation in the United States and/or other countries. RealAudio, RealNetworks, RealPlayer, RealSystem, RealText, and RealVideo are registered trademarks and RealMedia, RealProxy, and SureStream are trademarks of RealNetworks, Inc. in the United States and/or other countries.

All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such.

Network Appliance is a licensee of the CompactFlash and CF Logo trademarks.

Network Appliance NetCache is certified RealSystem compatible.

Table of Contents

Preface xi

Chapter 1 **Introduction to NetApp Storage Architecture** 1

Understanding storage architecture 2

Understanding the file system and its storage containers 11

Using traditional volumes 19

Chapter 2 **Quick setup for aggregates and volumes** 23

Planning your aggregate, volume, and qtree setup 24

Configuring data storage 29

Converting from one type of volume to another 35

Overview of aggregate and volume operations 36

Chapter 3 **Disk and Storage Subsystem Management** 45

Understanding disks 46

Disk configuration and ownership 53
- Initial configuration 54
- Hardware-based disk ownership 55
- Software-based disk ownership 58

Disk access methods 68
- Multipath I/O for Fibre Channel disks 69
- Clusters 75
- Combined head and disk shelf storage systems 76
- SharedStorage 77

Disk management 85
- Displaying disk information 86
- Managing available space on new disks 94
- Adding disks 97
- Removing disks 100
- Sanitizing disks 105

Disk performance and health 117

Storage subsystem management 122
- Viewing information 123

Changing the state of a host adapter 132

Chapter 4 **RAID Protection of Data** 135

Understanding RAID groups 136

Predictive disk failure and Rapid RAID Recovery 144

Disk failure and RAID reconstruction with a hot spare disk 145

Disk failure without a hot spare disk 146

Replacing disks in a RAID group 148

Setting RAID level and group size 149

Changing the RAID level for an aggregate 152

Changing the size of RAID groups 157

Controlling the speed of RAID operations 161
Controlling the speed of RAID data reconstruction 162
Controlling the speed of disk scrubbing 163
Controlling the speed of plex resynchronization 164
Controlling the speed of mirror verification 165

Automatic and manual disk scrubs 166
Scheduling an automatic disk scrub 167
Manually running a disk scrub 170

Minimizing media error disruption of RAID reconstructions 173
Handling of media errors during RAID reconstruction 174
Continuous media scrub 175
Disk media error failure thresholds 180

Viewing RAID status 181

Chapter 5 **Aggregate Management** 183

Understanding aggregates 184

Creating aggregates 187

Changing the state of an aggregate 193

Adding disks to aggregates 198

Destroying aggregates 204

Undestroying aggregates 206

Physically moving aggregates 208

Chapter 6 **Volume Management** 211

Traditional and FlexVol volumes. 212

Traditional volume operations 215
Creating traditional volumes 216
Physically transporting traditional volumes 221

FlexVol volume operations 224
Creating FlexVol volumes 225
Resizing FlexVol volumes 229
Cloning FlexVol volumes 231
Configuring FlexVol volumes to grow automatically 240
Displaying a FlexVol volume's containing aggregate 242

General volume operations 243
Migrating between traditional volumes and FlexVol volumes 244
Managing duplicate volume names 252
Managing volume languages. 253
Determining volume status and state. 256
Renaming volumes 262
Destroying volumes 263
Increasing the maximum number of files in a volume 265
Reallocating file and volume layout 267

Managing FlexCache volumes 268
How FlexCache volumes work. 269
Sample FlexCache deployments 275
Creating FlexCache volumes. 277
Sizing FlexCache volumes 279
Administering FlexCache volumes 281

Space management for volumes and files 283
Space guarantees 286
Space reservations 292
Fractional reserve 294
Space management policies 295

Chapter 7 **Qtree Management** 297

Understanding qtrees 298

Understanding qtree creation 300

Creating qtrees 302

Understanding security styles. 303

Changing security styles 306

Changing the CIFS oplocks setting 308
Displaying qtree status 311
Displaying qtree access statistics 312
Converting a directory to a qtree 313
Renaming or deleting qtrees 316

Chapter 8 **Quota Management** 319
Understanding quotas 320
When quotas take effect 323
Understanding default quotas 324
Understanding derived quotas 325
How Data ONTAP identifies users for quotas 328
Notification when quotas are exceeded 331
Understanding the /etc/quotas file 332
Overview of the /etc/quotas file 333
Fields of the /etc/quotas file 336
Sample quota entries 342
Special entries for mapping users 345
How disk space owned by default users is counted 349
Activating or reinitializing quotas 350
Modifying quotas 353
Deleting quotas 356
Turning quota message logging on or off 358
Effects of qtree changes on quotas 360
Understanding quota reports 362
Types of quota reports 363
Overview of the quota report format 364
Quota report formats 366
Displaying a quota report 370

Glossary 371

Index 379

Preface

Introduction

This guide describes how to configure, operate, and manage the storage resources of Network Appliance™ storage systems that run Data ONTAP® 7.1 software. It covers all models. This guide focuses on the storage resources, such as disks, RAID groups, plexes, and aggregates, and how file systems, or volumes, are used to organize and manage data.

Audience

This guide is for system administrators who are familiar with operating systems, such as the UNIX®, Windows NT®, Windows 2000®, Windows Server 2003 Software®, or Windows XP® operating systems, that run on the storage system's clients. It also assumes that you are familiar with how to configure the storage system and how Network File System (NFS), Common Internet File System (CIFS), and Hypertext Transport Protocol (HTTP) are used for file sharing or transfers. This guide doesn't cover basic system or network administration topics, such as IP addressing, routing, and network topology.

Terminology

NetApp® storage products (filers, FAS appliances, and NearStore® systems) are all storage systems—also sometimes called filers or storage appliances.

The terms "flexible volumes" and "FlexVol™ volumes" are used interchangeably in Data ONTAP documentation.

This guide uses the term *type* to mean pressing one or more keys on the keyboard. It uses the term *enter* to mean pressing one or more keys and then pressing the Enter key.

Command conventions

You can enter Data ONTAP commands on the system console or from any client computer that can access the storage system through a Telnet or Secure Socket Shell (SSH)-interactive session, or through the Remote LAN Manager (RLM).

In examples that illustrate commands executed on a UNIX workstation, the command syntax and output might differ, depending on your version of UNIX.

Keyboard conventions

When describing key combinations, this guide uses the hyphen (-) to separate individual keys. For example, *Ctrl-D* means pressing the *Control* and *D* keys simultaneously. Also, this guide uses the term *enter* to refer to the key that generates a carriage return, although the key is named “Return” on some keyboards.

Typographic conventions

The following table describes typographic conventions used in this guide.

Convention	Type of information
Italic font	Words or characters that require special attention. Placeholders for information you must supply. For example, if the guide says to enter the `arp -d` *hostname* command, you enter the characters **`arp -d`** followed by the actual name of the host. Book titles in cross-references.
`Monospaced` font	Command and daemon names. Information displayed on the system console or other computer monitors. The contents of files.
`Bold monospaced` font	Words or characters you type. What you type is always shown in lowercase letters, unless you must type it in uppercase letters.

Special messages

This guide contains special messages that are described as follows:

Note

A note contains important information that helps you install or operate the storage system efficiently.

Attention

An attention contains instructions that you must follow to avoid damage to the equipment, a system crash, or loss of data.

Introduction to NetApp Storage Architecture 1

About this chapter

This chapter provides an overview of how you use Data ONTAP 7.1 software to organize and manage the data storage resources (disks) that are part of a NetApp® storage system and the data that resides on those disks.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding storage architecture" on page 2
- "Understanding the file system and its storage containers" on page 11
- "Using traditional volumes" on page 19

Understanding storage architecture

About storage architecture

Storage architecture refers to how Data ONTAP utilizes storage systems to make data storage resources available to host or client systems and applications. Data ONTAP 7.0 and later versions distinguish between the physical layer of data storage resources and the logical layer that includes the file systems and the data that reside on the physical resources.

The physical layer includes disks, Redundant Array of Independent Disks (RAID) groups they are assigned to, plexes, and aggregates. The logical layer includes volumes, qtrees, Logical Unit Numbers (LUNs), and the files and directories that are stored in them. Data ONTAP also provides Snapshot™ technology to take point-in-time images of volumes and aggregates.

How storage systems use disks

Storage systems use disks from a variety of manufacturers. All new storage systems use block checksum disks (BCDs) for RAID parity checksums. These disks provide better performance for random reads than zoned checksum disks (ZCDs), which were used in older storage systems. For more information about disks, see “Understanding disks” on page 46.

How Data ONTAP uses RAID

Data ONTAP organizes disks into RAID groups, which are collections of data and parity disks to provide parity protection. Data ONTAP 7.1 supports the following RAID types for NetApp storage systems (including the R100 series, the R200 series, the F87, the F800 series, the FAS200 series, the FAS900 series, and the FAS3000 series storage systems).

- RAID4 technology: Before Data ONTAP 6.5, RAID4 was the only RAID protection scheme available for Data ONTAP aggregates. Within its RAID groups, it allots a single disk for holding parity data, which ensures against data loss due to a single disk failure within a group.
- RAID-DP™ technology (DP for double-parity): RAID-DP provides a higher level of RAID protection for Data ONTAP aggregates. Within its RAID groups, it allots one disk for holding parity data and one disk for holding double-parity data. Double-parity protection ensures against data loss due to a double disk failure within a group.

NetApp V-Series systems support systems that use RAID1, RAID5, and RAID10 levels, although the V-Series systems do not themselves use RAID1, RAID5, or RAID10. For information about V-Series systems and how they support RAID types, see the V-Series Systems *Planning Guide*.

Choosing the right size and the protection level for a RAID group depends on the kind of data you intend to store on the disks in that RAID group. For more information about RAID groups, see "Understanding RAID groups" on page 136.

What a plex is

A *plex* is a collection of one or more RAID groups that together provide the storage for one or more WAFL® (Write Anywhere File Layout) file system volumes. Data ONTAP uses plexes as the unit of RAID-level mirroring when the SyncMirror® feature is enabled. All RAID groups in one plex are of the same level, but may have a different number of disks.

What an aggregate is

An *aggregate* is a collection of one or two plexes, depending on whether you want to take advantage of RAID-level mirroring. If the aggregate is unmirrored, it contains a single plex. If the SyncMirror feature is licensed and enabled, Data ONTAP adds a second plex to the aggregate, which serves as a RAID-level mirror for the first plex in the aggregate.

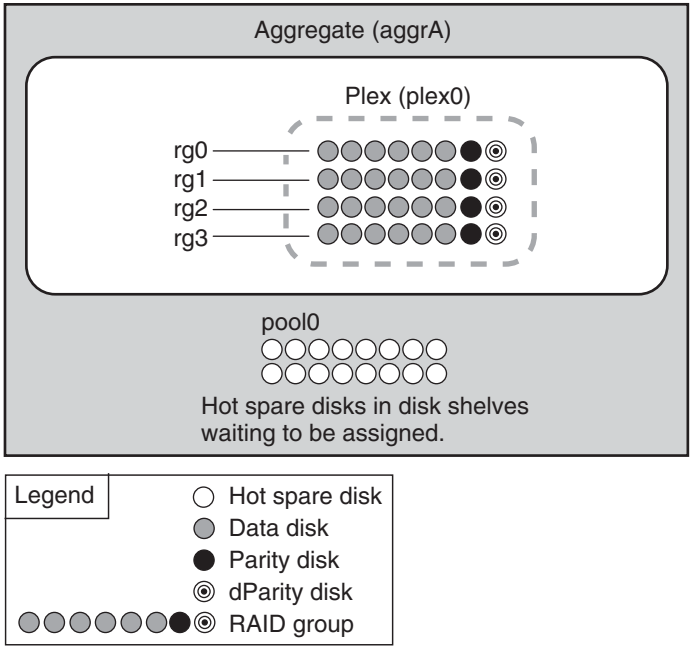
When you create an aggregate, Data ONTAP assigns data disks and parity disks to RAID groups, depending on the options you choose, such as the size of the RAID group (based on the number of disks to be assigned to it) or the level of RAID protection.

You use aggregates to manage plexes and RAID groups because these entities only exist as part of an aggregate. You can increase the usable space in an aggregate by adding disks to existing RAID groups or by adding new RAID groups. Once you've added disks to an aggregate, you cannot remove them to reduce storage space without first destroying the aggregate.

If the SyncMirror feature is licensed and enabled, you can convert an unmirrored aggregate to a mirrored aggregate and vice versa without any downtime.

An unmirrored aggregate: Consists of one plex, automatically named by Data ONTAP as *plex0*. This is the default configuration. In the following diagram, the unmirrored aggregate, arbitrarily named *aggrA* by the user, consists of one plex, which is made up of four double-parity RAID groups, automatically named *rg0*, *rg1*, *rg2,* and *rg3* by Data ONTAP.

Notice that RAID-DP requires that both a parity disk and a double parity disk be in each RAID group. In addition to the disks that have been assigned to RAID groups, there are sixteen hot spare disks in one pool of disks waiting to be assigned.



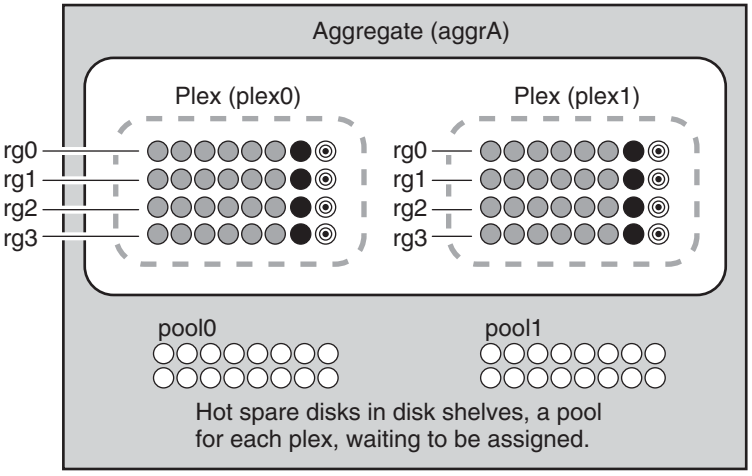
A mirrored aggregate: Consists of two plexes, which provide an even higher level of data redundancy via RAID-level mirroring. For an aggregate to be enabled for mirroring, the storage system's disk configuration must support RAID-level mirroring, and the storage system must have the necessary licenses installed and enabled, as follows.

- A single storage system must have the syncmirror_local license enabled.
- A clustered storage system pair where each node resides within 500 meters of the other must have the cluster and syncmirror_local licenses enabled on both systems.
- A clustered storage system pair where the nodes reside farther apart than 500 meters (known as a MetroCluster) must have the cluster, cluster_remote and syncmirror_local licenses installed. For information about MetroClusters, see the *Cluster Installation and Administration Guide*.

When you enable SyncMirror, Data ONTAP divides all the hot spare disks into two disk pools to ensure a single failure does not affect disks in both pools. This allows the creation of mirrored aggregates. Mirrored aggregates have two plexes. Data ONTAP uses disks from one pool to create the first plex, always named plex0, and another pool to create a second plex, typically named plex1. This provides fault isolation of plexes. A failure that affects one plex will not affect the other plex.

The plexes are physically separated (each plex has its own RAID groups and its own disk pool), and the plexes are updated simultaneously during normal operation. This provides added protection against data loss if there is a double-disk failure or a loss of disk connectivity, because the unaffected plex continues to serve data while you fix the cause of the failure. Once the plex that had a problem is fixed, you can resynchronize the two plexes and reestablish the mirror relationship.

In the following diagram, SyncMirror is enabled, so plex0 has been copied and automatically named *plex1* by Data ONTAP. Notice that plex0 and plex1 contain copies of one or more file systems and that the hot spare disks have been separated into two pools, Pool0 and Pool1.



For more information about aggregates, see “Understanding aggregates” on page 184.

What volumes are

A *volume* is a logical file system whose structure is made visible to users when you export the volume to a UNIX host through an NFS mount or to a Windows host through a CIFS share.

You assign the following attributes to every volume, whether it is a traditional or a FlexVol volume, except where noted:

- The name of the volume
- The size of the volume
- A security style, which determines whether a volume can contain files that use UNIX security, files that use NT file system (NTFS) file security, or both types of files
- Whether the volume uses CIFS oplocks (opportunistic locks)
- The type of language supported
- The level of space guarantees (for FlexVol volumes only)
- Disk space and file limits (quotas)
- A Snapshot copy schedule (optional)

  Data ONTAP automatically creates and deletes Snapshot copies of data in volumes to support commands related to Snapshot technology.

  For more information about the default Snapshot copy schedule, Snapshot copies, plexes, and SyncMirror, see the *Data Protection Online Backup and Recovery Guide*
- Whether the volume is designated as a SnapLock™ volume
- Whether the volume is a *root* volume

  With all new storage systems, Data ONTAP is installed at the factory with a root volume already configured. The root volume is named vol0 by default.

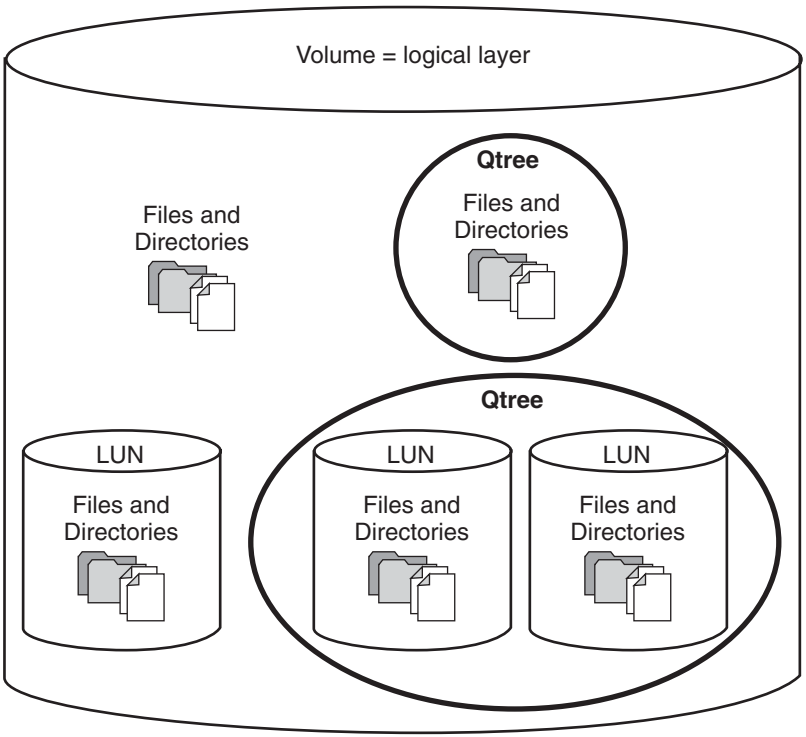
  - If the root volume is a FlexVol volume, its containing aggregate is named aggr0 by default.
  - If the root volume is a traditional volume, its containing aggregate is also named vol0 by default. In Data ONTAP 7.0 and later versions, a traditional volume and its containing aggregate always have the same name.

  The root volume contains the storage system's configuration files, including the /etc/rc file, which includes startup commands and log files. You use the root volume to set up and maintain the configuration files.

  Only one root volume is allowed on an storage system. The root volume contains log files, so for traditional volumes, make sure your root volume spans four to six disks to handle the increased traffic.

A volume is the most inclusive of the logical containers. It can store files and directories, qtrees, and LUNs. You can use qtrees to organize files and directories, as well as LUNs. You can use LUNs to serve as virtual disks in SAN environments to store files and directories. For information about qtrees, see "How qtrees are used" on page 11. For information about LUNs, see "How LUNs are used" on page 11.

The following diagram shows how you can use volumes, qtrees, and LUNs to store files and directories.



For more information about volumes, see Chapter 6, "Volume Management," on page 211.

How aggregates provide storage for volumes

Each volume depends on its containing aggregate for all its physical storage. The way a volume is associated with its containing aggregate depends on whether the volume is a traditional volume or a FlexVol volume.

Traditional volume: A traditional volume is contained by a single, dedicated, aggregate. A traditional volume is tightly coupled with its containing aggregate. The only way to increase the size of a traditional volume is to add entire disks to its containing aggregate. It is impossible to decrease the size of a traditional volume.

The smallest possible traditional volume must occupy all of two disks (for RAID4) or three disks (for RAID-DP). Thus, the minimum size of a traditional volume depends on the size and number of disks used to create the traditional volume.

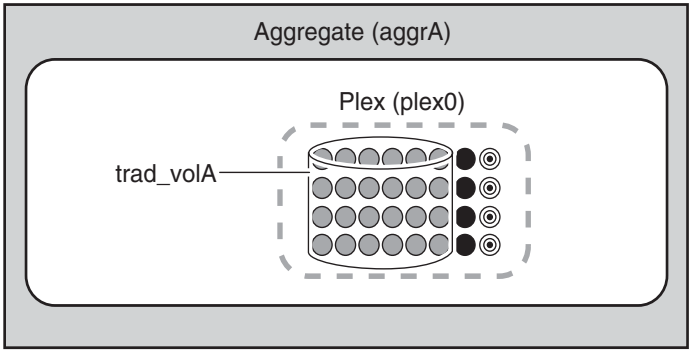
No other volume can use the storage associated with a traditional volume's containing aggregate.

When you create a traditional volume, Data ONTAP creates its underlying containing aggregate based on the parameters you choose with the `vol create` command or with the FilerView® Volume Wizard. Once created, you can manage the traditional volume's containing aggregate with the `aggr` command. You can also use FilerView to perform some management tasks.

The aggregate portion of each traditional volume is assigned its own pool of disks that are used to create its RAID groups, which are then organized into one or two plexes. Because traditional volumes are defined by their own set of disks and RAID groups, they exist outside of and independently of any other aggregates that might be defined on the storage system.

The following diagram illustrates how a traditional volume, trad_volA, is tightly coupled to its containing aggregate. When trad_volA was created, its size was determined by the amount of disk space requested, the number of disks and their capacity to be used, or a list of disks to be used.

A traditional volume with its tightly coupled containing aggregate



FlexVol volume: A FlexVol volume is loosely coupled with its containing aggregate. Because the volume is managed separately from the aggregate, FlexVol volumes give you a lot more options for managing the size of the volume. FlexVol volumes provide the following advantages:

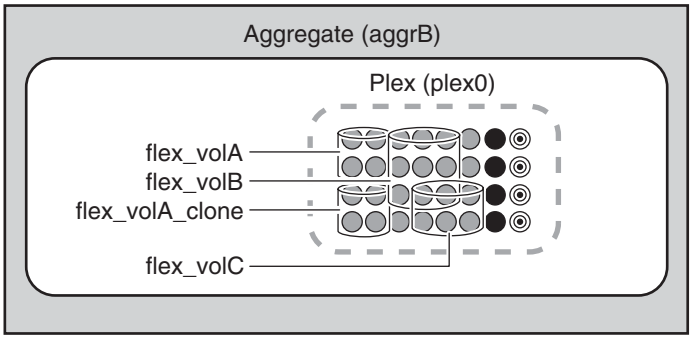
- You can create FlexVol volumes in an aggregate nearly instantaneously. They can be as small as 20 MB and as large as the volume capacity that is supported for your storage system. For information on the maximum raw volume size supported on the storage system, see the *System Configuration Guide* on the NetApp on the Web™ site (NOW) at http://netapp.now.com/.

  These volumes stripe their data across all the disks and RAID groups in their containing aggregate.
- You can increase and *decrease* the size of a FlexVol volume in small increments (as small as 4 KB), nearly instantaneously.
- You can increase the size of a FlexVol volume to be *larger* than its containing aggregate, which is referred to as aggregate overcommitment. For information about this feature, see "Aggregate overcommitment" on page 289.
- You can clone a FlexVol volume, which is then referred to as a FlexClone™ volume. For information about this feature, see "Cloning FlexVol volumes" on page 231.

A FlexVol volume can share its containing aggregate with other FlexVol volumes. Thus, a single aggregate is the shared source of all the storage used by the FlexVol volumes it contains.

In the following diagram, aggrB contains four FlexVol volumes of varying sizes. Note that one of the FlexVol volumes is a FlexClone.

Flexible volumes with their loosely coupled containing aggregate



Traditional volumes and FlexVol volumes can co-exist

You can create traditional volumes and FlexVol volumes on the same storage system, up to the maximum number of volumes allowed. For information about maximum limits, see "Maximum numbers of volumes" on page 26.

What snapshots are

A *snapshot* is a space-efficient, point-in-time image of the data in a volume or an aggregate. Snapshots are used for such purposes as backup and error recovery.

Data ONTAP automatically creates and deletes snapshots of data in volumes to support commands related to Snapshot technology. Data ONTAP also automatically creates snapshots of aggregates to support commands related to the SnapMirror® software, which provides volume-level mirroring. For example, Data ONTAP uses snapshots when data in two plexes of a mirrored aggregate need to be resynchronized.

You can accept the default snapshot schedule, or modify it. You can also create one or more snapshots at any time. For information about aggregate snapshots, see the *System Administration Guide*. For more information about snapshots, plexes, and SyncMirror, see the *Data Protection Online Backup and Recovery Guide*.

Understanding the file system and its storage containers

How volumes are used

A volume holds user data that is accessible via one or more of the access protocols supported by Data ONTAP, including Network File System (NFS), Common Internet File System (CIFS), HyperText Transfer Protocol (HTTP), Web-based Distributed Authoring and Versioning (WebDAV), Fibre Channel Protocol (FCP), and Internet SCSI (iSCSI). A volume can include files (which are the smallest units of data storage that hold user- and system-generated data) and, optionally, directories and qtrees in a Network Attached Storage (NAS) environment, and also LUNs in a Storage Area Network (SAN) environment.

For more information about volumes, see Chapter 6, "Volume Management," on page 211.

How qtrees are used

A *qtree* is a logically-defined file system that exists as a special top-level subdirectory of the root directory within a volume. You use qtrees to manage or partition data within a volume. You can specify the following features for a qtree.

- A security style like that of volumes
- Whether the qtree uses CIFS oplocks
- Whether the qtree has quotas (disk space and file limits)

  Using quotas enables you to manage storage resources on a per user, user group, or per project status. In this way, you can customize areas for projects and keep users and projects from monopolizing resources.

For more information about qtrees, see Chapter 7, "Qtree Management," on page 297.

How LUNs are used

NetApp storage architecture utilizes two types of LUNs:

- In SAN environments, NetApp storage systems are targets that have storage target devices, which are referred to as LUNs. With Data ONTAP, you configure the storage systems by creating traditional volumes to store LUNs or by creating aggregates to contain FlexVol volumes to store LUNs.

  LUNs created on NetApp storage systems and V-Series systems in a SAN environment are used as targets for external storage that is accessible from initiators, or hosts. You use these LUNs to store files and directories accessible through a UNIX or Windows host via FCP or iSCSI.

For more information about LUNs and how to use them, see your *Block Access Management Guide*.

- With the V-Series systems, LUNs provide external storage. They are created on the storage subsystems and are available for storage subsystems, such as IBM, HP, and Hitachi, to read data from or write data to.

  With the V-Series systems, LUNs on the storage subsystem play the role of disks on a NetApp storage system so that the LUNs on the storage subsystem provide the storage instead of the V-Series system. For more information, see the V-Series Systems *Planning Guide*.

How files are used

A *file* is the smallest unit of data management. Data ONTAP and application software create system-generated files, and you or your users create data files. You and your users can also create directories in which to store files. You create volumes in which to store files and directories. You create qtrees to partition your volumes. You manage file properties by managing the volume or qtree in which the file or its directory is stored.

How to use storage resources

The following table describes the storage resources available with Data ONTAP 7.0 and later versions and how you use them.

Storage Container	Description	How to Use
Disk	Advanced Technology Attachment (ATA), Fibre Channel, or SCSI disks are used, depending on the storage system model. Some disk management functions are specific to the storage system, depending on whether the storage system uses a hardware- or software-based disk ownership method.	Once disks are assigned to a storage system, you can choose one of the following methods to assign disks to each RAID group when you create an aggregate: ◆ You provide a list of disks. ◆ You specify a number of disks and let Data ONTAP assign the disks automatically. ◆ You specify the number of disks together with the disk size and/or speed, and let Data ONTAP assign the disks automatically. Disk-level operations are described in Chapter 3, “Disk and Storage Subsystem Management,” on page 45.
RAID group	Data ONTAP supports RAID4 and RAID-DP for all storage systems, and RAID-0 for V-Series systems. The number of disks that each RAID level uses by default is storage system model specific.	The smallest RAID group for RAID4 is two disks (one data and one parity disk); for RAID-DP, it’s three (one data and two parity disks). For information about performance, see “Larger versus smaller RAID groups” on page 142. You manage RAID groups with the `aggr` command and FilerView. (For backward compatibility, you can also use the `vol` command for traditional volumes.) RAID-level operations are described in Chapter 4, “RAID Protection of Data,” on page 135.

Storage Container	Description	How to Use
Plex	Data ONTAP uses plexes to organize file systems for RAID-level mirroring.	You can ◆ Configure and manage SyncMirror backup replication. For more information, see the *Data Protection Online Backup and Recovery Guide.* ◆ Split an aggregate in a SyncMirror relationship into its component plexes ◆ Rejoin split aggregates ◆ Remove and destroy plexes ◆ Change the state of a plex ◆ View the status of plexes
Aggregate	Consists of one or two plexes. A loosely coupled container for one or more FlexVol volumes. A tightly coupled container for exactly one traditional volume.	You use aggregates to manage disks, RAID groups, and plexes. You can create aggregates with the `aggr` command or with the FilerView browser interface. If you create a traditional volume, Data ONTAP automatically creates the underlying aggregate. Aggregate-level operations are described in Chapter 5, “Aggregate Management,” on page 183.

Storage Container	Description	How to Use
Volume (common attributes)	Both traditional and FlexVol volumes contain user-visible directories and files, and they can also contain qtrees and LUNs.	You can apply the following volume operations to both FlexVol volumes and traditional volumes. The operations are also described in "General volume operations" on page 243. ◆ Changing the language option for a volume ◆ Changing the state of a volume ◆ Changing the root volume ◆ Destroying volumes ◆ Exporting a volume using CIFS, NFS, and other protocols ◆ Increasing the maximum number of files in a volume ◆ Renaming volumes The following operations are described in the *Data Protection Online Backup and Recovery Guide*. ◆ Implementing SnapMirror ◆ Taking snapshots of volumes ◆ Implementing the SnapLock™ feature

Storage Container	Description	How to Use
FlexVol volume	A logical file system of user data, metadata, and snapshots that is loosely coupled to its containing aggregate. All FlexVol volumes share the underlying aggregate's disk array, RAID group, and plex configurations. Multiple FlexVol volumes can be contained within the same aggregate, sharing its disks, RAID groups, and plexes. FlexVol volumes can be modified and sized independently of their containing aggregate.	You can create FlexVol volumes after you have created the aggregates to contain them. You can increase and decrease the size of a FlexVol by adding or removing space in increments of 4 KB, and you can clone FlexVol volumes. FlexVol volume-level operations are described in Chapter 6, "FlexVol volume operations," on page 224.
Traditional volume	A logical file system of user data, metadata and snapshots that is tightly coupled to its containing aggregate. Exactly one traditional volume can exist within its containing aggregate, with the two entities becoming indistinguishable and functioning as a single unit. Traditional volumes are identical to volumes created with earlier than 7.0 versions of Data ONTAP. If you upgrade to Data ONTAP 7.0 and later versions, existing volumes are preserved as traditional volumes.	You can create traditional volumes, physically transport them, and increase them by adding disks. For information about creating and transporting traditional volumes, see "Traditional volume operations" on page 215. For information about increasing the size of a traditional volume, see "Adding disks to aggregates" on page 198.

Storage Container	Description	How to Use
Qtree	An optional, logically defined file system that you can create at any time within a volume. It is a subdirectory of the root directory of a volume. You store directories, files, and LUNs in qtrees. You can create up to 4,995 qtrees per volume.	You use qtrees as logical subdirectories to perform file system configuration and maintenance operations. Within a qtree, you can assign limits to the space that can be consumed and the number of files that can be present (through quotas) to users on a per-qtree basis, define security styles, and enable CIFS opportunity locks (oplocks). Qtree-level operations are described in Chapter 7, "Qtree Management," on page 297. Qtree-level operations related to configuring usage quotas are described in Chapter 8, "Quota Management," on page 319.
LUN (in a SAN environment)	Logical Unit Number; it is a logical unit of storage, which is identified by a number by the initiator accessing its data in a SAN environment. A LUN is a file that appears as a disk drive to the initiator.	You create LUNs within volumes and specify their sizes. For more information about LUNs, see your *Block Access Management Guide*.

Storage Container	Description	How to Use
LUN (with V-Series systems)	An area on the storage subsystem that is available for a V-Series system or non-V-Series system host to read data from or write data to. The V-Series system can virtualize the storage attached to it and serve the storage up as LUNs to customers outside the V-Series system (for example, through iSCSI). These LUNs are referred to as V-Series system-served LUNs. The clients are unaware of where such a LUN is stored.	See the V-Series *Software Setup, Installation, and Management Guide* for your storage subsystem for specific information about LUNs and how to use them for your system.
File	Files contain system-generated or user-created data. Files are the smallest unit of data management. Users organize files into directories. As a system administrator, you organize directories into volumes.	Configuring file space reservation is described in Chapter 6, "Volume Management," on page 211.

Using traditional volumes

Upgrading to Data ONTAP 7.0 or later

If you are upgrading to Data ONTAP 7.0 or later software from an earlier version, your existing volumes are preserved as traditional volumes. Your volumes and data remain unchanged, and the commands you used to manage your volumes and data are still supported for backward compatibility.

As you learn more about FlexVol volumes, you might want to migrate your data from traditional volumes to FlexVol volumes. For information about migrating traditional volumes to FlexVol volumes, see "Migrating between traditional volumes and FlexVol volumes" on page 244.

About traditional volumes

With traditional volumes, you can use the `aggr` and `aggr options` commands or FilerView to manage the volume's containing aggregate. For backward compatibility, you can also use the `vol` and the `vol options` commands to manage the traditional volume's containing aggregate.

The following table describes how to create and manage traditional volumes using either the `aggr` or the `vol` commands, and FilerView, depending on whether you are managing the physical or logical layers of that volume.

Traditional volume task, using FilerView, if available	**Using the `aggr` command**	**Using the `vol` command**
Create a volume In FilerView: Volumes > Add	`aggr create` *`trad_vol`* `-v -m {`*`disk-list`* `|` *`size`*`}` Creates a traditional volume and defines a set of disks to include in that volume or defines the size of the volume. The `-v` option designates that `trad_vol` is a traditional volume. Use `-m` to enable SyncMirror.	For backward compatibility: `vol create trad_vol -m {disk-list|size}`

Traditional volume task, using FilerView, if available	Using the `aggr` command	Using the `vol` command
Add disks In FilerView: Volumes > Manage. Click the *trad_vol* name you want to add disks to. The Volume Properties page appears. Click Add disks. The Volume Wizard appears.	`aggr add` *`trad_vol disks`*	For backward compatibility: `vol add` *`trad_vol disks`*
Create a SyncMirror replica In FilerView: For new aggregates: Aggregates > Add For existing aggregates: Aggregates > Manage. Click *trad_vol*. The Aggregate properties page appears. Click Mirror Click OK	`aggr mirror`	For backward compatibility: `vol mirror`
Set the root volume option This option can be used on only one volume per storage system. For more information on root volumes, see “How volumes are used” on page 11.	Not applicable.	`vol options` *`trad_vol`* `root` If the root option is set on a traditional volume, that volume becomes the root volume for the storage system on the next reboot.

Traditional volume task, using FilerView, if available	Using the `aggr` command	Using the `vol` command
Set RAID level (`raidtype`) options In FilerView: For new aggregates: Aggregates > Add For existing aggregates: Aggregates > Manage. Click *trad_vol*. Click Modify	`aggr options` *trad_vol* `{ raidsize` *number* `| raidtype` *level*`}`	For backward compatibility: `vol options` *trad_vol* `{ raidsize` *number* `| raidtype` *level*`}`
Set up SnapLock volume	`aggr create` *trad_vol* `-v -r -L` *disk-list*	For backward compatibility: `vol create trad_vol -r -L` *disk-list*
Split a SyncMirror relationship	`aggr split`	For backward compatibility: `vol split`

RAID level scrub
 	
In FilerView:
 	
Aggregates > Configure RAID
	`aggr scrub start`
	
	`aggr scrub suspend`
	
	`aggr scrub stop`
	
	`aggr scrub resume`
	
	`aggr scrub status`
	
	Manages RAID-level error scrubbing of the disks.
	
	See “Automatic and manual disk scrubs” on page 166.
		For backward compatibility:
		
		`vol scrub start`
		
		`vol scrub suspend`
		
		`vol scrub stop`
		
		`vol scrub resume`
		
		`vol scrub status`

Media level scrub	`aggr media_scrub` *t_vol* Manages media error scrubbing of disks in the traditional volume. See “Continuous media scrub” on page 175.	For backward compatibility: `vol media_scrub` *t_vol*

Traditional volume task, using FilerView, if available	Using the `aggr` command	Using the `vol` command
Verify that two SyncMirror plexes are identical	`aggr verify`	For backward compatibility: `vol verify`

Quick setup for aggregates and volumes 2

About this chapter

This chapter provides the information you need to plan and create aggregates and volumes.

After initial setup of your storage system's disk groups and file systems, you can manage or modify them using information in other chapters.

Topics in this chapter

This chapter discusses the following topics:

- "Planning your aggregate, volume, and qtree setup" on page 24
- "Configuring data storage" on page 29
- "Converting from one type of volume to another" on page 35
- "Overview of aggregate and volume operations" on page 36

Planning your aggregate, volume, and qtree setup

Planning considerations

How you plan to create your aggregates and FlexVol volumes, traditional volumes, qtrees, or LUNs depends on your requirements and whether your new version of Data ONTAP is a new installation or an upgrade from Data ONTAP 6.5.*x* or earlier. For information about upgrading a NetApp storage system, see the *Data ONTAP 7.1 Upgrade Guide*.

Considerations when planning aggregates

For new storage systems: If you purchased a new storage system with Data ONTAP 7.0 or later installed, the root FlexVol volume (vol0) and its containing aggregate (aggr0) are already configured.

The remaining disks on the storage system are all unallocated. You can create any combination of aggregates with FlexVol volumes, traditional volumes, qtrees, and LUNs, according to your needs.

Maximizing storage: To maximize the storage capacity of your storage system per volume, configure large aggregates containing multiple FlexVol volumes. Because multiple FlexVol volumes within the same aggregate share the same RAID parity disk resources, more of your disks are available for data storage.

SyncMirror replication: You can set up a RAID-level mirrored aggregate to contain volumes whose users require guaranteed SyncMirror data protection and access. SyncMirror replicates the volumes in plex0 to plex1. The disks used to store the second plex can be up to 30 km away if you use MetroCluster. If you set up SyncMirror replication, plan to allocate double the number of disks that you would otherwise need for the aggregate to support your users. For information about MetroClusters, see the *Cluster Installation and Administration Guide*.

All volumes contained in a mirrored aggregate are in a SyncMirror relationship, and all new volumes created within the mirrored aggregate inherit this feature. For more information on configuring and managing SyncMirror replication, see the *Data ONTAP Online Backup and Recovery Guide*.

If you set up SyncMirror replication, plan to allocate double the disks that you would otherwise need for the aggregate to support your users.

Size of RAID groups: When you create an aggregate, you can control the size of a RAID group. Generally, larger RAID groups maximize your data storage space by providing a greater ratio of data disks to parity disks. For information on RAID group size guidelines, see "Larger versus smaller RAID groups" on page 142.

Levels of RAID protection: Data ONTAP supports two levels of RAID protection, which you can assign on a per-aggregate basis: RAID4 and RAID-DP.

For more information on RAID4 and RAID-DP, see "Levels of RAID protection" on page 136.

Considerations when planning volumes

Root volume sharing: When technicians install Data ONTAP on your storage system, they create a root volume named *vol0*. The root volume is a FlexVol volume, so you can resize it. For information about the minimum size for a FlexVol root volume, see the section on root volume size in the *System Administration Guide*. For information about resizing FlexVol volumes, see "Resizing FlexVol volumes" on page 229.

Sharing storage: To share the storage capacity of your disks using the SharedStorage™ feature, you must decide whether you want to use the `vFiler no-copy migration` functionality. If so, you must configure your storage using traditional volumes. If you also want to take advantage of the migration software feature using SnapMover to reassign disks from a CPU-bound storage system to an underutilized storage system, you must have licenses for the MultiStore® and SnapMover® features. For more information, see "SharedStorage" on page 77.

SnapLock volume: The SnapLock feature enables you to keep a permanent snapshot by writing new data once to disks and then preventing the removal or modification of that data. You can create and configure a special traditional volume to provide this type of access, or you can create an aggregate to contain FlexVol volumes that provide this type of access. If an aggregate is enabled for SnapLock, all of the FlexVol volumes that it contains have mandatory SnapLock protection. For more information, see the *Data Protection Online Recovery and Backup Guide*.

Data sanitization: Disk sanitization is a Data ONTAP feature that enables you to erase sensitive data from storage system disks beyond practical means of physical recovery. Because data sanitization is carried out on the entire set of disks in an aggregate, configuring smaller aggregates to hold sensitive data that requires sanitization minimizes the time and disruption that sanitization

operations entail.You can create smaller aggregates and traditional volumes whose data you might have reason to sanitize at periodic intervals. For more information, see "Sanitizing disks" on page 105.

Maximum numbers of aggregates: You can create up to 100 aggregates per storage system, regardless of whether the aggregates contain FlexVol volumes or traditional volumes.

You can use the `aggr status` command or FilerView (by viewing the System Status window) to see how many aggregates exist. With this information, you can determine how many more aggregates you can create on the storage system, depending on available capacity. For more information about FilerView, see the *System Administration Guide*.

Maximum numbers of volumes: You can create up to 200 volumes per storage system. However, you can only create up to 100 traditional volumes because of the 100 aggregates per storage system limit. You can use the `vol status` command or FilerView (Volumes > Manage > Filter by) to see how many volumes exist, and whether they are FlexVol volumes or traditional volumes. With this information, you can determine how many more volumes you can create on that storage system, depending on available capacity.

Consider the following example. Assume you create:

- Ten traditional volumes. Each has exactly one containing aggregate.
- Twenty aggregates, and you then create four FlexVol volumes in each aggregate, for a total of eighty FlexVol volumes.

You now have a total of:

- Thirty aggregates (ten from the traditional volumes, plus the twenty created to hold the FlexVol volumes)
- Ninety volumes (ten traditional and eighty FlexVol) on the storage system

Thus, the storage system is well under the maximum limits for either aggregates or volumes.

If you have a combination of FlexVol volumes and traditional volumes, the 100-maximum limit of aggregates still applies. If you need more than 200 user-visible file systems, you can create qtrees within the volumes.

Considerations for FlexVol volumes

When planning the setup of your FlexVol volumes within an aggregate, consider the following issues.

General Deployment: FlexVol volumes have different best practices, optimal configurations, and performance characteristics compared to traditional volumes. Make sure you understand these differences and deploy the configuration that is optimal for your environment.

For information about deploying a storage solution with FlexVol volumes, including migration and performance considerations, see the technical report *Introduction to Data ONTAP Release 7G* (available from the NetApp Library at http://www.netapp.com/tech_library/ftp/3356.pdf).

FlexVol space guarantee: Setting a maximum volume size does not guarantee that the volume will have that space available if the aggregate space is oversubscribed. As you plan the size of your aggregate and the maximum size of your FlexVol volumes, you can choose to overcommit space if you are sure that the actual storage space used by your volumes will never exceed the physical data storage capacity that you have configured for your aggregate. This is called aggregate overcommitment. For more information, see "Aggregate overcommitment" on page 289.

Volume language: During volume creation you can specify the language character set to be used.

Backup: You can size your FlexVol volumes for convenient volume-wide data backup through SnapMirror, SnapVault®, and Volume Copy features. For more information, see the *Data ONTAP Online Backup and Recovery Guide*.

Volume cloning: Many database programs enable data cloning, that is, the efficient copying of data for the purpose of manipulation and projection operations. This is efficient because Data ONTAP allows you to create a duplicate of a volume by having the original volume and clone volume share the same disk space for storing unchanged data. For more information, see "Cloning FlexVol volumes" on page 231.

Considerations for traditional volumes

Upgrading: If you upgrade to Data ONTAP 7.0 or later from a previous version, the upgrade program preserves all of your existing volumes as traditional volumes.

Disk portability: You can create traditional volumes and aggregates whose disks you intend to physically transport from one storage system to another. This ensures that a specified set of physically transported disks will hold all the data associated with a specified volume and only the data associated with that volume. For more information, see "Physically transporting traditional volumes" on page 221.

Considerations when planning qtrees

Within a volume you have the option of creating qtrees to provide another level of logical file systems. This is especially useful if you are using traditional volumes. Some reasons to consider setting up qtrees include:

Increased granularity: Up to 4,995 qtrees—that is 4,995 virtually independent file systems—are supported per volume. For more information see Chapter 7, "Qtree Management," on page 297.

Sophisticated file and space quotas for users: Qtrees support a sophisticated file and space quota system that you can use to apply soft or hard space usage limits on individual users, or groups of users. For more information see Chapter 8, "Quota Management," on page 319.

Configuring data storage

About configuring data storage

You configure data storage by creating aggregates and FlexVol volumes, traditional volumes, and LUNs for a SAN environment. You can also use qtrees to partition data in a volume.

You can create up to 100 aggregates per storage system. Minimum aggregate size is two disks (one data disk, one parity disk) for RAID4 or three disks (one data, one parity, and one double parity disk) for RAID-DP. However, you are advised to configure the size of your RAID groups according to the anticipated load. For more information, see the chapter on system information and performance in the *System Administration Guide*.

Creating aggregates, FlexVol volumes, and qtrees

To create an aggregate and a FlexVol, complete the following steps.

Step	Action
1	(Optional) Determine the free disk resources on your storage system by entering the following command: **`aggr status -s`** `-s` displays a listing of the spare disks on the storage system. **Result:** Data ONTAP displays a list of the disks that are not allocated to an aggregate. With a new storage system, all disks except those allocated for the root volume's aggregate (explicit for a FlexVol and internal for a traditional volume) will be listed.

Step	Action
2	(Optional) Determine the size of the aggregate, assuming it is aggr0, by entering one of the following commands: For size in kilobytes, enter: **`df -A aggr0`** For size in 4096-byte blocks, enter: **`aggr status -b aggr0`** For size in number of disks, enter: **`aggr status { -d | -r } aggr0`** `-d` displays disk information `-r` displays RAID information **Note** If you want to expand the size of the aggregate, see "Adding disks to an aggregate" on page 199.

Step	Action
3	Create an aggregate by entering the following command: **`aggr create [-m] [-r raidsize] aggr ndisks[@disksize]`** **Example:** `aggr create aggr1 24@72G` **Result:** An aggregate named *aggr1* is created. It consists of 24 72-GB disks. `-m` instructs Data ONTAP to implement SyncMirror. `-r raidsize` specifies the maximum number of disks of each RAID group in the aggregate. The maximum and default values for *raidsize* are platform-dependent, based on performance and reliability considerations. By default, the RAID level is set to RAID-DP. If *raidsize* is sixteen (16), aggr1 consists of two RAID groups, the first group having fourteen (14) data disks, one (1) parity disk, and one (1) double parity disk, and the second group having six (6) data disks, one (1) parity disk, and one (1) double parity disk. If *raidsize* is eight (8), aggr1 consists of three RAID groups, each one having six (6) data disks, one (1) parity disk, and one (1)double parity disk.
4	(Optional) Verify the creation of this aggregate by entering the following command: **`aggr status aggr1`**

Step	Action
5	Create a FlexVol volume in the specified aggregate by entering the following command: **`vol create vol aggr size`** **Example:** `vol create new_vol aggr1 32g` **Result:** The FlexVol volume new_vol, with a maximum size of 32 GB, is created in the aggregate, aggr1. The default space guarantee setting for FlexVol volume creation is `volume`. The `vol create` command fails if Data ONTAP cannot guarantee 32 GB of space. To override the default, enter one of the following commands. For information about space guarantees, see "Space guarantees" on page 286. **`vol create vol -s none aggr size`** or **`vol create vol -s file aggr size`**
6	(Optional) To verify the creation of the FlexVol named *new_vol*, enter the following command: **`vol status new_vol -v`**
7	If you want to create additional FlexVol volumes in the same aggregate, use the `vol create` command as described in Step 5. Note the following constraints: ◆ Volumes must be uniquely named across all aggregates within the same storage system. If aggregate *aggr1* contains a volume named *volA*, no other aggregate on the storage system can contain a volume with the name *volA*. ◆ You can create a maximum of 200 FlexVol volumes in one storage system. ◆ Minimum size of a FlexVol volume is 20 MB.

Step	Action
8	To create qtrees within your volumes, enter the following command: **qtree create /vol/*vol*/*qtree*** **Example:** `qtree create /vol/new_vol/my_tree` **Result:** The qtree *my_tree* is created within the volume named new_vol. **Note** You can create up to 4,995 qtrees within one volume.
9	(Optional) To verify the creation of the qtree named my_tree, within the volume named new_vol, enter the following command: **qtree status new_vol -v**

Why continue using traditional volumes

If you upgrade to Data ONTAP 7.0 or later from a previous version of Data ONTAP, the upgrade program keeps your traditional volumes intact. You might want to maintain your traditional volumes and create additional traditional volumes because some operations are more practical on traditional volumes or require traditional volumes, such as:

- Performing disk sanitization operations
- Physically transferring volume data from one location to another (which is most easily carried out on small-sized traditional volumes)
- Migrating volumes using the SnapMover® feature
- Using the SharedStorage feature

Creating traditional volumes and qtrees

To create a traditional volume, complete the following steps:

Step	Action
1	(Optional) List the aggregates and traditional volumes on your storage system by entering the following command: **aggr status -v**

Step	Action
2	(Optional) Determine the free disk resources on your storage system by entering the following command: **aggr status -s**
3	Create a traditional volume by entering the following command: **aggr create *trad_vol* -v n*disks*[@*disksize*]** **Example:** `aggr create new_tvol -v 16@72g`
4	(Optional) Verify the creation of the traditional volume named new_tvol by entering the following command: **vol status new_tvol -v**
5	If you want to create additional traditional volumes, use the `aggr create` command as described in Step 3. Note the following constraints: ◆ All volumes, including traditional volumes, must be uniquely named within the same storage system. ◆ You can create a maximum of 100 traditional volumes within one storage system. ◆ Minimum traditional volume size depends on the disk capacity and RAID protection level.
6	Create qtrees within your volume by entering the following command: **qtree create /vol/*vol*/*qtree*** **Example:** `qtree create /vol/new_tvol/users_tree` **Result:** The qtree *users_tree* is created within the new_tvol volume. **Note** You can create up to 4,995 qtrees within one volume.
7	(Optional) Verify the creation of the qtree named users_tree within the new_tvol volume by entering the following command: **qtree status new_tvol -v**

Converting from one type of volume to another

What converting to another volume type involves

Converting from one type of volume to another is not a single-step procedure. It involves creating a new volume, migrating data from the old volume to the new volume, and verifying that the data migration was successful. You can migrate data from traditional volumes to FlexVol volumes or vice versa. For more information about migrating data, see "Migrating between traditional volumes and FlexVol volumes" on page 244.

When to convert from one type of volume to another

You might want to convert a traditional volume to a FlexVol volume because

- You upgraded an existing NetApp storage system that is running an earlier release than Data ONTAP 7.0 or later and you want to convert the traditional root volume to a FlexVol volume to reduce the amount of disks used to store the system directories and files.
- You purchased a new storage system but initially created traditional volumes and now you want to
  - Take advantage of FlexVol volumes
  - Take advantage of other advanced features, such as FlexClone volumes
  - Reduce lost capacity due to the number of parity disks associated with traditional volumes
  - Realize performance improvements by being able to increase the number of disks the data in a FlexVol volume is striped across

You might want to convert a FlexVol volume to a traditional volume because

- You want to revert to an earlier release of Data ONTAP.

Depending on the number and size of traditional volumes on your storage systems, this might require a significant amount of planning, resources, and time.

NetApp offers assistance

NetApp Professional Services staff, including Professional Services Engineers (PSEs) and Professional Services Consultants (PSCs) are trained to assist customers with converting volume types and migrating data, among other services. For more information, contact your local NetApp Sales representative, PSE, or PSC.

Overview of aggregate and volume operations

About aggregate and volume-level operations

The following table provides an overview of the operations you can carry out on an aggregate, a FlexVol volume, and a traditional volume.

Operation	Aggregate	FlexVol	Traditional volume
Adding disks to an aggregate	`aggr add` *`aggr disks`* Adds disks to the specified aggregate. See "Adding disks to aggregates" on page 198.	Not applicable.	`aggr add` *`trad_vol disks`* Adds disks to the specified traditional volume. See "Adding disks to aggregates" on page 198.
Changing the size of an aggregate	See "Displaying the number of hot spare disks with the Data ONTAP CLI" on page 95 and "Adding disks to aggregates" on page 198.	Not applicable.	See "Displaying the number of hot spare disks with the Data ONTAP CLI" on page 95 and "Adding disks to aggregates" on page 198.
Changing the size of a volume	Not applicable	`vol size` *`flex_vol newsize`* Modifies the size of the specified FlexVol volume. See "Resizing FlexVol volumes" on page 229.	To increase the size of a traditional volume, add disks to its containing aggregate. See "Changing the size of an aggregate" on page 36. You cannot decrease the size of a traditional volume.

Operation	Aggregate	FlexVol	Traditional volume

Changing states: online, offline, restricted	`aggr offline aggr`
	`aggr online aggr`
	`aggr restrict aggr`
	Takes the specified aggregate offline, brings it back online, or puts it in a restricted state.
	See "Changing the state of an aggregate" on page 193.
		`vol offline vol`
		`vol online vol`
		`vol restrict vol`
		Takes the specified volume offline, brings it back online (if its containing aggregate is also online), or puts it in a restricted state.
		See "Determining volume status and state" on page 256.
			`aggr offline vol`
			`aggr online vol`
			`aggr restrict vol`
			Takes the specified volume offline, brings it back online, or puts it in a restricted state.
			See "Determining volume status and state" on page 256.

Copying	`aggr copy start src_aggr dest_aggr` Copies the specified aggregate and its FlexVol volumes to a different aggregate on a new set of disks. See the *Data Protection Online Backup and Recovery Guide*.	`vol copy start src_vol dest_vol` Copies the specified source volume and its data content to a destination volume on a new set of disks. The source and destination volumes must be of the same type (either a FlexVol volume or a traditional volume). See the *Data Protection Online Backup and Recovery Guide*.	

Operation	Aggregate	FlexVol	Traditional volume
Creating an aggregate	`aggr create aggr [-f] [-m] [-n] [-t raidtype] [-r raidsize] [-T disk-type] [-R rpm] [-L] {ndisks[@size] | -d disk1 [disk2 ...] [-d diskn [diskn+1 ... ]]}` Creates a physical aggregate of disks, within which FlexVol volumes can be created. See "Creating aggregates" on page 187.	Not applicable.	See creating a volume.
Creating a volume	Not applicable.	`vol create flex_vol [-l language_code] [-s none | file | volume] aggr size` Creates a FlexVol volume within the specified containing aggregate. See "Creating FlexVol volumes" on page 225.	`aggr create trad_vol -v [-l language_code] [-f] [-n] [-m] [-L] [-t raidtype] [-r raidsize] [-R rpm] {ndisks@size] | -d disk1 [disk2 ...] [-d diskn [diskn+1 ... ]]}` Creates a traditional volume and defines a set of disks to include in that volume. See "Creating traditional volumes" on page 216.

Operation	Aggregate	FlexVol	Traditional volume
Creating a FlexClone	Not applicable.	`vol clone create` *`flex_vol clone_vol`* Creates a clone of the specified FlexVol volume. See "Cloning FlexVol volumes" on page 231.	Not applicable.
Creating a SnapLock volume	`aggr create` *`aggr`* `-L` *`disk-list`* See the *Data Protection Online Recovery and Backup Guide*.	FlexVol volumes inherit the SnapLock attribute from their containing aggregate. See the *Data Protection Online Recovery and Backup Guide*.	`aggr create` *`trad_vol`* `-v -L` *`disk-list`* See the *Data Protection Online Recovery and Backup Guide*.
Creating a SyncMirror replica	`aggr mirror` Creates a SyncMirror replica of the specified aggregate. See the *Data Protection Online Backup and Recovery Guide*.	Not applicable.	`aggr mirror` Creates a SyncMirror replica of the specified traditional volume. See the *Data Protection Online Backup and Recovery Guide*.
Destroying aggregates and volumes	`aggr destroy` *`aggr`* Destroys the specified aggregate and returns that aggregate's disks to the storage system's pool of hot spare disks. See "Destroying aggregates" on page 204.	`vol destroy` *`flex_vol`* Destroys the specified FlexVol volume and returns space to its containing aggregate. See "Destroying volumes" on page 263.	`aggr destroy` *`trad_vol`* Destroys the specified traditional volume and returns that volume's disks to the storage system's pool of hot spare disks See "Destroying volumes" on page 263.

Operation	Aggregate	FlexVol	Traditional volume
Displaying the containing aggregate	Not applicable.	`vol container flex_vol` Displays the containing aggregate of the specified FlexVol volume. See "Displaying a FlexVol volume's containing aggregate" on page 242.	Not applicable.
Displaying the language code	Not applicable	`vol lang [vol]` Displays the volume's language. See "Changing the language for a volume" on page 255.	
Displaying a media-level scrub	`aggr media_scrub status [aggr]` Displays media error scrubbing of disks in the aggregate. See "Continuous media scrub" on page 175	Not applicable.	`aggr media_scrub status [aggr]` Displays media error scrubbing of disks in the traditional volume. See "Continuous media scrub" on page 175.
Displaying the status	`aggr status [aggr]` Displays the offline, restricted, or online status of the specified aggregate. Online status is further defined by RAID state, reconstruction, or mirroring conditions. See "Changing the state of an aggregate" on page 193.	`vol status [vol]` Displays the offline, restricted, or online status of the specified volume, and the RAID state of its containing aggregate. See "Determining volume status and state" on page 256.	`aggr status [vol]` Displays the offline, restricted, or online status of the specified volume. Online status is further defined by RAID state, reconstruction, or mirroring conditions. See "Determining volume status and state" on page 256.

Operation	Aggregate	FlexVol	Traditional volume

Performing a RAID-level scrub	`aggr scrub start`
	`aggr scrub suspend`
	`aggr scrub stop`
	`aggr scrub resume`
	`aggr scrub status`
	Manages RAID-level error scrubbing of disks of the aggregate.
	See "Automatic and manual disk scrubs" on page 166.
		Not applicable.	`aggr scrub start`
			`aggr scrub suspend`
			`aggr scrub stop`
			`aggr scrub resume`
			`aggr scrub status`
			Manages RAID-level error scrubbing of disks of the traditional volume.
			See "Automatic and manual disk scrubs" on page 166

Renaming aggregates and volumes	`aggr rename` *`old_name new_name`* Renames the specified aggregate as *new_name*. See "Renaming an aggregate" on page 197.	`vol rename` *`old_name new_name`* Renames the specified flexible volume as *new_name.* See "Renaming volumes" on page 262.	`aggr rename` *`old_name new_name`* Renames the specified traditional volume as *new_name.* See "Renaming volumes" on page 262.
Setting the language code	Not applicable	`vol lang` *`vol language_code`* Sets the volume's language. See "Changing the language for a volume" on page 255.	
Setting the maximum directory size	Not applicable.	`vol option` *`vol`* `maxdirsize` *`size`* *size* specifies the maximum directory size allowed in the specified volume. See "Increasing the maximum number of files in a volume" on page 265.	

Operation	Aggregate	FlexVol	Traditional volume
Setting the RAID options	`aggr options aggr {raidsize number | raidtype level}` Modifies RAID settings on the specified aggregate. See "Setting RAID level and group size" on page 149 or "Changing the RAID level for an aggregate" on page 152.	Not applicable.	`aggr options trad_vol {raidsize number | raidtype level}` Modifies RAID settings on the specified traditional volume. See "Setting RAID level and group size" on page 149 or "Changing the RAID level for an aggregate" on page 152.
Setting the root volume	Not applicable.	`vol options flex_vol root`	`vol options trad_vol root`
Setting the UNICODE options	Not applicable.	`vol options vol {convert_ucode | create_ucode} {on|off}` Forces or specifies as default conversion to UNICODE format on the specified volume. For information about UNICODE, see the *System Administration Guide*.	
Splitting a SyncMirror relationship	`aggr split` Splits the relationship between two replicas in a SyncMirror relationship. See the *Data Protection Online Backup and Recovery Guide*.	Not applicable.	`aggr split` Splits the relationship between two replicas in a SyncMirror relationship. See the *Data Protection Online Backup and Recovery Guide*.
Verifying two SyncMirror replicas are identical	`aggr verify` Verifies that two replicas are identical. See the *Data Protection Online Backup and Recovery Guide*.	Not applicable.	`aggr verify` Verifies that two replicas are identical. See the *Data Protection Online Backup and Recovery Guide*.

Configuring volume-level options

The following table provides an overview of the options you can use to configure your aggregates, FlexVol volumes and traditional volumes.

Note

The `option` subcommands you execute remain in effect after the storage system is rebooted, so you do not have to add `aggr options` or `vol options` commands to the /etc/rc file.

Aggregate	FlexVol	Traditional volume
`aggr options` *`aggr`* [*`optname optvalue`*] Displays the option settings of *aggr*, or sets *optname* to *optvalue*. See the na_aggr man page.	`vol options` *`vol`* [*`optname optvalue`*] Displays the option settings of *vol*, or sets *optname* to *optvalue*. See the na_vol man page.	
	`convert_ucode on | off`	`convert_ucode on | off`
	`create_ucode on | off`	`create_ucode on | off`
	`fractional_reserve` *`percent`*	`fractional_reserve` *`percent`*
`fs_size-fixed on | off`	`fs_size-fixed on | off`	`fs_size-fixed on | off`
	`guarantee file | volume | none`	
`ignore_inconsistent on | off`		`ignore_inconsistent on | off`
`lost_write_protect`		
	`maxdirsize` *`number`*	`maxdirsize` *`number`*
	`minra on | off`	`minra on | off`
	`no_atime_update on | off`	`no_atime_update on | off`
`nosnap on | off`	`nosnap on | off`	`nosnap on | off`
	`nosnapdir on | off`	`nosnapdir on | off`
	`nvfail on | off`	`nvfail on | off`
`raidsize` *`number`*		`raidsize` *`number`*

Aggregate	FlexVol	Traditional volume
`raidtype raid4 | raid_dp | raid0`		`raidtype raid4 | raid_dp | raid0`
`resyncsnaptime` *number*		`resyncsnaptime` *number*
`root`	`root`	`root`
`snaplock_compliance` (read only)	`snaplock_compliance` (read only)	`snaplock_compliance` (read only)
	`snaplock_default_ period` (read only)	`snaplock_default_ period` (read only)
`snaplock_enterprise` (read only)	`snaplock_enterprise` (read only)	`snaplock_enterprise` (read only)
	`snaplock_minimum_ period`	`snaplock_minimum_ period`
	`snaplock_maximum_ period`	`snaplock_maximum_ period`
`snapmirrored off`	`snapmirrored off`	`snapmirrored off`
`snapshot_autodelete on | off`		
	`svo_allow_rman on | off`	`svo_allow_rman on | off`
	`svo_checksum on | off`	`svo_checksum on | off`
	`svo_enable on | off`	`svo_enable on | off`
	`svo_reject_errors`	`svo_reject_errors`

Disk and Storage Subsystem Management 3

About this chapter

This chapter discusses disk characteristics, how disks are configured, how they are assigned to NetApp storage systems, and how they are managed. This chapter also discusses how to manage other storage subsystem components connected to your storage system, including the host adapters, hubs, tape devices, and medium changer devices.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding disks" on page 46
- "Disk configuration and ownership" on page 53
- "Disk access methods" on page 68
- "Disk management" on page 85
- "Disk performance and health" on page 117
- "Storage subsystem management" on page 122

Understanding disks

About disks

Disks have several characteristics, which are either attributes determined by the manufacturer or attributes that are supported by Data ONTAP. Data ONTAP manages disks based on the following characteristics:

- Disk type (See "Disk type" on page 46)
- Disk capacity (See "Disk capacity" on page 48)
- Disk speed (See "Disk speed" on page 49)
- Disk checksum format (See "Disk checksum format" on page 49)
- Disk addressing (See "Disk addressing" on page 50)
- RAID group disk type (See "RAID group disk type" on page 52)

Disk type

Data ONTAP supports the following disk types, depending on the specific storage system, the disk shelves, and the I/O modules installed in the system:

- FC-AL—for F800, FAS200, FAS900, and FAS3000 series storage systems
- ATA (Parallel ATA)—for the NearStore storage systems (R100 series and R200) and for fabric-attached storage (FAS) storage systems that support the DS14mk2 AT disk shelf and the AT-FC or AT-FCX I/O module
- SCSI—for the F87 storage system

The following table shows what disk type is supported by which storage system, depending on the disk shelf and I/O module installed.

NetApp Storage System	Disk Shelf Supported	I/O Module	Disk Type
F87	Internal disk shelf	Not applicable.	SCSI
F800 series	Fibre Channel StorageShelf FC7, FC8, FC9	Not applicable.	FC
	DS14 DS14mk2 FC	LRC, ESH, ESH2	
FAS250	DS14mk2 FC (not expandable)	Not applicable.	FC

NetApp Storage System	Disk Shelf Supported	I/O Module	Disk Type
FAS270	DS14mk2 FC	LRC, ESH2	FC
FAS920 FAS940	Fibre Channel StorageShelf FC7, FC8, FC9	Not applicable.	FC
	DS14, DS14mk2 FC	LRC, ESH, ESH2	FC
FAS960	Fibre Channel StorageShelf FC7, FC8, FC9	Not applicable.	FC
	DS14, DS14mk2 FC	LRC, ESH, ESH2	FC
	DS14mk2 AT	AT-FCX	ATA
FAS980	Fibre Channel StorageShelf FC9	Not applicable.	FC
	DS14, DS14mk2 FC	LRC, ESH, ESH2	FC
FAS3020 FAS3050 FAS3070	DS14, DS14mk2 FC	LRC, ESH, ESH2	FC
	DS14mk2 AT	AT-FCX	ATA
R100	R1XX disk shelf	Not applicable.	ATA
R150	R1XX disk shelf	Not applicable.	ATA
	DS14mk2 AT	AT-FC	ATA
R200	DS14mk2 AT	AT-FC	ATA

For more information about disk support and capacity, see the *System Configuration Guide* on the NetApp on the Web (NOW) site at http://now.netapp.com/. When you access the *System Configuration Guide*, select the Data ONTAP version and storage system to find current information about all aspects of disk and disk shelf support and storage capacity.

Disk capacity

When you add a new disk, Data ONTAP reduces the amount of space on that disk available for user data by rounding down. This maintains compatibility across disks from various manufacturers. The available disk space listed by informational commands such as `sysconfig` is, therefore, less for each disk than its rated capacity (which you use if you specify disk size when creating an aggregate). The available disk space on a disk is rounded down as shown in the following table.

Disk	Right-sized Capacity	Available blocks
FC/SCSI disks		
4-GB disks	4 GB	8,192,000
9-GB disks	8.6 GB	17,612,800
18-GB disks	17 GB	34,816,000
35-GB disks (block checksum disks)	34 GB	69,632,000
36-GB disks (zoned checksum disks)	34.5 GB	70,656,000
72-GB disks	68 GB	139,264,000
144-GB disks	136 GB	278,528,000
288-GB disks	272 GB	557,056,000
ATA/SATA disks		
160-GB disks (available on R100 storage systems)	136 GB	278,258,000
250-GB disks (available on R150, R200, FAS900, and FAS3000 storage systems)	212 GB	434,176,000

Disk	Right-sized Capacity	Available blocks
320-GB disks (available on R200, FAS900, and FAS3000 storage systems)	274 GB	561,971,200

Disk speed

Disk speed is measured in revolutions per minute (RPM) and directly impacts input/output operations per second (IOPS) per drive as well as response time. Data ONTAP supports the following speeds for FC and ATA disk drives:

- FC disk drives
  - 10K RPM for FC disks of all capacities
  - 15K for FC disks with 36-GB and 72-GB capacities
- ATA disk drives
  - 5.4K RPM
  - 7.2K RPM

For more information about supported disk speeds, see the *System Configuration Guide*. For information about optimizing performance with 15K RPM FC disk drives, see the Technical Report (TR3285) on the NOW™ site at http://now.netapp.com/.

It is best to create homogenous aggregates with the same disk speed rather than mix drives with different speeds. For example, do not use10K and 15K FC disk drives in the same aggregate. If you plan to upgrade 10K FC disk drives to 15K FC disk drives, use the following process as a guideline:

1. Add enough 15K FC drives to create homogenous aggregates and FlexVol volumes (or traditional volumes) to store existing data.
2. Copy the existing data in the FlexVol volumes or traditional volumes from the 10K drives to the 15K drives.
3. Replace all existing 10K drives in the spares pool with 15K drives.

Disk checksum format

All new NetApp storage systems use block checksum disks (BCDs), which have a disk format of 520 bytes per sector. If you have an older storage system, it might have zoned checksum disks (ZCDs), which have a disk format of 512 bytes

per sector. When you run the `setup` command, Data ONTAP uses the disk checksum type to determine the checksum type of aggregates that you create. For more information about checksum types, see "How Data ONTAP enforces checksum type rules" on page 187.

Disk addressing

Disk addresses are represented in the following format:

HA.disk_id

HA refers to the host adapter (HA) number, which is the slot number on the storage system where the host adapter is attached, as shown in the following examples:

- `0a` —For a disk shelf attached to an onboard Fibre Channel host adapter
- `7` —For a disk shelf attached to a single-channel Fibre Channel host adapter installed in slot 7
- `7a` —For a disk shelf attached to a dual-channel Fibre Channel host adapter installed in slot 7, port A

disk_id is a protocol-specific identifier for attached disks. For Fibre Channel-Arbitrated Loop (FC-AL), the *disk_id* is an integer from 0 to 126. However, Data ONTAP only uses integers from 16 to 125. For SCSI, the *disk_id* is an integer from 0 to 15.

The *disk_id* corresponds to the disk shelf number and the bay in which the disk is installed, based on the disk shelf type. This results in a disk drive addressing map, which is typically included in the hardware guide for the disk shelf. The lowest *disk_id* is always in the far right bay of the first disk shelf. The next higher *disk_id* is in the next bay to the left, and so on. The following tables show examples of disk drive addressing maps for these disk shelves:

- Fibre Channel, DS14
- Fibre Channel, FC 7, 8, and 9
- NearStore, R100

Note

SCSI Enclosure Services (SES) is a program that monitors the disk shelf itself and requires that one or more bays always be occupied for SES to communicate with the storage system. These drives are referred to as SES drives.

Fibre Channel disk drive addressing maps:

The following table illustrates the shelf layout for the DS14 disk shelf. Note that the SES drives are in bay 0 and bay 1, and that the drive bay numbers begin with 16, on shelf ID 1.

DS14 Shelf ID	Bay													
	13	12	11	10	9	8	7	6	5	4	3	2	1	0
													SES drives	
7	125	124	123	122	121	120	119	118	117	116	115	114	113	112
6	109	108	107	106	105	104	103	102	101	100	99	98	97	96
5	93	92	91	90	89	88	87	86	85	84	83	82	81	80
4	77	76	75	74	73	72	71	70	69	68	67	66	65	64
3	61	60	59	58	57	56	55	54	53	52	51	50	49	48
2	45	44	43	42	41	40	39	38	37	36	35	34	33	32
1	29	28	27	26	25	24	23	22	21	20	19	18	17	16

The following table illustrates the shelf layout for the FC7, FC8, and FC9 disk shelves. Note that the SES drives are in bay 3 and bay 4, and that the drive bay numbers begin with 0, on shelf ID 0.

FC7, FC8, FC9 Shelf ID	Bay						
	6	5	4	3	2	1	0
			SES drives				
7	62	61	60	59	58	57	56
6	54	53	52	51	50	49	48
5	46	45	44	43	42	41	40
4	38	37	36	35	34	33	32
3	30	29	28	27	26	25	24

FC7, FC8, FC9 Shelf ID	Bay						
	6	5	4	3	2	1	0
			SES drives				
2	22	21	20	19	18	17	16
1	14	13	12	11	10	9	8
0	6	5	4	3	2	1	0

NearStore disk drive addressing map: The following table illustrates the shelf layout for the R100 and R150 disk shelves. Note that bays 4 through 7 are not shown.

R100, R150 Shelf ID	Bay											
	15	14	13	12	11	10	9	8	3	2	1	0
1	15	14	13	12	11	10	9	8	3	2	1	0

RAID group disk type

The RAID group disk type determines how the disk will be used in the RAID group. A disk cannot be used until it is configured as one of the following RAID group disk types and assigned to a RAID group.

- Data disk
- Hot spare disk
- Parity disk
- Double-parity disk

For more details on RAID group disk types, see “Understanding RAID groups” on page 136.

Disk configuration and ownership

About configuration and ownership

NetApp storage systems and components require initial configuration, most of which is performed at the factory. Once the storage system is configured, the disks must be assigned to a storage system using the hardware- or software-based disk ownership method to be accessed for data storage.

This section covers the following topics:

- "Initial configuration" on page 54
- "Hardware-based disk ownership" on page 55
- "Software-based disk ownership" on page 58

Initial configuration

How disks are initially configured

Disks are configured at the factory or at the customer site, depending on the hardware configuration and software licenses of the storage system. The configuration determines the method of disk ownership. A disk must be assigned to a storage system before it can be used as a spare or in a RAID group. If disk ownership is hardware based, disk assignment is performed by Data ONTAP. Otherwise, disk ownership is software based, and you must assign disk ownership.

Technicians install disks with the latest firmware. Then they configure some or all of the disks, depending on the storage system and which method of disk ownership is used.

- If the storage system uses hardware-based disk ownership, they configure all of the disks as spare disks, which are in a pool of hot spare disks, named Pool0 by default.
- If the storage system uses software-based disk ownership, they only configure enough disks to create a root volume. You must assign the remaining disks as spares at first boot before you can use them to create aggregates and volumes.

You might need to upgrade disk firmware for FC-AL or SCSI disks when new firmware is offered, or when you upgrade the Data ONTAP software. However, you cannot upgrade the firmware for ATA disks unless there is an AT-FCX module installed in the disk shelf.

Hardware-based disk ownership

Disk ownership supported by storage system model

Storage systems that support only hardware-based disk ownership include NearStore, F800 series and the FAS250 storage systems. Storage systems that support only software-based disk ownership include the FAS270 and V-Series storage systems.

The FAS900 and FAS3000 series storage systems can be either a hardware- or a software-based system. If a storage system that has CompactFlash also has the SnapMover license enabled, it becomes a software-based disk ownership storage system.

The following table lists the type of disk ownership that is supported by NetApp storage systems.

Storage System	Hardware-based	Software-based
R100 series R200 series	X non-clustered only	
FAS250	X non-clustered only	
FAS270		X
V-Series		X
F87	X	
F800 series	X	
FAS900 series	X	X (with SnapMover license)
FAS3000 series	X	X (with SnapMover license)

How hardware-based disk ownership works

Hardware-based disk ownership is determined by two conditions: how a storage system is configured and how the disk shelves are attached to it.

Without Multipath I/O: If the storage system is *not* configured for Multipath I/O, the disk ownership is based on the following rules:

- If clustering is not enabled, the single storage system owns all of the disks directly attached to it. This rule applies to direct-attached SCSI and NearStore ATA disks. For FC-AL disks, this rule applies to which port the disk shelf is attached to, which corresponds to the A loop or the B loop.
- If clustering is enabled, the local storage system owns direct FC-AL attached disks connected to it on the A loop and its partner owns the disks connected to it on the B loop.

Note

Clustering is considered enabled if an InterConnect card is installed in the storage system, it has a partner-sysid environment variable, or it has the clustering license installed and enabled.

- In either a single or clustered storage system with SAN switch-attached disks, a storage system with *even* switch port parity owns FCFLA attached disks whose A loop is attached to an even switch port or whose B loop is attached to an odd switch port.
- In either a single or clustered storage system with SAN switch-attached disks, a storage system with *odd* switch port parity owns FCFLA attached disks whose A loop is attached to an odd switch port or whose B loop is attached to an even switch port.
- In a clustered storage system with SAN disks attached with two switches, the above two rules apply to disks on both switches.
- For information about V-Series systems, see the V-Series *Software Setup, Installation and Administration Guide*.

With Multipath I/O: If the storage system *is* configured for Multipath I/O, there are three methods supported that use hardware-based disk ownership rules (using Multipath without SyncMirror, with SyncMirror, and with four separate host adapters). For detailed information on how to configure storage system using Multipath I/O, see "Multipath I/O for Fibre Channel disks" on page 69.

Functions performed for all hardware-based systems

For all hardware-based disk ownership storage systems, Data ONTAP performs the following functions:

- Recognizes all of the disks at bootup or when they are inserted into a disk shelf.

- Initializes all disks as spare disks.
- Automatically puts all disks into a pool until they are assigned to a RAID group.
- The disks remain spare disks until they are used to create aggregates and are designated as data disks or as parity disks by you or by Data ONTAP.

Note

Some storage systems that use hardware-based disk ownership do not support cluster failover, for example, NearStore (the R100 and R200 series) systems.

How disks are assigned to pools when SyncMirror is enabled

All spare disks are in pool0 unless the SyncMirror software is enabled. If SyncMirror is enabled on a hardware-based disk ownership storage system, all spare disks are divided into two pools, Pool0 and Pool1. For hardware-based disk ownership storage systems, disks are automatically placed in pools based on their location in the disk shelves, as follows:

- For all storage systems (except the FAS3000 series)
  - Pool0 - Host adapters in PCI slots 1-7
  - Pool1 - Host adapters in PCI slots 8-11
- For FAS3000 series
  - Pool0 - Onboard adapters 0a, 0b, and host adapters in PCI slots 1-2
  - Pool1 - Onboard adapters 0c, 0d, and host adapters in PCI slots 3-4

Software-based disk ownership

About software-based disk ownership

Software-based disk ownership software assigns ownership of a disk to a specific storage system by writing software ownership information on the disk rather than by using the topology of the storage system's physical connections. Software-based disk ownership is implemented in storage systems where a disk shelf can be accessed by more than one storage system. Configurations that use software-based disk ownership include

- FAS270 storage systems
- Any storage system with a SnapMover license
- Clusters configured for SnapMover vFiler™ migration. For more information, see the section on the SnapMover vFiler no copy migration feature in the *MultiStore Management Guide*.
- V-Series arrays. For more information, see the section on SnapMover in the V-Series Systems *Software Setup, Installation, and Management Guide.*
- FAS900 series or higher storage systems configured with SharedStorage

FAS270 storage systems: The NetApp FAS270 and FAS270c storage systems consist of a single disk shelf of 14 disks and either one internal system head (on the FAS270) or two clustered internal system heads (on the FAS270c) . By design, a disk located on this common disk shelf can, if the storage system has two system heads, be assigned to the ownership of either system head. The ownership of each disk is ascertained by an ownership record written on each disk.

NetApp delivers the FAS270 and FAS270c storage systems with each disk preassigned to the single FAS270 internal system head or preassigned to one of the two FAS270c system heads.

If you add one or more disk shelves to an existing FAS270 or FAS270c storage system, you might have to assign ownership of the disks contained on those shelves.

Software-based disk ownership tasks

You can perform the following tasks:

- Display disk ownership
- Assign disks
- Modify disk assignments

- Re-use disks that are configured for software-based disk ownership
- Erase software-based disk ownership prior to removing a disk
- Automatically erase disk ownership information
- Undo accidental conversion to software-based disk ownership

Displaying disk ownership

To display the ownership of all disks, complete the following step.

Step	Action
1	Enter the following command to display a list of all the disks visible to the storage system, whether they are owned or not. `sh1> **disk show -v**`

Note

You must use `disk show` to see unassigned disks. Unassigned disks are not visible using higher level commands such as the `sysconfig` command.

Sample output: The following sample output of the `disk show -v` command on an FAS270c shows disks 0b.16 through 0b.29 assigned in odd/even fashion to the internal cluster nodes (or system heads) sh1 and sh2. The fourteen disks on the add-on disk shelf are still unassigned to either system head.

```
sh1> disk show -v
 DISK      OWNER              POOL   SERIAL NUMBER
 --------- ---------------    -----  -------------
 0b.43     Not Owned          NONE   41229013
 0b.42     Not Owned          NONE   41229012
 0b.41     Not Owned          NONE   41229011
 0b.40     Not Owned          NONE   41229010
 0b.39     Not Owned          NONE   41229009
 0b.38     Not Owned          NONE   41229008
 0b.37     Not Owned          NONE   41229007
 0b.36     Not Owned          NONE   41229006
 0b.35     Not Owned          NONE   41229005
 0b.34     Not Owned          NONE   41229004
 0b.33     Not Owned          NONE   41229003
 0b.32     Not Owned          NONE   41229002
 0b.31     Not Owned          NONE   41229001
 0b.30     Not Owned          NONE   41229000
 0b.29     sh1 (84165672)     Pool0  41226818
 0b.28     sh2 (84165664)     Pool0  41221622
```

```
0b.27     sh1 (84165672)   Pool0  41226333
0b.26     sh2 (84165664)   Pool0  41225544
0b.25     sh1 (84165672)   Pool0  41221700
0b.24     sh2 (84165664)   Pool0  41224003
0b.23     sh1 (84165672)   Pool0  41227932
0b.22     sh2 (84165664)   Pool0  41224591
0b.21     sh1 (84165672)   Pool0  41226623
0b.20     sh2 (84165664)   Pool0  41221819
0b.19     sh1 (84165672)   Pool0  41227336
0b.18     sh2 (84165664)   Pool0  41225345
0b.17     sh1 (84165672)   Pool0  41225446
0b.16     sh2 (84165664)   Pool0  41201783
```

Additional `disk show` parameters are listed below.

`disk show` parameters	**Information displayed**
`disk show -a`	Displays all assigned disks
`disk show -n`	Displays all disk that are not assigned
`disk show -o` *`ownername`*	Displays all disks owned by the storage system or system head whose name is specified by *ownername*
`disk show -s` *`sysid`*	Displays all disks owned by the storage system or system specified by its serial number, *sysid*
`disk show -v`	Displays all the visible disks

Assigning disks

To assign disks that are currently labeled “not owned,” complete the following steps.

Step	**Action**
1	Use the `disk show -n` command to view all disks that do not have assigned owners.

Step	Action
2	Use the following command to assign the disks that are labeled “Not Owned” to one of the system heads. If you are assigning unowned disks to a non-local storage system, you must identify the storage system by using either the `-o` *`ownername`* or the `-s` *`sysid`* parameters or both. **`disk assign {`*`disk_name`*` |all| -n count} [-p pool] [-o ownername] [-s sysid] [-c block|zoned] [-f]`** *disk_name* specifies the disk that you want to assign to the storage system or system head. `all` specifies all of the unowned disks are assigned to the storage system or system head. `-n` *`count`* specifies the number of unassigned disks to be assigned to the storage system or system head, as specified by `count`. `-p` *`pool`* specifies which SyncMirror pool the disks are assigned to. The value of pool is either 0 or 1. `-o` *`ownername`* specifies the storage system or the system head that the disks are assigned to. `-s` *`sysid`* specifies the storage system or the system head that the disks are assigned to. `-c` specifies the checksum type (either block or zoned) for a LUN in V-Series systems. `-f` must be specified if the storage system or system head already owns the disk. **Example:** The following command assigns six disks on the FAS270c to the system head sh1: `sh1>` **`disk assign 0b.43 0b.41 0b.39 0b.37 0b.35 0b.33`** **Result:** The specified disks are assigned as disks to the system head on which the command was executed.
3	Use the `disk show -v` command to verify the disk assignments that you have just made.

After you have assigned ownership to a disk, you can assign that disk to the aggregate on the storage system that owns it, or leave it as a spare disk on that storage system.

Note

You cannot download firmware to unassigned disks.

Modifying disk assignments

You can also use the `disk assign` command to modify the ownership of any disk assignment that you have made. For example, on the FAS270c, you can reassign a disk from one system head to the other. On either the FAS270 or FAS270c storage system, you can change an assigned disk back to "Not Owned" status.

Attention

You should only modify disk assignments for spare disks. Disks that have already been assigned to an aggregate cannot be reassigned without endangering all the data and the structure of that entire aggregate.

To modify disk ownership assignments, complete the following steps.

Step	Action
1	View the spare disks whose ownership can safely be changed by entering the following command: **`aggr status -r`**

Step	Action
2	Use the following command to modify assignment of the spare disks. **`disk assign {`*`disk1`* **`[`***`disk2`***`] [...]|-n`** ***`num_disks`***`} -f {-o`** ***`ownername`*** **`| -s unowned	-s`** ***`sysid`***`}`** *disk1* [*disk2*] [...] are the names of the spare disks whose ownership assignment you want to modify. `-n` *num_disks* specifies a number of disks, rather than a series of disk names, to assign ownership to. `-f` forces the assignment of disks that have already been assigned ownership. `-o` *ownername* specifies the host name of the storage system head to which you want to reassign the disks in question. `-s unowned` modifies the ownership assignment of the disks in question back to "Not Owned." `-s` *sysid* is the factory-assigned NVRAM number of the storage system head to which you want to reassign the disks. It is displayed with the `sysconfig` command. **Example:** The following command unassigns four disks on the FAS270c from the storage system sh1: `sh1>` **`disk assign 0b.30 0b.29 0b.28 0b.27 -s 003303542 unowned`**
3	Use the `disk show -v` command to verify the disk assignment modifications that you have just made.

Re-using disks that are configured for software-based disk ownership

If you want to re-use disks from storage systems that have been configured for software-based disk ownership, you should take precautions if you reinstall these disks in storage systems that do not use software-based disk ownership.

Attention

Disks with unerased software-based ownership information that are installed in an unbooted storage system that does not use software-based disk ownership will cause that storage system to fail on reboot.

Take the following precautions, as appropriate:

- Erase the software-based disk ownership information from a disk prior to removing it from its original storage system. See "Erasing software-based disk ownership prior to removing a disk" on page 64.
- Transfer the disks to the target storage system while that storage system is in operation. See "Automatically erasing disk ownership information" on page 65.
- If you accidentally cause a boot failure by installing software-assigned disks, undo this mishap by running the `disk remove_ownership` command in maintenance mode. See "Undoing accidental conversion to software-based disk ownership" on page 66.

Erasing software-based disk ownership prior to removing a disk

If possible, you should erase software-based disk ownership information on the target disks before removing them from their current storage system and prior to transferring them to another storage system.

To undo software-based disk ownership on a target disk prior to removing it, complete the following steps.

Step	Action
1	At the prompt of the storage system whose disks you want to transfer, enter the following command to list all the storage system disks and their RAID status. **`aggr status -r`** Note the names of the disks that you want to transfer. **Note** In most cases, (unless you plan to physically move an entire aggregate of disks to a new storage system) you should plan to transfer only disks listed as hot spare disks.
2	For each disk that you want to remove, enter the following command: **`disk remove_ownership`** ***`disk_name`*** *disk_name* is the name of the disk whose software-based ownership information you want to remove.

Step	Action
3	Enter the following command to confirm the removal of the disk ownership information from the specified disk. `disk show -v` **Result:** The specified disk and any other disk that is labeled "not owned" is ready to be moved to other storage systems.
4	Remove the specified disk from its original storage system and install it into its target storage system.

Automatically erasing disk ownership information

If you physically transfer disks from a storage system that uses software-based disk ownership to a running storage system that does not, you can do so without using the `disk remove_ownership` command if that storage system is running Data ONTAP 6.5.1 or higher.

To automatically erase disk ownership information by physically transferring disks to a non-software-based storage system, complete the following steps.

Step	Action	
1	Do not shut down the target storage system.	
2	On the target storage system, enter the following command to confirm the version of Data ONTAP on the target storage system. `version`	
3	**If**	**Then**
	The Data ONTAP version on the target storage system is 6.5.1 or later	Go to Step 4.
	The Data ONTAP version on the target storage system is earlier than 6.5.1	Do not continue this procedure; instead, erase the software-based disk ownership information on the source storage system, as described in "Erasing software-based disk ownership prior to removing a disk" on page 64.

Step	Action
4	Remove the disks from their original storage system and physically install them in the running target storage system. If Data ONTAP 6.5.1 or later is installed, the running target storage system automatically erases any existing software-based disk ownership information on the transferred disks.
5	On the target storage system, use the `aggr status -r` command to verify that the disks you have added are successfully installed.

Undoing accidental conversion to software-based disk ownership

If you transfer disks from a storage system configured for software-based disk ownership (such as the FAS270 storage system, or a cluster enabled for SnapMover vFiler™ migration) to another storage system that does not use software-based disk ownership, you might accidentally mis-configure that target storage system as a result of the following circumstances.

- You neglect to remove software-based disk ownership information from the target disks before you remove them from their original storage system.
- You add the disks to a target storage system that does not use software-based disk ownership while the target storage system is off.
- The target storage system is upgraded to Data ONTAP 6.5.1 or later.

Under these circumstances, if you reboot the target storage system in normal mode, the remaining disk ownership information causes the target storage system to convert to a mis-configured software-based disk ownership setup. It will fail to reboot.

To undo this accidental conversion to software-based disk ownership, complete the following steps.

Step	Action
1	Turn on or reboot the target storage system. When prompted to do so, press Ctrl-C to display the boot menu.
2	Enter the choice for booting in maintenance mode.

Step	Action
3	In maintenance mode, enter the following command: **`disk remove_ownership all`** The software-based disk ownership information is erased from all disks that have them.
4	Halt the storage system to exit maintenance mode by entering the following command: **`halt`**
5	Reboot the target storage system. The storage system will reboot in normal mode with software-based disk ownership disabled.

Disk access methods

About disk access methods

Several disk access methods are supported on NetApp storage systems, which are covered in the following topics:

- "Multipath I/O for Fibre Channel disks" on page 69
- "Clusters" on page 75
- "Combined head and disk shelf storage systems" on page 76
- "SharedStorage" on page 77

Multipath I/O for Fibre Channel disks

Understanding Multipath I/O

The Multipath I/O feature for Fibre Channel disks enables you to create two paths, a primary path and a secondary path, from a single storage system to a disk loop. You can use this feature with or without SyncMirror.

Although it is not necessary to have a dual-port disk host adapter to set up Multipath I/O, you should use two dual-port host adapters to connect to two disk shelf loops, thus preventing either host adapter from being a single point of failure. In addition, using dual-port host adapters conserves Peripheral Component Interconnect (PCI) slots.

If your environment requires additional fault tolerance, you can use Multipath I/O with SyncMirror and configure it with four separate host adapters, connecting one path from each host adapter to one channel of a disk shelf. With this configuration, not only is each path supported by a separate host adapter, but each host adapter is on a separate bus. If there is a bus failure, or a host adapter failure, only one path is lost.

Advantages of Multipath I/O

By providing redundant paths to the same disk on a single storage system, the Multipath I/O feature offers the following advantages:

- Overall reliability and uptime of the storage subsystem is increased.
- Disk availability is higher.
- Bandwidth is increased (each loop provides an additional 200 MB/second of bandwidth).
- Storage subsystem hardware can be maintained with no downtime.

  When a primary host adapter is brought down, all traffic moves from that host adapter to the secondary host adapter. As a result, you can perform maintenance tasks, such as replacing a malfunctioning Loop Resiliency Circuit (LRC) module or cables connecting that host adapter to the disk shelves, without affecting the storage subsystem service.

Requirements to enable Multipath I/O on the storage system

The Multipath I/O feature is enabled automatically, subject to the following restrictions:

- Only the following storage systems support Multipath I/O:
  - F800 series
  - FAS900 series
  - FAS3000 series

  Note

  None of the NearStore storage systems (R100, R150, or R200 series) support Multipath I/O.

- Only the following host adapters support Multipath I/O:
  - QLOGIC 2200 (P/N X2040B)
  - QLOGIC 2212 (X2044A, 2044B)
  - QLOGIC 2342 (X2050A)
  - LSI 929X (X2050B)

  Note

  Although the 2200 and 2212 host adapters can co-exist with older (2100 and 2000) adapters on a storage system, Multipath I/O is not supported on older storage systems.

  To determine the slot number where a host adapter can be installed in your storage system, see the *System Configuration Guide* at the NOW site (http://now.netapp.com/).

- FC7 and FC8 disk shelves do not support Multipath I/O.
- FC9 must have two LRC modules to support Multipath I/O.
- DS14 and DS14mk2 FC disk shelves must have either two LRC modules or two Embedded Switch Hub (ESH) modules to support Multipath I/O.
- Older 9-GB disks (ST19171FC) and older 18-GB disks (ST118202FC) do not support Multipath I/O.
- Storage systems in a MetroCluster configuration support Multipath I/O

  Multipath I/O setup and clustering setup both require the A and B ports of the disk shelves. Therefore, it is not possible to have both features enabled simultaneously.

  Note

  Storage systems configured in a standard (CFO) cluster do not support Multipath I/O.

- Hardware connections must be set up for Multipath I/O as specified in the corresponding *Fibre Channel StorageShelf* guide.
- SharedStorage configurations require Multipath I/O.

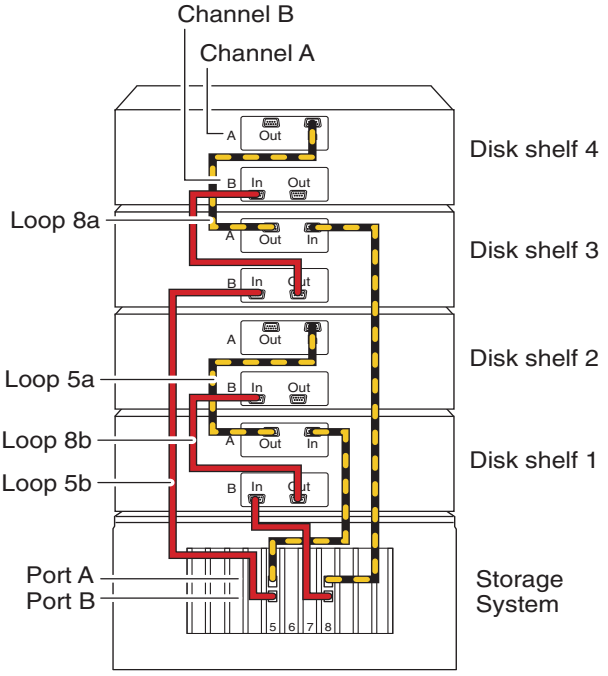
Supported configurations

Multipath I/O supports the following configurations:

- "Multipath I/O without SyncMirror" on page 71
- "Multipath I/O with SyncMirror using hardware-based disk ownership" on page 72
- "Multipath I/O with SyncMirror using software-based disk ownership" on page 73
- "Multipath I/O with SyncMirror, using four separate adapters" on page 74

Multipath I/O without SyncMirror: Configure a single storage system for Multipath I/O without SyncMirror by connecting a primary path from one host adapter to one disk loop and a secondary path from another host adapter to that disk loop, as shown in the following illustration. To display the paths using the `storage show disk -p` command, see "Example 1" on page 89.

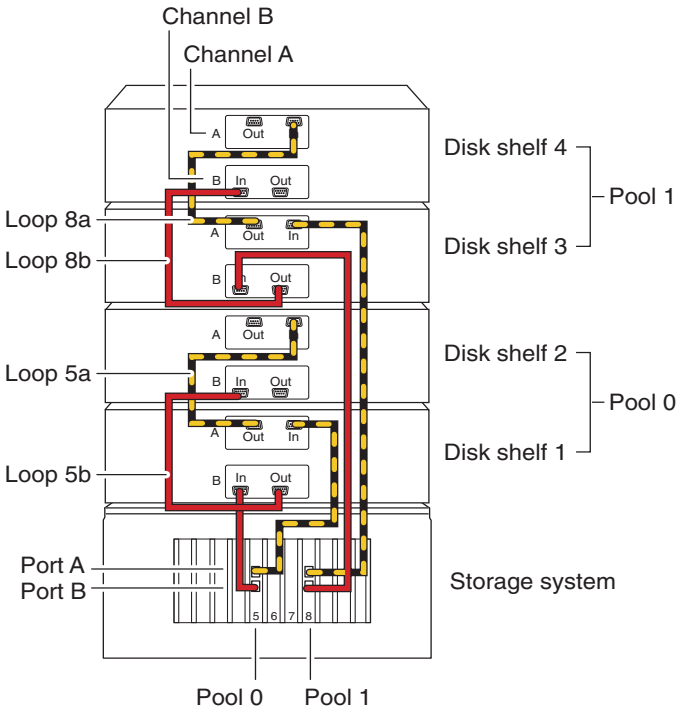
- The first loop is configured as follows:
  - Primary path: from system port 5a to disk shelves 1 and 2, A channels
  - Secondary path: from system port 8b to disk shelves 1 and 2, B channels
- The second loop is configured as follows:
  - Primary path: from system port 8a to disk shelves 3 and 4, A channels
  - Secondary path: from system port 5b to disk shelves 3 and 4, B channels



MPIO without SyncMirror

Multipath I/O with SyncMirror using hardware-based disk ownership: If your storage system does not support software-based disk ownership, you need to know which slots the adapters are in because that is what pool ownership is determined by. For example, with the FAS900 series, slots 1 through 7 own Pool0, and slots 8 through 11 own Pool1. In this case, you should configure the system to have a primary path and a secondary path connected from one adapter to the first disk loop and a primary and a secondary path from the other adapter to the second disk loop, as shown in the following illustration. To display the paths using the `storage show disk -p` command, see "Example 2" on page 90.

- The first loop is configured as follows:
  - Primary path: from system port 5a to disk shelves 1 and 2. A channels.
  - Secondary path: from system port 5b to disk shelves 1 and 2, B channels
- The second loop is configured as follows:
  - Primary path: from system port 8a to disk shelves 3 and 4, A channels
  - Secondary path: from system port 8b to disk shelves 3 and 4, B channels



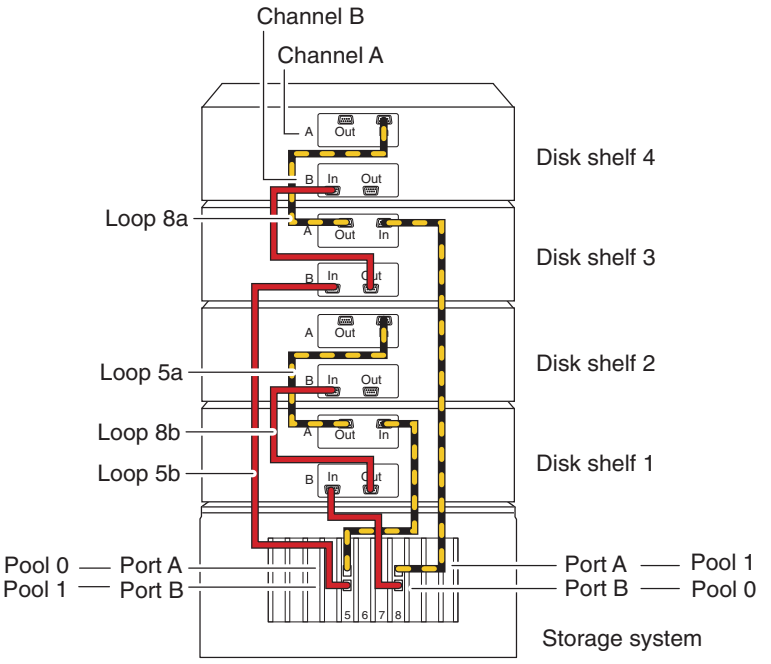
Multipath I/O with SyncMirror with hardware-based disk ownership

Multipath I/O with SyncMirror using software-based disk ownership:

If your storage system supports software-based disk ownership, you should configure the system to have a primary path and a secondary path from two different adapters to the first disk loop and a primary and a secondary path from the two adapters to the second disk loop, as shown in the following illustration. To display the paths using the `storage show disk -p` command, see "Example 3" on page 91.

- The first loop is configured as follows:
  - Primary path: from system port 5a to disk shelves 1 and 2, A channels
  - Secondary path: from system port 8b to disk shelves 1 and 2, A channels
    You can configure this as Pool0.

- The second loop is configured as follows:
  - Primary path: from system port 8a to disk shelves 3 and 4, A channels
  - Secondary path: from system port 5b to disk shelves 3 and 4, B channels
    You can configure this as Pool1.



MPIO with SyncMirror with software-based disk ownership

Multipath I/O with SyncMirror, using four separate adapters: If you want to provide the highest level of availability, you can configure Multipath I/O with SyncMirror using four separate adapters, one for each disk shelf. For the latest information about which slots to use for adapters in your specific storage system, see the *System Configuration Guide*.

Clusters

About clusters

NetApp clusters are two storage systems, or nodes, in a partner relationship where each node can access the other's disk shelves as a secondary owner. Each partner maintains two Fibre Channel Arbitrated Loops (or loops): a primary loop for a path to its own disks, and a secondary path to its partner's disk. The primary loop, loop A, is created by connecting the A ports of one or more disk shelves to the node's disk adapter card, and the secondary loop, loop B, is created by connecting the B ports of one or more disk shelves to the node's disk adapter card.

If one of the clustered nodes fails, its partner can start an emulated storage system that takes over serving the failed partner's disk shelves, providing uninterrupted access to its partner's disks as well as its own disks. For more information on installing or managing clusters, see the *Cluster Installation and Administration Guide*.

Moving data outside of a cluster

You can move data outside a cluster without having to copy data using the vFiler migrate feature (for NFS only). You place a traditional volume into a vFiler unit and move the volume using the `vfiler migrate` command. For more information, see the *MultiStore Management Guide*.

Combined head and disk shelf storage systems

About combined head and disk shelf storage systems

Some storage systems combine one or two system heads and a disk shelf into a single unit. For example, the FAS270c consists of two clustered system heads that share control of a single shelf of fourteen disks.

Primary clustered system head ownership of each disk on the shelf is determined by software-based disk ownership information stored on each individual disk, not by A loop and B loop attachments. You use software-based disk ownership commands to assign each disk to the FAS270 system heads, or any system with a SnapMover license.

For more information on software-based disk ownership assignment, see "Software-based disk ownership" on page 58.

SharedStorage

Understanding SharedStorage

Data ONTAP 7.1 supports SharedStorage, the ability to share a pool of disks amongst a community of NetApp storage systems, made up of two to four homogeneous NetApp FAS900 series and higher storage systems, without requiring any of the storage systems to be in a cluster. SharedStorage does not support using more than one kind of model in one community. For example, you cannot mix a FAS960 storage system with a FAS980 storage system.

You can configure SharedStorage with or without the vFiler no-copy migration functionality. If you do *not* want to use the vFiler no-copy migration functionality, you can create aggregates and FlexVol volumes in the community. If you want to use the vFiler no-copy migration functionality, you are restricted to creating only traditional volumes that are associated with a vFiler unit. For more information about how to use this functionality, see "vFiler no-copy migration software" on page 83.

The SharedStorage feature enables you to perform the following tasks:

- Increase disk capacity independently of the storage systems

  You can add disks (up to a maximum of 336) to any of the disk shelves and leave them unassigned. This allows you to provision spare disks amongst the community of storage systems rather than provision disks for each storage system individually.
- Assign or provision individual disks across up to four storage systems to expand traditional volumes and aggregates
- Assign dual paths to clustered storage systems in the community
- Assign independent paths to each shelf in the community (however, you cannot daisy-chain shelves)

In addition, SharedStorage uses a standardized back-end architecture, which provides the following benefits:

- Easy-to-use all-optical cabling and storage controllers
- Reduced spares cost, because only one FRU is needed
- Cabling flexibility, because there are multiple distance options for optical cables
- Optimized bandwidth, because there are dedicated 2-Gb optical dual paths to all shelves

How SharedStorage works

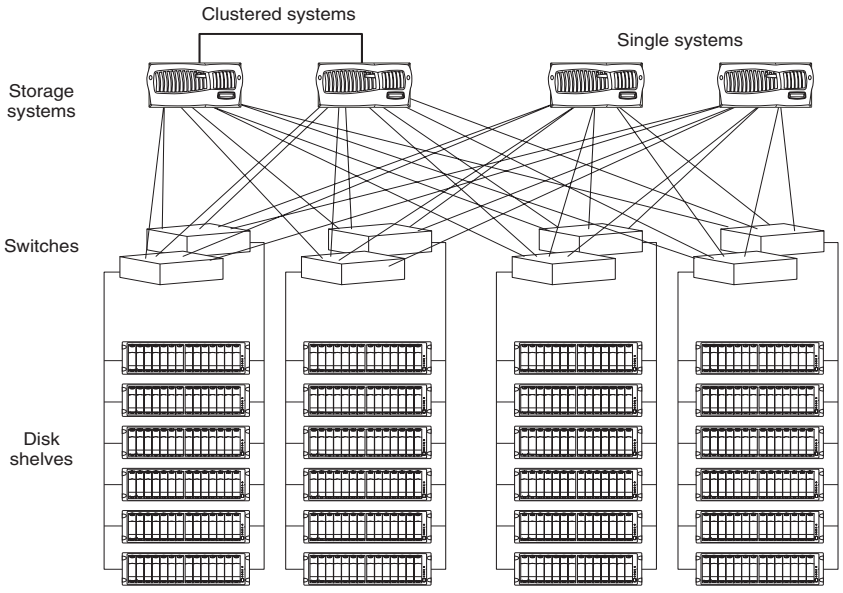
SharedStorage uses external Fibre Channel hubs to connect all of the disks to all of the storage systems in the community. Each storage system can also use the hub to communicate with every other storage system. Each storage system is both an initiator and a target, so all of the storage systems can submit and receive FC requests. The storage systems can also share SES information and controls as well as state information when performing upgrades of disk firmware and other tasks.

Two hubs are connected to each storage system and each one controls an FC-AL loop, either an A loop or a B loop, to provide redundancy. Each storage system supports up to four A and four B loops. Up to six disk shelves can be directly connected to a loop switch port on each hub, so that all connected ports are logically on the same FC-AL loop.

You can set up the storage systems in the following configurations with full multiprotocol support, including NFS, CIFS, FCP, and iSCSI:

- One or two clusters
- One cluster with one or two single storage systems
- Two to four single storage systems

The following diagram shows four storage systems, with the first two configured as a cluster. The nodes in the cluster are directly connected to each other with IB cluster adapter cables (notice that the cluster interconnect cables are not attached to the hubs).



You use software-based disk ownership to assign disks to storage systems. Each disk is dually connected, and the paths to each disk go through different disk adapters, which means that loss of a single adapter, hub, cable connection, or I/O module can be tolerated.

All of the storage systems can communicate with each other as well as all of the disk shelves and the disks in the community. Up to two storage systems can control the SES disk drives of a given disk shelf. In each shelf, at least one SES drive bay must be occupied by a disk. This allows any storage system to turn on any disk shelf's LED lights, check its environment, receive shelf status. or perform upgrades of disk firmware.

SyncMirror is supported with SharedStorage. For information about the SyncMirror rules regarding pools, see "How disks are assigned to pools when SyncMirror is enabled" on page 57.

How to install a SharedStorage community

Installing a SharedStorage community requires SupportEdge Premium Support service, and the Installation Service is mandatory. For information, contact your NetApp Sales representative.

The requirements for using SharedStorage include the following components:

- Two to four homogenous NetApp FAS900 series storage systems with four dual-ported QLogic FC HBAs (as clustered pairs or not, in any combination)
- DS14mk2 shelves, up to six shelves per storage system
- ESH or ESH2 shelf modules (two per each shelf)
- Emulex InSpeed 370 20-port loop switches, (two per each storage system)
- Up to 336 disks (4 pairs of loops, 6 shelves, 14 disks per shelf)

For wiring information, see the *Installation and Setup Instructions for NetApp SharedStorage*. These instructions include the software setup procedure for booting the storage systems the first time.

After you have completed the setup procedure, verify the following:

- The lights on all of used hub ports are green.
- Each storage system sees all disks, which all have a primary and a secondary path (use the `storage show disk -p` command to display both paths).
- Each storage system sees all host adapters (use the `storage show adapter` command to display information about all or the specified adapter that is installed in a given slot).

- Each storage system sees the all of the other storage systems (use the `storage show initiator` command to see a list of the initiator systems in the community).

Using software-based disk ownership

SharedStorage uses software-based disk ownership. For information on how to manage disks using software-based ownership, see "Software-based disk ownership" on page 58.

You assign disks in a community using the same command as you do for single or clustered storage systems under most circumstances. However, there are a few exceptions:

You can unassign disk ownership of a disk that is owned by a storage system by assigning it as `unowned`, as shown in the following example:

```
shared_1> disk assign 0b.16 -s unowned -f
```

The result of this command is that the disk is returned to the unowned pool.

You can also assign ownership of spare disks from one storage system to another, as shown in the following example:

```
shared_1> disk assign 0b.17 -o shared_2 -f
```

If there is a communication problem between the two storage systems, you will see warnings about "rescan messages".

Managing disks with SharedStorage

If you use the Data ONTAP command-line interface (CLI), you should assign disks and spares to each storage system and leave the rest in a large unowned pool. Assign disks from the unowned pool when you want to

- Increase the size of an aggregate or a traditional volume if you are using the vFiler no-copy migration feature
- Add a new aggregate or a traditional volume if you are using the vFiler no-copy migration feature
- Replace a failed disk

If you use the FilerView or DataFabric® Manager graphical user interface, which do not recognize an unowned disk pool, you should assign all of the disks as spares to one storage system. This makes it easier to reassign disks for the tasks listed above.

Note

If you always use volumes of the same size, you can reassign all volumes and all vFiler units, and migrate a vFiler unit to the required storage system, when necessary.

Managing spare disks: If Data ONTAP needs a spare disk to replace a failed disk, it selects one that is assigned to that storage system. You should assign as many spares as possible to storage systems that are experiencing a higher disk failure rate. If necessary, you can assign more disks from the unowned pool of spare disks.

Allocating disks: If a storage system needs more storage, use the `disk assign` command to reassign spare disks to that storage system. The newly reassigned disks are then added to the traditional volume.

Note

You cannot assign disks to qtrees or FlexVol volumes.

Displaying information about disks: To see information about the disks owned by one storage system, complete the following step.

Step	Action

1	Enter the following command:
	`shared_1>` **`disk show`**
	`DISK OWNER POOL SERIAL NUMBER`
	`------ ----------- (--------) ------ --------------`
	`9b.19 shared_1 (0050408412) Pool0 3HZ6RA1B0000742SWC9`
	`3a.22 shared_1 (0050408412) Pool0 3HZ6DGM000074310Z3A`
	`2b.104 shared_1 (0050408412) Pool0 414W5505`
	`2b.106 shared_1 (0050408412) Pool0 414X5475`

About initiators and targets

Each storage system can behave as an initiator or a target. The storage system behaves as an initiator when it reads and writes data to disks. The storage system behaves as a target when it communicates with disks and disk shelves to download firmware, share SES information with other storage systems or share information with an FC adapter card.

Displaying initiators

To display the initiators in a SharedStorage community, complete the following step.

Step	Action

1	Enter the following command:
	`shared_1>` **`storage show initiators`**
	`HOSTNAME  SYSTEM ID`
	`----------  ----------`
	`shared_1  0050408412 (self)`
	`shared_2  0050408123`
	`shared_3  0050408133`
	`shared_4  0050408717`

To display the all of the initiators in the loop, complete the following step.

Step	Action

1	Enter the following command:
	`shared_1>` **`fcadmin device_map`**
	`Loop Map for channel 3b:`
	`Translated Map: Port Count 73`
	`0 7 16 17 18 19 20 21 22 23 24 25 26 27 28 29`
	`32 33 34 35 36 37 38 39 40 41 42 43 44 45 48 49`
	`50 51 52 53 54 55 56 57 58 59 60 61 80 81 82 83`
	`84 85 86 87 88 89 90 91 92 93 96 97 98 99 100 101`
	`102 103 104 105 106 107 108 109 1 2`
	
	`Shelf mapping:`
	`Shelf 1: 29 28 27 26 25 24 23 22 21 20 19 18 17 16`
	`Shelf 2: 45 44 43 42 41 40 39 38 37 36 35 34 33 32`
	`Shelf 3: 61 60 59 58 57 56 55 54 53 52 51 50 49 48`
	`Shelf 5: 93 92 91 90 89 88 87 86 85 84 83 82 81 80`
	`Shelf 6: 109 108 107 106 105 104 103 102 101 100 99 98 97 96`
	
	`Initiators on this looop:`
	`0 (self) 1 (shared_2) 7 (shared_3) 2 (shared_4)`

vFiler no-copy migration software

The vFiler no-copy migration software supports NFS (non-disruptive) and CIFS. If you want to use the vFiler no-copy migration software, you are restricted to creating only traditional volumes and you must have the following licenses installed on your storage systems:

- SnapMover
- MultiStore

There are a few limitations with the vFiler migrate feature:

- Root volumes in vFiler units cannot be migrated.
- vFiler functionality is not supported for iSCSI or FCP.
- When you move a volume to another storage system, VSM, QSM, and NDMP relationships must be re-established on the new storage system.
- You can only move entire vFiler units. Use a one-to-one ratio for mapping traditional volumes to vFiler units.

With vFiler no-copy migration software installed, you can perform the following tasks:

- Perform non-disruptive maintenance

  You can isolate storage systems and disks, take them offline, perform maintenance and bring them back online without taking a loop out of service.

  The SharedStorage hubs allow for multiple paths to the storage, which allow for hot swappable ESH controller modules and the ability to take one path to the storage offline, even in a CFO pair.

  With vFiler no-copy migration functionality, you can migrate a traditional volume from one storage system to another, thereby isolating the first storage system to perform system maintenance while the target storage system continues to serve data.
- Coordinate disk and shelf firmware downloads

  SharedStorage technology ensures there is no disruption of service to all of the storage systems in the community when disk or disk shelf firmware is being downloaded to any disk or disk shelf.
- Balance workloads amongst the storage systems using vFiler no-copy migration

Balancing workloads amongst the community

You can balance workloads amongst the storage systems in the community by migrating traditional volumes that are associated with vFiler units. If one storage system in the community is CPU-bound with the workload from one vFiler unit, you can migrate that unit to another storage system within seconds using the no-

copy migration feature of SnapMover. For example, if you have four storage systems and one has a heavier load than the other three, use SnapMover to re-assign disks from CPU-bound head to another storage system. First you create a vFiler unit. Then you move its IP address by migrating from the overburdened storage system to an under-utilized storage system. The disks containing the volume change ownership from the overburdened storage system to the new one. As a result, you can balance traditional volumes across multiple storage systems in the community. For more information, see the *MultiStore Management Guide.*

Disk management

About disk management

You can perform the following tasks to manage disks:

- “Displaying disk information” on page 86
- “Managing available space on new disks” on page 94
- “Adding disks” on page 97
- “Removing disks” on page 100
- “Sanitizing disks” on page 105

Displaying disk information

Types of disk information

You can display a lot of information about disks by using the Data ONTAP CLI or FilerView.

Using the Data ONTAP CLI

The following table describes the Data ONTAP commands you can use to display status about disks.

Data ONTAP command	To display information about...
`df`	Disk space usage for file systems.
`disk maint status`	The status of disk maintenance tests that are in progress, after the `disk maint start` command has been executed.
`disk sanitize status`	The status of the disk sanitization process, after the `disk sanitize start` command has been executed.
`disk shm_stats`	SMART data from ATA disks.
`disk show`	Ownership. A list of disks owned by a storage system, or unowned disks. (for software-basked disk ownership systems only)
`fcstat device_map`	A physical representation of where the disks reside in a loop and a mapping of the disks to the disk shelves.
`fcstat fcal_stats`	Error and exceptions conditions, and handler code paths executed.
`fcstat link_stats`	Link event counts.

Data ONTAP command	To display information about...
`storage show disk`	The disk ID, shelf, bay, serial number, vendor, model, and revision level of all disks, or by the host disks associated with the specified host adapter (where *name* can be an electrical name, such as 4a.16, or a World Wide Name.
`storage show disk -a`	All information in a report form that is easily interpreted by scripts. This form also appears in the STORAGE section of an AutoSupport report.
`storage show disk -p`	Primary and secondary paths to a disk.
`sysconfig -d`	Disk address in the Device column, followed by the host adapter (HA) slot, shelf, bay, channel, and serial number.
`sysstat`	The number of kilobytes per second (kB/s) of disk traffic being read and written.

Examples of usage The following examples show how to use some of the Data ONTAP commands.

Displaying disk attributes: To display disk attribute information about all the disks connected to your storage system, complete the following step.

Step	Action

1	Enter one of the following commands:
	storage show disk
	Result: The following information is displayed.
	`system_0> storage show disk`
	`DISK  SHELF  BAY  SERIAL  VENDOR  MODEL  REV`
	`----  -----  ---  --------  ------  ---------  ----`
	`7a.16  1  0  414A3902  NETAP  X272_HJURE  NA14`
	`7a.17  1  1  414B5632  NETAP  X272_HJURE  NA14`
	`7a.18  1  2  414D3420  NETAP  X272_HJURE  NA14`
	`7a.19  1  3  414G4031  NETAP  X272_HJURE  NA14`
	`7a.20  1  4  414A4164  NETAP  X272_HJURE  NA14`
	`....`
	`7a.26  1  10  414D4510  NETAP  X272_HJURE  NA14`
	`7a.27  1  11  414C2993  NETAP  X272_HJURE  NA14`
	`7a.28  1  12  414F5867  NETAP  X272_HJURE  NA14`
	`7a.29  1  13  414C8334  NETAP  X272_HJURE  NA14`
	`7a.32  2  0  3HZY38RT0000732 NETAP  X272_SCHI6  NA05`
	`7a.33  2  2  3HZY38RT0000732 NETAP  X272_SCHI6  NA05`

Displaying the primary and secondary paths to the disks: To display the primary and secondary paths to all the disks connected to your storage system, complete the following step.

Step	Action
1	Enter the following command: **storage show disk -p**

Note

The disk addresses shown for the primary and secondary paths to a disk are aliases of each other.

In the following examples, dual host adapters, with the ports labeled as A and B, are installed in the PCI expansion slot 5 and slot 8 of a storage system. However, when Data ONTAP displays information about the adapter port label, it uses the lowercase a and b. Each disk shelf also has two ports, labeled A and B. When Data ONTAP displays information about the disk shelf port label, it uses the uppercase A and B.

The adapter in slot 8 is connected from its A port to port A of disk shelf 1, and the adapter in slot 5 is connected from its B port to port B of disk shelf 2. While it is not necessary to connect the adapter to the disk shelf using the same port label, doing so can help you keep track of adapter-to-shelf connections.

Each example displays the output of the `storage show disk -p` command, which shows the primary and secondary paths to all disks connected to the storage system. Each example represents a different configuration of Multipath I/O.

Example 1: In the following example, system_1 is configured for Multipath I/O without SyncMirror, as described at "Multipath I/O without SyncMirror" on page 71.

The first and third columns, labeled PRIMARY and SECONDARY, designate the primary and secondary paths from the adapter's slot number, host adapter port, and disk number.

```
system_1> storage show disk -p
PRIMARY PORT  SECONDARY  PORT  SHELF BAY
------- ----  ---------- ----  ----- ---
5a.16    A     8b.16    B      1     0
5a.17    A     8b.17    B      1     1
5a.18    B     8b.18    A      1     2
5a.19    A     8b.19    B      1     3
5a.20    A     8b.20    B      1     4
5a.21    B     8b.21    A      1     5
5a.22    A     8b.22    B      1     6
5a.23    A     8b.23    B      1     7
5a.24    B     8b.24    A      1     8
5a.25    B     8b.25    A      1     9
5a.26    A     8b.26    B      1    10
5a.27    A     8b.27    B      1    11
5a.28    B     8b.28    A      1    12
5a.29    A     8b.29    B      1    13

5a.32    B     8b.32    A      2     0
5a.33    A     8b.33    B      2     1
```

```
5a.34    A     8b.34    B       2     2
...
5a.43    A     8b.43    B       2    11
5a.44    B     8b.44    A       2    12
5a.45    A     8b.45    B       2    13

8a.48    B     5b.48    A       3     0
8a.49    A     5b.49    B       3     1
8a.50    B     5b.50    A       3     2
...
8a.59    A     5b.59    B       3    11
8a.60    B     5b.60    A       3    12
8a.61    B     5b.61    A       3    13

8a.64    B     5b.64    A       4     0
8a.65    A     5b.65    B       4     1
8a.66    A     5b.66    B       4     2
...
8a.75    A     5b.75    B       4    11
8a.76    A     5b.76    B       4    12
8a.77    B     5b.77    A       4    13
```

Example 2: In the following example, system_2 is configured for Multipath I/O with SyncMirror using hardware-based disk ownership, as described at "Multipath I/O with SyncMirror using hardware-based disk ownership" on page 72.

```
system_2> storage show disk -p
PRIMARY PORT  SECONDARY  PORT  SHELF BAY
------- ----  ---------- ----  ----- ---
5a.16    A     5b.16    B       1     0
5a.17    A     5b.17    B       1     1
5a.18    B     5b.18    A       1     2
5a.19    A     5b.19    B       1     3
5a.20    A     5b.20    B       1     4
5a.21    B     5b.21    A       1     5
5a.22    A     5b.22    B       1     6
5a.23    A     5b.23    B       1     7
5a.24    B     5b.24    A       1     8
5a.25    B     5b.25    A       1     9
5a.26    A     5b.26    B       1    10
5a.27    A     5b.27    B       1    11
5a.28    B     5b.28    A       1    12
5a.29    A     5b.29    B       1    13

5a.32    B     5b.32    A       2     0
```

```
5a.33    A     5b.33    B       2     1
5a.34    A     5b.34    B       2     2
...
5a.43    A     5b.43    B       2    11
5a.44    B     5b.44    A       2    12
5a.45    A     5b.45    B       2    13

8a.48    B     8b.48    A       3     0
8a.49    A     8b.49    B       3     1
8a.50    B     8b.50    A       3     2
...
8a.59    A     8b.59    B       3    11
8a.60    B     8b.60    A       3    12
8a.61    B     8b.61    A       3    13

8a.64    B     8b.64    A       4     0
8a.65    A     8b.65    B       4     1
8a.66    A     8b.66    B       4     2
...
8a.75    A     8b.75    B       4    11
8a.76    A     8b.76    B       4    12
8a.77    B     8b.77    A       4    13
```

Example 3: In the following example, system_3 is configured for Multipath I/O with SyncMirror using software-based disk ownership, as described at "Multipath I/O with SyncMirror using software-based disk ownership" on page 73.

```
system_3> storage show disk -p
PRIMARY PORT  SECONDARY  PORT  SHELF BAY
------- ----  ---------- ----  ----- ---
5a.16    A     8b.16    B       1     0
5a.17    A     8b.17    B       1     1
5a.18    B     8b.18    A       1     2
5a.19    A     8b.19    B       1     3
5a.20    A     8b.20    B       1     4
5a.21    B     8b.21    A       1     5
5a.22    A     8b.22    B       1     6
5a.23    A     8b.23    B       1     7
5a.24    B     8b.24    A       1     8
5a.25    B     8b.25    A       1     9
5a.26    A     8b.26    B       1    10
5a.27    A     8b.27    B       1    11
5a.28    B     8b.28    A       1    12
5a.29    A     8b.29    B       1    13
```

```
5a.32    B     8b.32    A      2     0
5a.33    A     8b.33    B      2     1
5a.34    A     8b.34    B      2     2
...
5a.43    A     8b.43    B      2    11
5a.44    B     8b.44    A      2    12
5a.45    A     8b.45    B      2    13

8a.48    B     5b.48    A      3     0
8a.49    A     5b.49    B      3     1
8a.50    B     5b.50    A      3     2
...
8a.59    A     5b.59    B      3    11
8a.60    B     5b.60    A      3    12
8a.61    B     5b.61    A      3    13

8a.64    B     5b.64    A      4     0
8a.65    A     5b.65    B      4     1
8a.66    A     5b.66    B      4     2
...
8a.75    A     5b.75    B      4    11
8a.76    A     5b.76    B      4    12
8a.77    B     5b.77    A      4    13
```

Using FilerView

You can also use FilerView to display information about disks, as described in the following table.

To display information about...	Open FilerView and go to...
How many disks are on a storage system	Filer > Show Status **Result:** The following information is displayed: total number of disks, the number of spares, and the number of disks that have failed.

To display information about...	Open FilerView and go to...
All disks, spare disks, broken disks, zeroing disks, and reconstructing disks	Storage > Disks > Manage, and select the type of disk from the pull-down list **Result:** The following information about disks is displayed: Disk ID, type (parity, data, dparity, spare, and partner), checksum type, shelf and bay location, channel, size, physical size, pool, and aggregate.

Managing available space on new disks

Displaying free disk space

You use the `df` command to display the amount of free disk space in the specified volume or aggregate or all volumes and aggregates (shown as Filesystem in the command output) on the storage system. This command displays the size in 1,024-byte blocks, unless you specify another value, using one of the following options: `-h` (causes Data ONTAP to scale to the appropriate size), `-k` (kilobytes), `-m` (megabytes), `-g` (gigabytes), or `-t` (terabytes).

On a separate line, the `df` command also displays statistics about how much space is consumed by the snapshots for each volume or aggregate. Blocks that are referenced by both the active file system and by one or more snapshots are counted only in the active file system, not in the snapshot line.

Disk space report discrepancies

The total amount of disk space shown in the `df` output is less than the sum of available space on all disks installed in an aggregate.

In the following example, the `df` command is issued on a traditional volume with three 72-GB disks installed, with RAID-DP enabled, and the following data is displayed:

```
toaster> df /vol/vol0

Filesystem            kybtes    used      avail  capacity  Mounted on
/vol/vol0           67108864  382296   66726568        1%   /vol/vol0
/vol/vol0/.snapshot 16777216   14740   16762476        0%  /vol/vol0/
.snapshot
```

When you add the numbers in the kbytes column, the sum is significantly less than the total disk space installed. The following behavior accounts for the discrepancy:

- The two parity disks, which are 72-GB disks in this example, are not reflected in the output of the `df` command.
- The storage system reserves 10 percent of the total disk space for efficiency, which `df` does not count as part of the file system space.

Note

The second line of output indicates how much space is allocated to snapshots. Snapshot reserve, if activated, can also cause discrepancies in the disk space report. For more information, see the *Data Protection Online Backup and Recovery Guide*.

Displaying the number of hot spare disks with the Data ONTAP CLI

To ascertain how many hot spare disks you have on your storage system using the Data ONTAP CLI, complete the following step.

Step	Action

1	Enter the following command:
	`aggr status -s`
	Result: If there are hot spare disks, a display like the following appears, with a line for each spare disk, grouped by checksum type:
	`Pool1 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 traditonal volumes or aggregates`
	`spare  9a.24  9a  1  8  FC:A  1  FCAL  10000  34000/69532000  34190/70022840`
	`spare  9a.29  9a  1  13  FC:A  1  FCAL  10000  34000/69532000  34190/70022840`
	`Pool0 spare disks (empty)`

Displaying the number of hot spare disks with FilerView

To ascertain how many hot spare disks you have on your storage system using FilerView, complete the following steps.

Step	Action
1	Open a browser and point to FilerView (for instructions on how to do this, see the chapter on accessing the storage system in the *System Administration Guide*).
2	Click the button to the left of FilerView to view a summary of system status, including the number of disks, and the number of spare and failed disks.

Adding disks

Considerations when adding disks to a storage system

The number of disks that are initially configured in RAID groups affects read and write performance. A greater number of disks means a greater number of independently seeking disk-drive heads reading data, which improves performance. Write performance can also benefit from more disks; however, the difference can be masked by the effect of nonvolatile RAM (NVRAM) and the manner in which WAFL manages write operations.

As more disks are configured, the performance increase levels off. Performance is affected more with each new disk you add until the striping across all the disks levels out. When the striping levels out, there is an increase in the number of operations per second and a reduced response time.

For overall improved performance, add enough disks for a complete RAID group. The default RAID group size is storage system-specific.

When you add disks to a storage system that is a target in a SAN environment, you should also perform a full reallocation scan. For more information, see your *Block Access Management Guide*.

Reasons to add disks

You add disks for the following reasons:

- You want to add storage capacity to the storage system to meet current or future storage requirements
- You are running out of hot spare disks
- You want to replace one or more disks

Meeting storage requirements: To meet current storage requirements, add disks before a file system is 80 percent to 90 percent full.

To meet future storage requirements, add disks before the applied load places stress on the existing array of disks, even though adding more disks at this time will not significantly improve the storage system's current performance immediately.

Running out of hot spare disks: You should periodically check the number of hot spares you have in your storage system. If there are none, then add disks to the disk shelves so they become available as hot spares. For more information, see "Hot spare disks" on page 139.

Replacing one or more disks: You might want to replace a disk because it has failed or has been put out-of-service. You might also want to replace a number of disks with ones that have more capacity or have a higher RPM.

Prerequisites for adding new disks

Before adding new disks to the storage system, be sure that the storage system supports the type of disk you want to add. For the latest information on supported disk drives, see the *Data ONTAP Release Notes* and the *System Configuration Guide* on the NOW site (http://now.netapp.com/).

Note

You should always add disks of the same size, the same checksum type, preferably block checksum, and the same RPM.

How Data ONTAP recognizes new disks

When the disks are installed, they become hot-swappable spare disks, which means they can be replaced while the storage system and shelves remain powered on.

Once the disks are recognized by Data ONTAP, you, or Data ONTAP, can add the disks to a RAID group in an aggregate with the `aggr add` command. For backward compatibility, you can also use the `vol add` command to add disks to the RAID group in the aggregate that contains a traditional volume.

Physically adding disks to the storage system

When you add disks to a storage system, you need to insert them in a disk shelf according to the instructions in the disk shelf manufacturer's documentation or the disk shelf guide provided by NetApp. For detailed instructions about adding disks or determining the location of a disk in a disk shelf, see your disk shelf documentation or the hardware and service guide for your storage system.

To add new disks to the storage system, complete the following steps.

<table>
<tr><th>Step</th><th colspan="2">Action</th></tr>
<tr><td rowspan="3">1</td><td>If the disks are...</td><td>Then...</td></tr>
<tr><td>Native Fibre Channel disks in Fibre Channel-attached shelves, or ATA disks on Fibre Channel-attached shelves</td><td>Go to Step 2.</td></tr>
<tr><td>Native SCSI disks or ATA disks in SCSI-attached shelves</td><td>Enter the following command, and go to Step 2:
<code>disk swap</code></td></tr>
<tr><td>2</td><td colspan="2">Install one or more disks according to the hardware guide for your disk shelf or the specific hardware and service guide for your storage system.

Note
On FAS270 and FAS270c storage systems or storage systems licensed for SnapMover, a disk ownership assignment might need to be carried out. For more informaiton, see “Software-based disk ownership” on page 58.

Result: The storage system displays a message confirming that one or more disks were installed and then waits 15 seconds as the disks are recognized. The storage system recognizes the disks as hot spare disks.

Note
If you add multiple disks, the storage system might require 25 to 40 seconds to bring the disks up to speed as it checks the device addresses on each adapter.</td></tr>
<tr><td>3</td><td colspan="2">Verify that the disks were added by entering the following command:
<code>aggr status -s</code>

Result: The number of hot spare disks in the RAID Disk column under Spare Disks increases by the number of disks you installed.</td></tr>
</table>

Removing disks

Reasons to remove disks

You remove a disk for the following reasons:

- You want to replace the disk because
  - It is a failed disk. You cannot use this disk again.
  - It is a data disk that is producing excessive error messages, and is likely to fail. You cannot use this disk again.
  - It is an old disk with low capacity or slow RPMs and you are upgrading your storage system.
- You want to reuse the disk. It is a hot spare disk in good working condition, but you want to use it elsewhere.

Note

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 file system that has fewer data disks.

Removing a failed disk

To remove a failed disk, complete the following steps.

Step	Action
1	Find the disk ID of the failed disk by entering the following command: `aggr status -f` **Result:** The ID of the failed disk is shown next to the word *failed*. The location of the disk is shown to the right of the disk ID, in the column HA SHELF BAY.

Step	Action	
2	**If the disk is a...**	**Then...**
	Fibre Channel disk or in a Fibre Channel-attached shelf	Go to Step 3.
	SCSI disk or in a SCSI-attached shelf	Enter the following command and go to Step 3: `disk swap`
3	Remove the disk from the disk shelf according to the disk shelf manufacturer's instructions.	

Removing a hot spare disk

To remove a hot spare disk, complete the following steps.

Step	Action	
1	Find the disk IDs of hot spare disks by entering the following command: `aggr status -s` **Result:** 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 following command to spin down the disk: `disk remove disk_name` *disk_name* is the name of the disk you want to remove (from the output of Step 1).	
3	**If the disk is...**	**Then...**
	A Fibre Channel disk or in a Fibre Channel-attached shelf	Go to Step 4.
	A SCSI disk or in a SCSI-attached shelf	Enter the following command, and go to Step 4: `disk swap`

Step	Action
4	Wait for the disk to stop spinning. See the hardware guide for your disk shelf model for information about how to tell when a disk stops spinning.
5	Remove the disk from the disk shelf, following the instructions in the hardware guide for your disk shelf model. **Result:** When replacing FC disks, there is no service interruption. When replacing SCSI and ATA disks, file service resumes 15 seconds after you remove the disk.

Removing a data disk

To remove a data disk, complete the following steps.

Step	Action
1	Find the disk name in the log messages that report disk errors by looking at the numbers that follow the word *Disk*.
2	Enter the following command: **`aggr status -r`**
3	Look at the Device column of the output of the `sysconfig -r` command. It shows the disk ID of each disk. The location of the disk appears to the right of the disk ID, in the column HA SHELF BAY.
4	Enter the following command to fail the disk: **`disk fail [-i]`** ***`disk_name`*** `-i` specifies to fail the disk immediately. *disk_name* is the disk name from the output in Step 1.

<table>
<tr><th>Step</th><th colspan="2">Action</th></tr>
<tr><td></td><td>If you...</td><td>Then...</td></tr>
<tr><td></td><td>Do not specify the <code>-i</code> option</td><td>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.

This copy might take several hours, depending on the size of the disk and the load on the storage system.

Attention
You must wait for the disk copy to complete before going to the next step.

If the copy operation is successful, then the pre-failed disk is failed and the new replacement disk takes its place.</td></tr>
<tr><td></td><td>Specify the <code>-i</code> option or if the disk copy operation fails</td><td>The pre-failed disk fails and the storage system operates in degraded mode until the RAID system reconstructs a replacement disk.</td></tr>
<tr><td>5</td><td>If the disk is...</td><td>Then...</td></tr>
<tr><td></td><td>A Fibre Channel disk or in a Fibre Channel-attached shelf</td><td>Go to Step 6.</td></tr>
<tr><td></td><td>A SCSI disk or in a SCSI-attached shelf</td><td>Enter the following command, then go to Step 6:
<code>disk swap</code></td></tr>
</table>

Step	Action
6	Remove the failed disk from the disk shelf, following the instructions in the hardware guide for your disk shelf model. **Result:** File service resumes 15 seconds after you remove the disk.

Cancelling a disk swap command

To cancel the swap operation and continue service, complete the following step.

Step	Action
1	Enter the following command: `disk unswap`

Sanitizing disks

About disk sanitization

Disk sanitization is the process of physically obliterating data by overwriting disks with specified byte patterns or random data in a manner that prevents recovery of the original data by any known recovery methods. You sanitize disks if you want to ensure that data currently on those disks is physically unrecoverable. For example, you might have some disks that you intend to remove from one storage system and you want to re-use those disks in another storage system or simply dispose of the disks. In either case, you want to ensure no one can retrieve any data from those disks.

The Data ONTAP `disk sanitize` command enables you to carry out disk sanitization by using three successive default or user-specified byte overwrite patterns for up to seven cycles per operation. You can start, stop, and display the status of the disk sanitization process, which runs in the background. Depending on the capacity of the disk and the number of patterns and cycles specified, this process can take several hours to complete. When the process has completed, the disk is in a sanitized state. You can return a sanitized disk to the spare disk pool with the `disk sanitize release` command.

What this section covers

The section covers the following topics:

- "Disk sanitization limitations" on page 105
- "Licensing disk sanitization" on page 106
- "Sanitizing spare disks" on page 107
- "Stopping disk sanitization" on page 110
- "Selectively sanitizing data" on page 110
- "Reading disk sanitization log files" on page 115

Disk sanitization limitations

The following list describes the limitations of disk sanitization operations. Disk sanitization

- Is not supported on older disks.

  To determine if disk sanitization is supported on a specified disk, run the `storage show disk` command. If the vendor for the disk in question is listed as NETAPP, disk sanitization is supported.

- Is not supported on V-Series systems.
- Is not supported in takeover mode on clustered storage systems. (If a storage system is disabled, it remains disabled during the disk sanitization process.)
- Cannot be carried out on disks that were failed due to readability or writability problems.
- Cannot be carried out 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.
- Cannot perform the formatting phase of the disk sanitization process on ATA drives.
- Cannot be carried out on more than one SES drive at a time.

Licensing disk sanitization

Before you can use the disk sanitization feature, you must install the disk sanitization license.

Attention

Once installed on a storage system, the license for disk sanitization is permanent.

Attention

The disk sanitization license prohibits the following admin command from being used on the storage system:

- `dd` (to copy blocks of data)

Attention

The disk sanitization license prohibits the following diagnostic commands from being used on the storage system:

- `dumpblock` (to print dumps of disk blocks)
- `setflag wafl_metadata_visible` (to allow access to internal WAFL files)

To install the disk sanitization license, complete the following step:

Step	Action
1	Enter the following command: **`license add`** ***`disk_sanitize_code`*** *disk_sanitize_code* is the disk sanitization license code that NetApp provides.

Sanitizing spare disks

You can sanitize any disk that has *spare* status. This includes disks that exist in the storage system as spare disks after the aggregate that they belong to has been destroyed. It also includes disks that were removed from the spare disk pool by the `disk remove` command but have been returned to spare status after a system reboot.

To sanitize a disk or a set of disks in a storage system, complete the following steps.

Step	Action
1	Print a list of all disks assigned to RAID groups, failed, or existing as spares, by entering the following command. **`sysconfig -r`** Do this to verify that the disk or disks that you want to sanitize do not belong to any existing RAID group in any existing aggregate.
2	Enter the following command to sanitize the specified disk or disks of all existing data. **`disk sanitize start [-p pattern1|-r [-p pattern2|-r] [-p pattern|-r]]] [-c cycle_count] disk_list`** `-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 hex pattern specification is `-p 0x55 -p 0xAA -p 0x3c`. `-r` replaces a patterned overwrite with a random overwrite for any or all of the cycles, for example: `-p 0x55 -p 0xAA -r` `-c` *cycle_count* specifies the number of cycles for applying the specified overwrite patterns. The default value is one cycle. The maximum value is seven cycles. **Note** To be in compliance with United States Department of Defense and Department of Energy security requirements, you must set *cycle_count* to six cycles per operation. *disk_list* specifies a space-separated list of spare disks to be sanitized.

Step	Action
	Example: The following command applies the default three disk sanitization overwrite patterns for one cycle (for a total of 3 overwrites) to the specified disks, 7.6, 7.7, and 7.8. **`disk sanitize start 7.6 7.7 7.8`** If you set *cycle_count* to 6, this example would result in three disk sanitization overwrite patterns for six cycles (for a total of 18 overwrites) to the specified disks.
	Result: The specified disks are sanitized, put into the pool of broken disks, and marked as *sanitized*. A list of all the sanitized disks is stored in the storage system's `/etc` directory. **Note** If you need to abort the sanitization operation, enter **`disk sanitize abort [`*`disk_list`*`]`** If the sanitization operation is in the process of formatting the disk, the abort will wait until the format is complete. The larger the drive, the more time this process takes to complete. **Attention** Do not turn off the storage system, disrupt the disk loop, or remove target disks during the sanitization process. If the sanitization process is disrupted, the target disks that are in the formatting stage of disk sanitization will require reformatting before their sanitization can be completed. See "If formatting is interrupted" on page 110.
3	To check the status of the disk sanitization process, enter the following command: **`disk sanitize status [`*`disk_list`*`]`**

Step	Action
4	To release sanitized disks from the pool of broken disks for reuse as spare disks, enter the following command: **`disk sanitize release` *`disk_list`*** **Attention** The `disk sanitize release` command removes the *sanitized* label from the affected disks and returns them to spare state. Rebooting the storage system or removing the disk also removes the sanitized label from any sanitized disks and returns them to spare state. **Verification:** To list all disks in the storage system and verify the release of the sanitized disks into the pool of spares, enter **`sysconfig -r.`**

Process description: After you enter the `disk sanitize start` command, Data ONTAP begins the sanitization process on each of the specified disks. The process consists of a disk format operation, followed by the specified overwrite patterns repeated for the specified number of cycles.

Note
The formatting phase of the disk sanitization process is skipped on ATA disks.

The time to complete the sanitization process for each disk depends on the size of the disk, the number of patterns specified, and the number of cycles specified.

For example, the following command invokes one format overwrite pass and 18 pattern overwrite passes of disk 7.3.

```
disk sanitize start -p 0x55 -p 0xAA -p 0x37 -c 6 7.3
```

- If disk 7.3 is 36 GB and each formatting or pattern overwrite pass on it takes 15 minutes, then the total sanitization time is 19 passes times 15 minutes, or 285 minutes (4.75 hours).
- If disk 7.3 is 73 GB and each formatting or pattern overwrite pass on it takes 30 minutes, then total sanitization time is 19 passes times 30 minutes, or 570 minutes (9.5 hours).

If disk sanitization is interrupted: If the sanitization process is interrupted by power failure, system panic, or a user-invoked `disk sanitize abort` command, the `disk sanitize` command must be re-invoked and the process repeated from the beginning in order for the sanitization to take place.

If formatting is interrupted: If the formatting phase of disk sanitization is interrupted, Data ONTAP attempts to reformat any disks that were corrupted by an interruption of the formatting. 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 such a disk is found, Data ONTAP attempts to reformat that disk, and writes a message to the console informing you that a corrupted disk has been found and will be reformatted. After the disk is reformatted, it is returned to the hot spare pool. You can then rerun the `disk sanitize` command on that disk.

Stopping disk sanitization

You can use the `disk sanitize abort` command to stop an ongoing sanitization process on one or more specified disks. If you use the `disk sanitize abort` command, the specified disk or disks are returned to spare state and the sanitized label is removed. To stop a disk sanitization process, complete the following step.

Step	Action
1	Enter the following command: **`disk sanitize abort`** ***`disklist`*** **Result:** Data ONTAP displays the message "Sanitization abort initiated." If the specified disks are undergoing the disk formatting phase of sanitization, the abort will not occur until the disk formatting is complete. Once the process is stopped, Data ONTAP displays the message "Sanitization aborted for *diskname*."

Selectively sanitizing data

Selective data sanitization consists of physically obliterating data in specified blocks while preserving all other data located on the affected aggregate for continued user access.

Summary of the selective sanitization process: Because data for any one file in a storage system is physically stored on any number of data disks in the aggregate containing that data, and because the physical location of data within an aggregate can change, sanitization of selected data, such as files or directories, requires that you sanitize every disk in the aggregate where the data is

located (after first migrating the aggregate data that you do *not* want to sanitize to disks on another aggregate). To selectively sanitize data contained in an aggregate, you must carry out three general tasks.

1. Delete the selected files or directories (and any aggregate snapshots that contain those files or directories) from the aggregate that contains them.
2. Migrate the remaining data (the data that you want to preserve) in the affected aggregate to a new set of disks in a destination aggregate on the same storage system using the `ndmpcopy` command.
3. Destroy the original aggregate and sanitize all the disks that were RAID group members in that aggregate.

Requirements for selective sanitization: Successful completion of this process requires the following conditions:

- You must install a disk sanitization license on your storage system.
- You must have enough storage space on your storage system to create an additional destination aggregate to which you can migrate the data that you want to preserve from the original aggregate. This destination aggregate must have a storage capacity at least as large as that of the original aggregate.
- You must use the `ndmpcopy` command to migrate data in the affected aggregate to a new set of disks in a destination aggregate on the same storage system. For information about the `ndmpcopy` command, see the *Data Protection Online Backup and Recovery Guide*.

Aggregate size and selective sanitization: Because sanitization of any unit of data in an aggregate still requires you to carry out data migration and disk sanitization processes on that entire aggregate, you should use small aggregates to store data that requires sanitization. Use of small aggregates for storage of data requiring sanitization minimizes the time, disk space, and bandwidth that sanitization will requires.

Backup and data sanitization: Absolute sanitization of data means physical sanitization of *all* instances of aggregates containing sensitive data; it is therefore advisable to maintain your sensitive data in aggregates that are *not* regularly backed up to aggregates that also back up large amounts of nonsensitive data.

Procedure for selective sanitization: To carry out selective sanitization of data within an aggregate or a traditional volume, complete the following steps.

Step	Action
1	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 -rf /nixdir/nixfile.doc`**
2	From the NetApp storage system, enter the following commands to delete all snapshots of the aggregates and volumes (both traditional and FlexVol volumes) that contain the files or directories that you just deleted. ◆ To delete all snapshots associated with the aggregate, enter the following command: **`snap delete -a aggr_name -A`** *aggr_name* is the aggregate that contains the files or directories that you just deleted. For example: **`snap delete -a nixsrcaggr -A`** ◆ To delete all snapshots associated with the volume, enter the following command: **`snap delete -a vol_name -V`** *vol_name* is the traditional volume or FlexVol that contains the files or directories that you just deleted. For example: **`snap delete -a nixsrcvol -V`** ◆ To delete a specific snapshot for either an aggregate or a volume, enter one of the following commands: **`snap delete aggr_name snapshot_name -A`** **`snap delete vol_name snapshot_name -V`** Examples: **`snap delete nixsrcaggr nightly0 A`** **`snap delete nixsrcvol nightly0 -V`**

Step	Action
3	Enter the following command to determine the size of the aggregate from which you deleted data: **`aggr status aggr_name -b`** For backward compatibility, you can also use the following command for traditional volumes. **`vol status vol_name -b`** **Example:** **`aggr status nixsrcaggr -b`** Calculate the aggregate size in bytes by multiplying the bytes per block (block size) by the blocks per aggregate (aggregate size).
4	Enter the following command to create an aggregate to which you will migrate undeleted data. This aggregate must be of equal or greater storage capacity than the aggregate from which you just deleted file, directory, or snapshot data: **`aggr create dest_aggr ndisks`** For backward compatibility with traditional volumes, you can also enter: **`vol create dest_vol disklist`** **Example:** **`aggr create nixdestaggr 8@72G`** **Note** The purpose of this new aggregate is to provide a migration destination that is absolutely free of the data that you want to sanitize.

Step	Action
5	Enter the following command to copy the data you want to preserve to the destination aggregate from the source aggregate you want to sanitize: **ndmpcopy *src_aggr dest_aggr*** *src_aggr* is the source aggregate. *dest_aggr* is the destination aggregate. **Attention** Be sure that you have deleted the files or directories that you want to sanitize (and any affected snapshots) from the source aggregate *before* you run the `ndmpcopy` command. **Example: ndmpcopy nixsrcvol nixdestvol**
6	Record the disks currently in the source aggregate. (After that aggregate is destroyed, you will sanitize these disks.) To list the disks in the source aggregate, enter the following command: **aggr status *src_aggr* -r** **Example: aggr status nixsrcaggr -r** The disks that you are going to sanitize are listed in the Device column of the `aggr status -r` output.
7	In maintenance mode, enter the following command to take the source aggregate offline: **aggr offline *src_aggr*** **Example: aggr offline nixsrcaggr**
8	Enter the following command to destroy the source aggregate: **aggr destroy *src_aggr*** **Example: aggr destroy nixsrcaggr**

Step	Action
9	Enter the following command to rename the destination aggregate, giving it the name of the source aggregate that you just destroyed: **`aggr rename dest_aggr src_aggr`** **Example: `aggr rename nixdestaggr nixsrcaggr`**
10	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`** **Result:** Users who were accessing files in the original volume will continue to access those files in the renamed destination volume with no remapping of their connections required.
11	Use the `disk sanitize` command to sanitize the disks that used to belong to the source aggregate. Follow the procedure described in "Sanitizing spare disks" on page 107.

Reading disk sanitization log files

The disk sanitization process outputs two types of log files.

- One file, /etc/sanitized_disks, lists all the drives that have been sanitized.
- For each disk being sanitized, a file is created where the progress information will be written.

Listing the sanitized disks: The /etc/sanitized_disks file contains the serial numbers of all drives that have been successfully sanitized. For every invocation of the `disk sanitize start` command, the serial numbers of the newly sanitized disks are appended to the file.

The /etc/sanitized_disks file shows output similar to the following:

```
admin1> rdfile /etc/sanitized_disks
Tue Jun 24 02:54:11 Disk 8a.44 [S/N 3FP0RFAZ00002218446B]
sanitized.
Tue Jun 24 02:54:15 Disk 8a.43 [S/N 3FP20XX400007313LSA8]
sanitized.
```

```
Tue Jun 24 02:54:20 Disk 8a.45 [S/N 3FP0RJMR0000221844GP]
sanitized.
Tue Jun 24 03:22:41 Disk 8a.32 [S/N 43208987] sanitized.
```

Reviewing the disk sanitization progress: A progress file is created for each drive sanitized and the results are consolidated to the /etc/sanitization.log file every 15 minutes during the sanitization operation. Entries in the log resemble the following:

```
Tue Jun 24 02:40:10 Disk sanitization initiated on drive 8a.43 [S/N
3FP20XX400007313LSA8]
Tue Jun 24 02:40:10 Disk sanitization initiated on drive 8a.44 [S/N
3FP0RFAZ00002218446B]
Tue Jun 24 02:40:10 Disk sanitization initiated on drive 8a.45 [S/N
3FP0RJMR0000221844GP]
Tue Jun 24 02:53:55 Disk 8a.44 [S/N 3FP0RFAZ00002218446B] format
completed in 00:13:45.
Tue Jun 24 02:53:59 Disk 8a.43 [S/N 3FP20XX400007313LSA8] format
completed in 00:13:49.
Tue Jun 24 02:54:04 Disk 8a.45 [S/N 3FP0RJMR0000221844GP] format
completed in 00:13:54.
Tue Jun 24 02:54:11 Disk 8a.44 [S/N 3FP0RFAZ00002218446B] cycle 1
pattern write of 0x47 completed in 00:00:16.
Tue Jun 24 02:54:11 Disk sanitization on drive 8a.44 [S/N
3FP0RFAZ00002218446B] completed.
Tue Jun 24 02:54:15 Disk 8a.43 [S/N 3FP20XX400007313LSA8] cycle 1
pattern write of 0x47 completed in 00:00:16.
Tue Jun 24 02:54:15 Disk sanitization on drive 8a.43 [S/N
3FP20XX400007313LSA8] completed.
Tue Jun 24 02:54:20 Disk 8a.45 [S/N 3FP0RJMR0000221844GP] cycle 1
pattern write of 0x47 completed in 00:00:16.
Tue Jun 24 02:54:20 Disk sanitization on drive 8a.45 [S/N
3FP0RJMR0000221844GP] completed.
Tue Jun 24 02:58:42 Disk sanitization initiated on drive 8a.43 [S/N
3FP20XX400007313LSA8]
Tue Jun 24 03:00:09 Disk sanitization initiated on drive 8a.32 [S/N
43208987]
Tue Jun 24 03:11:25 Disk 8a.32 [S/N 43208987] cycle 1 pattern write
of 0x47 completed in 00:11:16.
Tue Jun 24 03:12:32 Disk 8a.43 [S/N 3FP20XX400007313LSA8]
sanitization aborted by user.
Tue Jun 24 03:22:41 Disk 8a.32 [S/N 43208987] cycle 2 pattern write
of 0x47 completed in 00:11:16.
Tue Jun 24 03:22:41 Disk sanitization on drive 8a.32 [S/N 43208987]
completed.
```

Disk performance and health

About monitoring disk performance and health

Data ONTAP continually monitors disks to assess their performance and health. When Data ONTAP encounters specific activities on a disk, it will take corrective action by either taking a disk offline temporarily or by taking it out of service to run further tests. When this occurs, the disk is in the maintenance center.

When Data ONTAP takes disks offline temporarily

Data ONTAP temporarily stops I/O activity to a disk and takes a disk offline when

- You update disk firmware
- ATA disks take a long time to recover from a bad media patch

While the disk is offline, Data ONTAP reads from other disks within the RAID group while writes are logged. The offline disk is brought back online after re-synchronization is complete. This process generally takes a few minutes and incurs a negligible performance impact. For ATA disks, this reduces the probability of forced disk failures due to bad media patches or transient errors because taking a disk offline provides a software-based mechanism for isolating faults in drives and for performing out-of-band error recovery.

The disk offline feature is only supported for spares and data disks within RAID-DP and mirrored-RAID4 aggregates. 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.

You view the status of disks with the `aggr status -r` or `aggr status -s` commands, as shown in the following examples. You can see what disks are offline with either option.

Note

For backward compatibility, you can also use the `vol status -r` or `vol status -s` commands.

Example 1:

```
system> aggr status -r aggrA
Aggregate aggrA (online, raid4-dp degraded) (block checksums)
Plex /aggrA/plex0 (online, normal, active)
RAID group /aggrA/plex0/rg0 (degraded)
RAID Disk Device HA SHELF BAY CHAN Pool Type RPM Used (MB/blks)
Phys (MB/blks)
```

```
--------- ------ ------------- ---- ---- ---- ----- --------------
--------------
parity 8a.20 8a 1 4 FC:A - FCAL 10000 1024/2097152 1191/2439568
data 6a.36 6a 2 4 FC:A - FCAL 10000 1024/2097152 1191/2439568
data 6a.19 6a 1 3 FC:A - FCAL 10000 1024/2097152 1191/2439568
data 8a.23 8a 1 7 FC:A - FCAL 10000 1024/2097152 1191/2439568
(offline)
```

Example 2:

```
system> aggr status -s
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 8a.24 8a 1 8 FC:A - FCAL 10000 1024/2097152 1191/2439568
spare 8a.25 8a 1 9 FC:A - FCAL 10000 1024/2097152 1191/2439568
spare 8a.26 8a 1 10 FC:A - FCAL 10000 1024/2097152 1191/2439568
(offline)
spare 8a.27 8a 1 11 FC:A - FCAL 10000 1024/2097152 1191/2439568
spare 8a.28 8a 1 12 FC:A - FCAL 10000 1024/2097152 1191/2439568
```

When Data ONTAP takes a disk out of service

When Data ONTAP detects disk errors, it takes corrective action. For example, if a disk experiences a number of errors that exceed predefined thresholds for that disk type, Data ONTAP takes one of the following actions:

- If the `disk.maint_center.spares_check` option is set to on (which it is by default) and there are two or more spares 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 there are less than two spares available, Data ONTAP does not assign the disk to the maintenance center. It simply fails the 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.main_center.spares_check` option has no affect on putting disks into the maintenance center from the command line interface.

Once the disk is in the maintenance center, it is subjected to a number of tests, depending on what type of disk it is. If the disk passes all of the tests, it is returned to the spare pool. If it is ever sent back to the maintenance center, it is automatically failed. If a disk doesn't pass the tests the first time, it is automatically failed.

Data ONTAP also informs you of these activities by sending messages to

- The console
- A log file at /etc/maintenance.log
- A binary file that is sent with weekly AutoSupport messages

This feature is controlled by the `disk.maint_center.enable` option. It is on by default.

Manually running maintenance tests

You can initiate maintenance tests on a disk by using the `disk maint start` command. The following table summarizes how to use this command.

`disk maint` parameter	**Information displayed**
`disk maint list`	Shows all of the available tests.

`disk maint` parameter	**Information displayed**

`disk maint start`
`[-t test_list]`
`[-c cycle_count]`
`[-f]`
`[-i]`
`-d` *`disk_list`*
	Starts the test.
	
	`-t` *`test_list`* specifies which tests to run. The default is all.
	
	`-c` *`count_cycle`* specifies the number of cycles the tests will run on the disk. The default is 1.
	
	`-f` suppresses the warning message and forces execution of the command without confirmation (data disks only).
	
	If this option is *not* specified, the command issues a warning message and waits for confirmation before proceeding.
	
	`-i` instructs Data ONTAP to immediately remove the disk (data disks only) from the RAID group and begin the maintenance tests. As as a result, the RAID group enters degraded mode. If a suitable spare disk is available, the contents of the removed disk will be reconstructed onto that spare disk.
	
	If this option is *not* specified, Data ONTAP marks the disk as pending. If an appropriate spare is available, it is selected for Rapid RAID Recovery, and the data disk is copied to the spare. After the copy is completed, the data disk is removed form the RAID configuration and the testing begins.
	
	`-d disk_list` specifies a list of disks to run the tests on.

`disk maint` parameter	**Information displayed**
`disk maint status [-v] [`*`disk_list`*`]`	Shows the status of the disks in the maintenance center. `-v` specifies verbose. *disk_list* specifies a list of disks in the maintenance center to display the status of. The default is all.
`disk maint abort` *`disk_list`*	Stops the tests that are running on a disk in the maintenance center. If the specified disks were ones that you initiated the test on, they are returned to the spare pool. Those that were sent to the maintenance center by Data ONTAP are failed if aborted before testing is completed.

Storage subsystem management

About managing storage subsystem components

You can perform the following tasks on storage subsystem components:

- “Viewing information” on page 123
- “Changing the state of a host adapter” on page 132

Viewing information

Commands you use to view information

You can use the `environment`, `storage show`, and `sysconfig` commands to view information about the following storage subsystem components connected to your storage system. The components you can view status about with FilerView are also noted.

- Disks (status viewable with FilerView)
- Host Adapters (status viewable with FilerView)
- Hubs (status viewable with FilerView)
- Media changer devices
- Shelves (status viewable with FilerView)
- Switches
- Switch ports
- Tape drive devices

The following table provides a brief description of the subsystem component commands. For detailed information about these commands and their options, see the na_environment(1), na_storage(1), and na_sysconfig(1) man pages on the storage system.

Note

The options `alias` and `unalias` for the `storage` command are discussed in detail in the *Data Protection Guide Tape Backup and Recovery Guide*.

Data ONTAP command	To display information about...
`environment shelf`	Environmental information for each host adapter, including SES configuration, SES path.
`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.

Data ONTAP command	To display information about...
`storage show adapter`	Host adapter attributes, including a description, firmware revision level, PCI bus width, PCI clock speed, FC node name, cacheline size, FC packet size, link data rate, SRAM parity, external GIC, state, in use, redundant.
`storage show hub`	Hub attributes: hub name, channel, loop, shelf ID, shelf UID, term switch, shelf state, ESH state, and hub activity per disk ID: loop up count, invalid 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 tape`	All tape drive devices that are installed in the system.
`storage show tape supported [-v]`	All tape drives supported. With `-v`, information about density and compressions settings is also displayed.
`sysconfig -A`	All `sysconfig` reports, including configuration errors, disk drives, media changers, RAID details, tape devices, and aggregates.
`sysconfig -m`	Tape libraries.
`sysconfig -t`	Tape drives.

Viewing information about disks and host adapters

To view information about disks and host adapters, complete the following step.

Step	Action
1	Enter the following command: **`storage show`**

Example: The following example shows information about the adapters and disks connected to the storage system tpubs-cf1:

```
tpubs-cf1> storage show
Slot:            7
Description:     Fibre Channel Host Adapter 7 (QLogic 2100 rev. 3)
Firmware Rev:    1.19.14
PCI Bus Width:   32-bit
PCI Clock Speed: 33 MHz
FC Node Name:    2:000:00e08b:006a15
Cacheline Size:  128
FC Packet Size:  512
Link Data Rate:  2 Gbit
SRAM Parity:     No
External GBIC:   No
State:           Enabled
In Use:          Yes
Redundant:       No


DISK  SHELF  BAY  SERIAL       VENDOR   MODEL      REV
----- -----  ---  ------       ------   ---------  ----
7.6     0     6   LA774453     SEAGATE  ST19171FC  FB59
7.5     0     5   LA694863     SEAGATE  ST19171FC  FB59
7.4     0     4   LA781085     SEAGATE  ST19171FC  FB59
7.3     0     3   LA773189     SEAGATE  ST19171FC  FB59
7.14    1     6   LA869459     SEAGATE  ST19171FC  FB59
7.13    1     5   LA781479     SEAGATE  ST19171FC  FB59
7.12    1     4   LA772259     SEAGATE  ST19171FC  FB59
7.11    1     3   LA783073     SEAGATE  ST19171FC  FB59
7.10    1     2   LA700702     SEAGATE  ST19171FC  FB59
7.9     1     1   LA786084     SEAGATE  ST19171FC  FB59
7.8     1     0   LA761801     SEAGATE  ST19171FC  FB59
7.2     0     2   LA708093     SEAGATE  ST19171FC  FB59
7.1     0     1   LA773443     SEAGATE  ST19171FC  FB59
7.0     0     0   LA780611     SEAGATE  ST19171FC  FB59
```

Viewing information about host adapters

To view information about host adapters, complete the following step.

Step	Action	
1	**If you want to view...**	**Then...**
	Information about all the host adapters	Enter the following command: **`storage show adapter`**
	Information about a specific host adapter	Enter the following command: **`storage show adapter`** ***`name`*** *name* is the adapter name.

Example 1: The following example shows information about all the adapters installed in the storage system tpubs-cf2:

```
tpubs-cf2> storage show adapter
Slot:            7a
Description:     Fibre Channel Host Adapter 7a (QLogic 2100 rev. 3)
Firmware Rev:    1.19.14
PCI Bus Width:   32-bit
PCI Clock Speed: 33 MHz
FC Node Name:    2:000:00e08b:00fb15
Cacheline Size:  128
FC Packet Size:  512
Link Data Rate:  2 Gbit
SRAM Parity:     No
External GBIC:   No
State:           Enabled
In Use:          Yes
Redundant:       No
Slot:            7b
Description:     Fibre Channel Host Adapter 7b (QLogic 2100 rev. 3)
Firmware Rev:    1.19.14
PCI Bus Width:   32-bit
PCI Clock Speed: 33 MHz
FC Node Name:    2:000:00e08b:006b15
Cacheline Size:  128
FC Packet Size:  512
Link Data Rate:  2 Gbit
SRAM Parity:     No
External GBIC:   No
State:           Enabled
In Use:          Yes
Redundant:       No
```

Example 2: The following example shows information about adapter 7b in the storage system tpubs-cf2:

```
tpubs-cf2> storage show adapter 7b
Slot:            7b
Description:     Fibre Channel Host Adapter 7b (QLogic 2100 rev. 3)
Firmware Rev:    1.19.14
PCI Bus Width:   32-bit
PCI Clock Speed: 33 MHz
FC Node Name:    2:000:00e08b:006b15
Cacheline Size:  128
FC Packet Size:  512
Link Data Rate:  2 Gbit
SRAM Parity:     No
External GBIC:   No
State:           Enabled
In Use:          Yes
Redundant:       No
```

Viewing information about hubs

To view information about hubs, complete the following step.

Step	Action	
1	**If you want to view...**	**Then...**
	Information about all hubs	Enter the following command: **`storage show hub`**
	Information about a specific hub	Enter the following command: **`storage show hub`** ***`name`*** *name* is the hub name.

Example: The following example shows information about hub 8a.shelf1:

```
storage show hub 8a.shelf1
Hub name:    8a.shelf1
Channel:     8a
Loop:        A
Shelf id:    1
Shelf UID:   50:05:0c:c0:02:00:12:3d
Term switch: OFF
Shelf state: ONLINE
ESH state:   OK
```

```
                  Loop  Invalid  Invalid Clock  Insert Stall Util
Disk  Disk   Port      up       CRC     Word   Delta  Count Count %
Id    Bay    State     Count    Count   Count
-------------------------------------------------------------------
[IN ]        OK   3    0        0       128    1      0     0
[ 16]  0     OK   4    0        0       128    0      0     0
[ 17]  1     OK   4    0        0       128    0      0     0
[ 18]  2     OK   4    0        0       128    0      0     0
[ 19]  3     OK   4    0        0       128    0      0     0
[ 20]  4     OK   4    0        0       128    0      0     0
[ 21]  5     OK   4    0        0       128    0      0     0
[ 22]  6     OK   4    0        0       128    0      0     0
[ 23]  7     OK   4    0        0       128    0      0     0
[ 24]  8     OK   4    0        0       128    0      0     0
[ 25]  9     OK   4    0        0       128    0      0     0
[ 26]  10    OK   4    0        0       128    0      0     0
[ 27]  11    OK   4    0        0       128    0      0     0
[ 28]  12    OK   4    0        0       128    0      0     0
[OUT]        OK   4    0        0       128    0      0     0

Hub name:    8b.shelf1
Channel:     8b
Loop:        B
Shelf id:    1
Shelf UID:   50:05:0c:c0:02:00:12:3d
Term switch: OFF
Shelf state: ONLINE
ESH state:   OK
                  Loop  Invalid  Invalid Clock  Insert Stall Util
Disk  Disk   Port      up       CRC     Word   Delta  Count Count %
Id    Bay    State     Count    Count   Count
-------------------------------------------------------------------
[IN ]        OK   3    0        0       128    1      0     0
[ 16]  0     OK   4    0        0       128    0      0     0
[ 17]  1     OK   4    0        0       128    0      0     0
[ 18]  2     OK   4    0        0       128    0      0     0
[ 19]  3     OK   4    0        0       128    0      0     0
[ 20]  4     OK   4    0        0       128    0      0     0
[ 21]  5     OK   4    0        0       128    0      0     0
[ 22]  6     OK   4    0        0       128    0      0     0
[ 23]  7     OK   4    0        0       128    0      0     0
[ 24]  8     OK   4    0        0       128    0      0     0
[ 25]  9     OK   4    0        0       128    0      0     0
[ 26]  10    OK   4    0        0       128    0      0     0
[ 27]  11    OK   4    0        0       128    0      0     0
[ 28]  12    OK   4    0        0       128    0      0     0
```

```
[ 29]  13     OK    4     0          0       128     0       0       0
[OUT]         OK    4     0          0       128     0       0       0
```

Note

Hub 8b.shelf1 is also listed by the `storage show hub 8a.shelf1` command in the example, because the two hubs are part of the same shelf and the disks in the shelf are dual-ported disks. Effectively, the command is showing the disks from two perspectives.

Viewing information about medium changers

To view information about medium changers attached to your storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage show mc [`*`name`*`]`** *name* is the name of the medium changer for which you want to view information. If no medium changer name is specified, information for all medium changers is displayed.

Viewing information about switches

To view information about switches attached to the storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage show switch [`*`name`*`]`** *name* is the name of the switch for which you want to view information. If no switch name is specified, information for all switches is displayed.

Viewing information about switch ports

To view information about ports for switches attached to the storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage show port [name]`** *name* is the name of the port for which you want to view information. If no port name is specified, information for all ports is displayed.

Viewing information about tape drives

To view information about tape drives attached to your storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage show tape [tape]`** *tape* is the name of the tape drive for which you want to view information. If no tape name is specified, information for all tape drives is displayed.

Viewing supported tape drives

To view information about tape drives that are supported by your storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage show tape supported [-v]`** `-v` displays all information about supported tape drives, including their density and compression settings. If no option is given, only the names of supported tape drives are displayed.

Viewing tape drive statistics

To view storage statistics for tape drives attached to the storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage stats tape`** ***`name`*** *name* is the name of the tape drive for which you want to view storage statistics.

Resetting tape drive statistics

To reset storage statistics for a tape drive attached to the storage system, complete the following step.

Step	Action
1	Enter the following command: **`storage stats tape zero`** ***`name`*** *name* is the name of the tape drive.

Changing the state of a host adapter

About the state of a host adapter

A host adapter can be enabled or disabled. You can change the state of an adapter by using the `storage` command.

When to change the state of an adapter

Disable: You might want to disable an adapter if

- You are replacing any of the hardware components connected to the adapter, such as cables and Gigabit Interface Converters (GBICs)
- You are replacing a malfunctioning I/O module or bad cables

You can disable an adapter only if all disks connected to it can be reached through another adapter. Consequently, SCSI adapters and adapters connected to single-attached devices cannot be disabled.

If you try to disable an adapter that is connected to disks with no redundant access paths, you will get the following error message:

```
"Some device(s) on host adapter n can only be
 accessed through this adapter; unable to disable 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.

Enable: You might want to enable a disabled adapter after you have performed maintenance.

Enabling or disabling an adapter

To enable or disable an adapter, complete the following steps.

Step	Action
1	Enter the following command to identify the name of the adapter whose state you want to change: **`storage show adapter`** **Result:** The field that is labeled "Slot" lists the adapter name.

Step	Action	
2	**If you want to...**	**Then...**
	Enable the adapter	Enter the following command: **`storage enable adapter`** ***`name`*** *name* is the adapter name.
	Disable the adapter	Enter the following command: **`storage disable adapter`** ***`name`*** *name* is the adapter name.

RAID Protection of Data 4

About this chapter

This chapter describes how to manage RAID protection on storage system aggregates. Throughout this chapter, the term aggregate refers to an aggregate that contains either FlexVol volumes or traditional volumes.

Data ONTAP uses RAID Level 4 or RAID-DP (double-parity) protection to ensure data integrity within a group of disks even if one or two of those disks fail.

Note

The RAID principles and management operations described in this chapter do not apply to V-Series systems. Data ONTAP uses RAID-0 for V-Series systems since the LUNs that they use are RAID protected by the storage subsystem.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding RAID groups" on page 136
- "Predictive disk failure and Rapid RAID Recovery" on page 144
- "Disk failure and RAID reconstruction with a hot spare disk" on page 145
- "Disk failure without a hot spare disk" on page 146
- "Replacing disks in a RAID group" on page 148
- "Setting RAID level and group size" on page 149
- "Changing the RAID level for an aggregate" on page 152
- "Changing the size of RAID groups" on page 157
- "Controlling the speed of RAID operations" on page 161
- "Automatic and manual disk scrubs" on page 166
- "Minimizing media error disruption of RAID reconstructions" on page 173
- "Viewing RAID status" on page 181

Understanding RAID groups

About RAID groups in Data ONTAP

A RAID group consists of one or more data disks, across which client data is striped and stored, plus one or two parity disks. The purpose of a RAID group is to provide parity protection from data loss across its included disks. RAID4 uses one parity disk to ensure data recoverability if one disk fails within the RAID group. RAID-DP uses two parity disks to ensure data recoverability even if two disks within the RAID group fail.

RAID group disk types

Data ONTAP assigns and makes use of four different disk types to support data storage, parity protection, and disk replacement.

Disk	Description
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).
Hot 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 functions as a hot spare disk.
Parity disk	Stores data reconstruction information within RAID groups.
dParity disk	Stores double-parity information within RAID groups, if RAID-DP is enabled.

Levels of RAID protection

Data ONTAP supports two levels of RAID protection, RAID4 and RAID-DP, which you can assign on a per-aggregate basis.

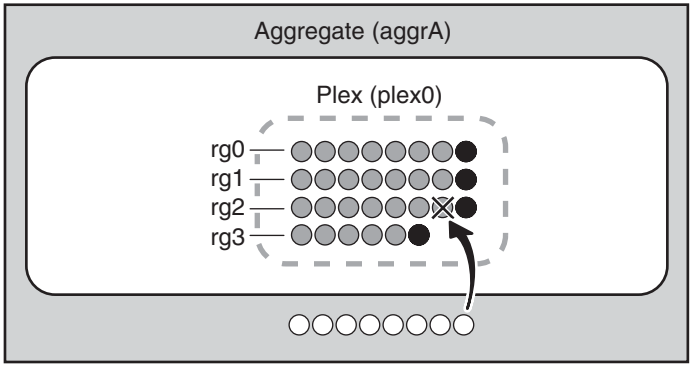
- If an aggregate is configured for RAID4 protection, Data ONTAP reconstructs the data from a single failed disk within a RAID group and transfers that reconstructed data to a spare disk.
- If an aggregate is configured for RAID-DP protection, Data ONTAP reconstructs the data from one or two failed disks within a RAID group and transfers that reconstructed data to one or two spare disks as necessary.

RAID4 protection: RAID4 provides single-parity disk protection against single-disk failure within a RAID group. 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 there are no spare disks available, Data ONTAP goes into a 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.

The following figure diagrams a traditional volume configured for RAID4 protection.



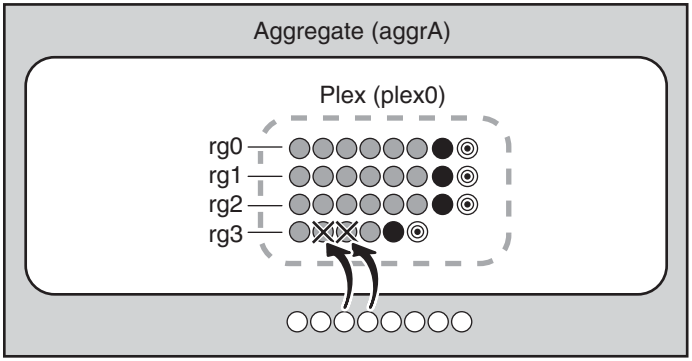
RAID-DP protection: RAID-DP provides double-parity disk protection when the following conditions occur:

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

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 (or dParity) disk.

If there is a data-disk 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. The following figure diagrams a traditional volume configured for RAID-DP protection.



How Data ONTAP organizes RAID groups automatically

When you create an aggregate or add disks to an aggregate, Data ONTAP creates new RAID groups as each RAID group is filled with its maximum number of disks. Within each aggregate, RAID groups are named rg0, rg1, rg2, and so on in order of their creation. The last RAID group formed might contain fewer disks than are specified for the aggregate's RAID group size. In that case, any disks added to the aggregate are also added to the last RAID group until the specified RAID group size is reached.

- If an aggregate is configured for RAID4 protection, Data ONTAP assigns the role of parity disk to the largest disk in each RAID group.

  Note

  If an existing RAID4 group is assigned an additional disk that is larger than the group's existing parity disk, then Data ONTAP reassigns the new disk as parity disk for that RAID group. If all disks are of equal size, any one of the disks can be selected for parity.

- If an aggregate is configured for RAID-DP protection, Data ONTAP assigns the role of dParity disk and regular parity disk to the largest and second largest disk in the RAID group.

Note

If an existing RAID-DP group is assigned an additional disk that is larger than the group's existing dParity disk, then Data ONTAP reassigns the new disk as the regular parity disk for that RAID group and restricts its capacity to be no greater than that of the existing dParity disk. If all disks are of equal size, any one of the disks can be selected for the dParity disk.

Hot spare disks

A hot spare disk is a disk that has not been assigned to a RAID group. It does not yet hold data but is ready for use. In the event of disk failure within a RAID group, Data ONTAP automatically assigns hot spare disks to RAID groups to replace the failed disks. Hot spare disks do not have to be in the same disk shelf as other disks of a RAID group to be available to a RAID group.

Hot spare disk size recommendations: Keep at least one hot spare disk for each disk size and disk type installed in your storage system. This allows the storage system to use a disk of the same size and type as a failed disk when reconstructing a failed disk. If a disk fails and a hot spare disk of the same size is not available, the storage system uses a spare disk of the next available size up. See "Disk failure and RAID reconstruction with a hot spare disk" on page 145 for more information.

Note

If no spare disks exist in a storage system, Data ONTAP can continue to function in degraded mode. Data ONTAP supports degraded mode in the case of single-disk failure for aggregates configured with RAID4 protection and in the case of single- or double- disk failure in aggregates configured for RAID-DP protection. For details see "Disk failure without a hot spare disk" on page 146.

Maximum number of RAID groups

Data ONTAP supports up to 400 RAID groups per storage system or cluster. When configuring your aggregates, keep in mind that each aggregate requires at least one RAID group and that the total of all RAID groups in a storage system cannot exceed 400.

RAID4, RAID-DP, and SyncMirror

RAID4 and RAID-DP can be used in combination with the Data ONTAP SyncMirror feature, which also offers protection against data loss due to disk 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 disk failure in one plex is repaired by the undamaged data in the opposite plex. The advantages and disadvantages of using RAID4 or RAID-DP, with and without the SyncMirror feature, are listed in the following tables.

Advantages and disadvantages of using RAID4:

Factor affected by RAID level	RAID4	RAID4 with SyncMirror
What RAID and SyncMirror protect against	Single-disk failure within one or multiple RAID groups	Single-disk failure within one or multiple RAID groups in one plex and single-, double-, or greater-disk failure in the other plex. A double-disk failure in a RAID group results in a failed plex. If this occurs, a double-disk failure on any RAID group on the other plex fails the aggregate. See "Advantages of RAID4 with SyncMirror" on page 141. Storage subsystem failures (HBA, cables, shelf) on the storage system
Required disk resources per RAID group	*n* data disks + 1 parity disk	2 x (*n* data disks + 1 parity disk)
Performance cost	None	Low mirroring overhead; can improve performance
Additional cost and complexity	None	SyncMirror license and configuration Possible cluster license and configuration

Advantages and disadvantages of using RAID-DP:

Factor affected by RAID level	RAID-DP	RAID-DP with SyncMirror
What RAID and SyncMirror protect against	Single- or double-disk failure within one or multiple RAID groups	Single-disk failure and media errors on another disk. Single- or double-disk failure within one or multiple RAID groups in one plex and single-, double-, or greater disk failure in the other plex. SyncMirror and RAID-DP together cannot protect against more than two disk failures on both plexes. It can protect against more than two disk failures on one plex with up to two disk failures on the second plex. A triple disk failure in a RAID group results in a failed plex. If this occurs, a triple disk failure on any RAID group on the other plex will fail the aggregate. See "Advantages of RAID-DP with SyncMirror" on page 142. Storage subsystem failures (HBA, cables, shelf) on the storage system
Required disk resources per RAID group	*n* data disks + 2 parity disks	2 x (*n* data disks + 2 parity disks)
Performance cost	Almost none	Low mirroring overhead; can improve performance
Additional cost and complexity	None	SyncMirror license and configuration Possible cluster license and configuration

Advantages of RAID4 with SyncMirror: On SyncMirror-replicated aggregates using RAID4, any combination of multiple disk failures within single RAID groups in one plex is restorable, as long as multiple disk failures are not concurrently occurring in the opposite plex of the mirrored aggregate.

Advantages of RAID-DP with SyncMirror: On SyncMirror-replicated aggregates using RAID-DP, any combination of multiple disk failures within single RAID groups in one plex is restorable, as long as concurrent failures of more than two disks are not occurring in the opposite plex of the mirrored aggregate.

For more SyncMirror information: For more information on the Data ONTAP SyncMirror feature, see the *Data Protection Online Backup and Recovery Guide*.

Larger versus smaller RAID groups

You can specify the number of disks in a RAID group and the RAID level of protection, or you can use the default for the specific storage system. Adding more data disks to a RAID group increases the striping of data across those disks, which typically improves I/O performance. However, with more disks, there is a greater risk that one of the disks might fail.

Configuring an optimum RAID group size for an aggregate 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. For a list of default and maximum RAID group sizes, see "Maximum and default RAID group sizes" on page 157.

Advantages of large RAID groups: Large RAID group configurations offer the following advantages:

- More data drives available. An aggregate configured into a few large RAID groups requires fewer drives reserved for parity than that same aggregate configured into many small RAID groups.
- Small improvement in storage system performance. Write operations are generally faster with larger RAID groups than with smaller RAID groups.

Advantages of small RAID groups: Small RAID group configurations offer the following advantages:

- Shorter disk reconstruction times. In case of disk failure within a small RAID group, data reconstruction time is usually shorter than it would be within a large RAID group.
- Decreased risk of data loss due to multiple disk failures. The probability of data loss through double-disk failure within a RAID4 group or through triple-disk failure within a RAID-DP group is lower within a small RAID group than within a large RAID group.

For example, whether you have a RAID group with fourteen disks or two RAID groups with seven disks, you still have the same number of disks available for striping. However, with multiple smaller RAID groups, you minimize the risk of the performance impact during reconstruction and you minimize the risk of multiple disk failure within each RAID group.

Advantages of RAID-DP over RAID4

With RAID-DP, you can use larger RAID groups because they offer more protection. A RAID-DP group is more reliable than a RAID4 group that is half its size, even though a RAID-DP group has twice as many disks. Thus, the RAID-DP group provides better reliability with the same parity overhead.

Predictive disk failure and Rapid RAID Recovery

How Data ONTAP handles failing disks

Data ONTAP monitors disk performance so that when certain conditions occur, it can predict that a disk is likely to fail. For example, under some circumstances, if 100 or more media errors occur on a disk in a one-week period, Data ONTAP identifies this disk as a possible risk. When this occurs, Data ONTAP implements a process called Rapid RAID Recovery, and performs the following tasks:

1. Places the disk in question in pre-fail mode. This can occur at any time, regardless of what state the RAID group containing the disk is in.
2. Swaps in the spare replacement disk.
3. Copies the pre-failed disk's contents to a hot spare disk on the storage system before an actual failure occurs.
4. Once the copy is complete, fails the disk that is in pre-fail mode.

Steps 2 through 4 can only occur when the RAID group is in a normal state.

By executing a copy, fail, and disk swap operation on a disk that is predicted to fail, Data ONTAP avoids three problems that a sudden disk failure and subsequent RAID reconstruction process involves:

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

If the disk that is in pre-fail mode fails on its own before copying to a hot spare disk is complete, Data ONTAP starts the normal RAID reconstruction process.

Disk failure and RAID reconstruction with a hot spare disk

About this section

This section describes how the storage system reacts to a single- or double-disk failure when a hot spare disk is available.

Data ONTAP replaces failed disk with spare and reconstructs data

If a disk fails, 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). Data ONTAP first attempts to use a hot spare disk of the same size as the failed disk. If no disk of the same size is available, Data ONTAP replaces the failed disk with a spare disk of the next available size up.
- In the background, reconstructs the missing data onto the hot spare disk or disks
- Logs the activity in the /etc/messages file on the root volume
- Sends an AutoSupport message

Note

If RAID-DP is enabled, the above processes can be carried out even in the event of simultaneous failure on two disks in a RAID group.

During reconstruction, file service might slow down.

Attention

After Data ONTAP is finished reconstructing data, 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. For information about replacing a disk, see Chapter 3, “Disk and Storage Subsystem Management,” on page 45.

If a disk fails and no hot spare disk is available, contact technical support.

You should keep at least one matching hot spare disk for each disk size and disk type installed in your storage system. This allows the storage system to use a disk of the same size and type as a failed disk when reconstructing a failed disk. If a disk fails and a hot spare disk of the same size is not available, the storage system uses a spare disk of the next available size up.

Disk failure without a hot spare disk

About this section

This section describes how the storage system reacts to a disk failure when hot spare disks are not available.

Storage system runs in degraded mode

When there is a single-disk failure in RAID4 enabled aggregates or a double-disk failure in RAID-DP enabled aggregates, and there are no hot spares available, the storage system continues to run without losing any data, but performance is somewhat degraded.

Attention

You should replace the failed disks as soon as possible, because additional disk failure might cause the storage system to lose data in the file systems contained in the affected aggregate.

Storage system logs warning messages in /etc/messages

The storage system logs a warning message in the /etc/messages file on the root volume once per hour after a disk fails.

Storage system shuts down automatically after 24 hours

To ensure that you notice the failure, the storage system automatically shuts itself off in 24 hours, by default, or at the end of a period that you set with the `raid.timeout` option of the `options` command. You can restart the storage system without fixing the disk, but it continues to shut itself off periodically until you repair the problem.

Storage system sends messages about failures

Check the /etc/messages file on the root volume once a day for important messages. You can automate checking of this file from a remote host with a script that periodically searches the file and alerts you of problems.

Alternatively, you can monitor AutoSupport messages. Data ONTAP sends AutoSupport messages when a disk fails.

Storage system reconstructs data after disk is replaced

After you replace a disk, the storage system detects the new disk immediately and uses it for reconstructing the failed disk. The storage system starts file service and reconstructs the missing data in the background to minimize service interruption.

Replacing disks in a RAID group

Replacing data disks

If you need to replace a disk—for example a mismatched data disk in a RAID group—you use the `disk replace` command. This command 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

Data ONTAP does not allow mixing disk types in the same aggregate.

To replace a disk in a RAID group, complete the following step.

Step	Action
1	Enter the following command: `disk replace start [-f] old_disk new_spare` `-f` suppresses confirmation information being displayed. It also allows a less than optimum replacement disk to be used. For example, the replacement disk might not have a matching RPM, or it might not be in the right spare pool.

Stopping the disk replacement operation

To stop the disk replace operation, or to prevent the operation if copying did not begin, complete the following step.

Step	Action
1	Enter the following command: `disk replace stop old_disk`

Setting RAID level and group size

About RAID group level and size

Data ONTAP provides default values for the RAID group level and RAID group size parameters when you create aggregates and traditional volumes. You can use these defaults or you can specify different values.

Specifying the RAID level and size when creating aggregates or FlexVol volumes

To specify the level and size of an aggregate's or traditional volume's RAID groups, complete the following steps.

Step	Action
1	View the spare disks to know which ones are available to put in a new aggregate by entering the following command: **`aggr status -s`** **Result:** The device number, shelf number, and capacity of each spare disk on the storage system is listed.
2	For an aggregate, specify RAID group level and RAID group size by entering the following command: **`aggr create`** ***`aggr`*** **`[-m] [-t {raid4|raid_dp}] [-r`** ***`raid_group_size`*** **`]`** ***`disk_list`*** *aggr* is the name of the aggregate you want to create. or For a traditional volume, specify RAID group level and RAID group size by entering the following command: **`aggr create`** ***`vol`*** **`[-v] [-m] [-t {raid4|raid_dp}] [-r`** ***`raid_group_size`*** **`]`** ***`disk_list`*** *vol* is the name of the traditional volume you want to create.

Step	Action
	`-m` specifies the optional creation of a SyncMirror-replicated volume if you want to supplement RAID protection with SyncMirror protection. A SyncMirror license is required for this feature. `-t {raid4|raid_dp}` specifies the level of RAID protection (RAID4 or RAID-DP) that you want to provide. If no RAID level is specified, the default value `raid_dp` is applied to an aggregate or the default value `raid4` is applied to a traditional volume. RAID-DP is the default for both aggregates and traditional volumes on storage systems that support ATA disks. `-r` *raid-group-size* is the number of disks per RAID group that you want. If no RAID group size is specified, the default value for your storage system model is applied. For a listing of default and maximum RAID group sizes, see "Maximum and default RAID group sizes" on page 157. *disk_list* specifies the disks to include in the volume that you want to create. It can be expressed in the following formats: ◆ *ndisks*[@*disk-size*] *ndisks* is the number of disks to use. It must be at least 2. *disk-size* is the disk size to use, in gigabytes. You must have at least *ndisks* available disks of the size you specify. ◆ `-d` *disk_name1 disk_name2... disk_nameN* *disk_name1*, *disk_name2,* and *disk_nameN* are disk IDs of one or more available disks; use a space to separate multiple disks. **Example:** The following command creates the aggregate newaggr. Since RAID-DP is the default, it does not have to be specified. RAID group size is 16 disks. Since the aggregate consists of 32 disks, those disks will form two RAID groups, rg0 and rg1: `aggr create newaggr -r 16 32@72`

Step	Action
3	(Optional) To verify the RAID structure of the aggregate that you just created, enter the appropriate command: `aggr status aggr -r` **Result:** The parity and data disks for each RAID group in the aggregate just created are listed. In aggregates and traditional volumes with RAID-DP protection, you will see parity, dParity, and data disks listed for each RAID group. In aggregates and traditional volumes with RAID4 protection, you will see parity and data disks listed for each RAID group.
4	(Optional) To verify that spare disks of sufficient number and size exist on the storage system to serve as replacement disks in event of disk failure in one of the RAID groups in the aggregate that you just created, enter the following command: `aggr status -s`

Changing the RAID level for an aggregate

Changing the RAID group level

You can change the level of RAID protection configured for an aggregate. When you change an aggregate's RAID level, Data ONTAP reconfigures all the existing RAID groups to the new level and applies the new level to all subsequently created RAID groups in that aggregate.

Changing from RAID4 to RAID-DP protection

Before you change an aggregate's RAID protection from RAID4 to RAID-DP, you need to ensure that hot spare disks of sufficient number and size are available. During the conversion, Data ONTAP adds an additional disk to each existing RAID group from the storage system's hot spare disk pool and assigns the new disk the dParity disk function for the RAID-DP group. In addition, the aggregate's `raidsize` option is changed to RAID-DP as the default on this storage system. The `raidsize` option also controls the size of new RAID groups that might be created in the aggregate.

Changing an aggregate's RAID level: To change an existing aggregate's RAID protection from RAID4 to RAID-DP, complete the following steps.

Step	Action
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	Enter the following command to make sure that a hot spare disk exists on the storage system for each RAID group listed for the aggregate in question, and make sure that these hot spare disks match the size and checksum type of the existing parity disks in those RAID groups. **`aggr status -s`** If necessary, add hot spare disks of the appropriate number of appropriate number, size, and checksum type to the storage system. See "Prerequisites for adding new disks" on page 98.

Step	Action
3	Enter the following command: **aggr options *aggr_name* raidtype raid_dp** *aggr_name* is the aggregate whose RAID level (`raidtype`) you are changing. **Example:** The following command changes the RAID level of the aggregate this aggr to RAID-DP: `aggr options thisaggr raidtype raid_dp` For backward compatibility, you can enter the following command: **vol options *vol_name* raidtype raid_dp**

Associated RAID group size changes: When you change the RAID protection level of an existing aggregate from RAID4 to RAID-DP, the following associated RAID group size changes take place:

- A second parity disk (dParity) is automatically added to each existing RAID group from the hot spare disk pool, thus increasing the size of each existing RAID group by one.

  If hot spare disks available on the storage system are of insufficient number or size to support the RAID level conversion, Data ONTAP issues a warning before executing the command to set the RAID level to RAID-DP (either `aggr options` *aggr_name* `raidtype raid_dp` or `vol options` *vol_name* `raidtype raid_dp`).

  If you continue the operation, RAID-DP protection is implemented on the aggregate in question, but some of its RAID groups for which no second parity disk was available remain degraded. In this case, the protection offered is no improvement over RAID4, and no hot spare disks are available in case of disk failure since all were reassigned as dParity disk.
- The aggregate's `raidsize` option, which sets the size for any new RAID groups created in this aggregate, is automatically reset to one of the following RAID-DP defaults:
  - On all non-NearStore storage systems, 16
  - On an R100 storage system, 12
  - On an R150 storage system, 12
  - On an R200 storage system, 14
  - On all NetApp storage systems that support ATA disks, 14

Note

After the `aggr options` *aggr_name* `raidtype raid_dp` operation is complete, you can manually change the `raidsize` option through the `aggr options` *aggr_name* `raidsize` command. See "Changing the maximum size of RAID groups" on page 158.

For backward compatibility, you can also use the following commands for traditional volumes:
`vol options` *vol_name* `raidtype raid_dp` operation
`vol options` *vol_name* `raidsize` command

Changing from RAID-DP to RAID4 protection

Changing an aggregate's RAID level: While it is possible to change an aggregate from RAID-DP to RAID4, there is a restriction, as described in the following note.

Note

You cannot change an aggregate from RAID-DP to RAID4 if the aggregate contains a RAID group larger than the maximum allowed for RAID4.

To change an existing aggregate's RAID protection from RAID-DP to RAID4, complete the following step.

Step	Action
1	Enter the following command: **`aggr options aggr_name raidtype raid4`** *aggr_name* is the aggregate whose RAID level (`raidtype`) you are changing. **Example:** The following command changes the RAID level of the aggregate thataggr to RAID4: `aggr options thataggr raidtype raid4`

Associated RAID group size changes: The RAID group size determines the size of any new RAID groups created in an aggregate. When you change the RAID protection of an existing aggregate from RAID-DP to RAID4, Data ONTAP automatically carries out the following associated RAID group size changes:

- In each of the aggregate's existing RAID groups, the RAID-DP second parity disk (dParity) is removed and placed in the hot spare disk pool, thus reducing each RAID group's size by one parity disk.
- For NearStore storage systems, Data ONTAP changes the aggregate's `raidsize` option to the RAID4 default sizes, as indicated on the following storage systems:
  - R100 (8)
  - R150 (6)
  - R200 (7)
- For non-NearStore storage systems, Data ONTAP changes the setting for the aggregate's `raidsize` option to the size of the largest RAID group in the aggregate. However, there are two exceptions:
  - If the aggregate's largest RAID group is larger than the maximum RAID4 group size on non-NearStore storage systems (14), then the aggregate's `raidsize` option is set to 14.
  - If the aggregate's largest RAID group is smaller than the default RAID4 group size on non-NearStore storage systems (8), then the aggregate's `raidsize` option is set to 8.
- For storage systems that support ATA disks, Data ONTAP changes the setting for the aggregate's `raidsize` option to 7.

Note

For storage systems that support ATA disks, the restriction about not being able to change an aggregate from RAID-DP to RAID4 if the aggregate contains a RAID group larger than the maximum allowed for RAID4 also applies to traditional volumes.

After the `aggr options` *aggr_name* `raidtype raid_dp` operation is complete, you can manually change the `raidsize` option through the `aggr options` *aggr_name* `raidsize` command. See "Changing the maximum size of RAID groups" on page 158.

For backward compatibility, you can also use the following commands for traditional volumes:

`vol options` *vol_name* `raidtype raid_dp`
`vol options` *vol_name* `raidsize`

Verifying the RAID level

To verify the RAID level of an aggregate, complete the following step.

Step	Action
1	Enter the following command: `aggr status` *`aggr_name`* or `aggr options` *`aggr_name`* For backward compatibility, you can also enter the following command: `vol options` *`vol_name`*

Changing the size of RAID groups

Maximum and default RAID group sizes

You can change the size of RAID groups that will be created on an aggregate or a traditional volume.

Maximum and default RAID group sizes vary according to the storage system and level of RAID group protection provided. The default RAID group sizes are generally recommended.

Maximum and default RAID-DP group sizes and defaults: The following table lists the minimum, maximum, and default RAID-DP group sizes supported on NetApp storage systems.

Storage system	Minimum group size	Maximum group size	Default group size
R200	3	16	14
R150	3	16	12
R100	3	12	12
Aggregates with ATA disks on other NetApp storage systems	3	16	14
All other NetApp storage systems	3	28	16

Maximum and default RAID4 group sizes and defaults: The following table lists the minimum, maximum, and default RAID4 group sizes supported on NetApp storage systems.

Storage system	Minimum group size	Maximum group size	Default group size
R200	2	7	7
R150	2	6	6
R100	2	8	8
FAS250	2	14	7

Storage system	Minimum group size	Maximum group size	Default group size
Aggregates with ATA disks on other NetApp storage systems	2	7	7
All other NetApp storage systems	2	14	8

Note

If, as a result of a software upgrade from an older version of Data ONTAP, traditional volumes exist that contain RAID4 groups larger than the maximum group size for the storage system, convert the traditional volumes in question to RAID-DP as soon as possible.

Changing the maximum size of RAID groups

The `aggr option raidsize` option specifies the maximum RAID group size that can be reached by adding disks to an aggregate. For backward compatibility, you can also use the `vol option raidsize` option when you change the `raidsize` option of a traditional volume's containing aggregate.

- You can increase the `raidsize` option to allow more disks to be added to the most recently created RAID group.
- The new `raidsize` setting also applies to subsequently created RAID groups in an aggregate. Either increasing or decreasing `raidsize` settings will apply to future RAID groups.
- You cannot decrease the size of already created RAID groups.
- Existing RAID groups remain the same size they were before the `raidsize` setting was changed.

Changing the raidsize setting: To change the `raidsize` setting for an existing aggregate, complete the following step.

Step	Action

1	Enter the following command:
	`aggr options aggr_name raidsize size`
	aggr_name is the aggregate whose `raidsize` setting you are changing.
	size is the number of disks you want in the most recently created and all future RAID groups in this aggregate.
	Example: The following command changes the `raidsize` setting of the aggregate yeraggr to 16 disks:
	`aggr options yeraggr raidsize 16`
	For backward compatibility, you can also enter the following command for traditional volumes:
	`vol options vol_name raidsize size`
	Example: The following command changes the `raidsize` setting of the traditional volume yervol to 16 disks:
	`vol options yervol raidsize 16`

For information about adding disks to existing RAID groups, see "Adding disks to aggregates" on page 198.

Verifying the raidsize setting

To verify the size of `raidsize` setting in an aggregate, enter the `aggr options` *`aggr_name`* command.

For backward compatibility, you can also enter the `vol options` *`vol_name`* command for traditional volumes.

Changing the size of existing RAID groups

If you increased the `raidsize` setting for an aggregate or a traditional volume, you can also use the `-g` *raidgroup* option in the `aggr add` command or in the `vol add` command to add disks to an existing RAID group. For information about adding disks to existing RAID groups, see "Adding disks to a specific RAID group in an aggregate" on page 201.

Controlling the speed of RAID operations

RAID operations you can control

You can control the speed of the following RAID operations with RAID options:

- RAID data reconstruction
- Disk scrubbing
- Plex resynchronization
- Synchronous mirror verification

Effects of varying the speed on storage system performance

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 FC loop on FC-based storage systems), 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 that the operation has on client access to the storage system, change the specific RAID option from `medium` (the default) to `low`. This also causes the operation to slow down.
- If you want to speed up the operation, change the RAID option from `medium` to `high`. This might decrease the performance of the storage system in response to client access.

Detailed information

The following sections discuss how to control the speed of RAID operations:

- "Controlling the speed of RAID data reconstruction" on page 162
- "Controlling the speed of disk scrubbing" on page 163
- "Controlling the speed of plex resynchronization" on page 164
- "Controlling the speed of mirror verification" on page 165

Controlling the speed of RAID data reconstruction

About RAID data reconstruction

If a disk fails, the data on it is reconstructed on a hot spare disk if one is available. Because RAID data reconstruction consumes CPU resources, increasing the speed of data reconstruction sometimes slows storage system network and disk operations.

Changing RAID data reconstruction speed

To change the speed of data reconstruction, complete the following step.

Step	Action
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—CPU time, disks, and FC loop bandwidth (on FC-based storage systems)—available for RAID data reconstruction; this setting can heavily affect storage system performance. Low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance. The default speed is medium. **Note** The setting for this option also controls the speed of Rapid RAID recovery.

RAID operations affecting RAID data reconstruction speed

When RAID data reconstruction and plex resynchronization 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 of both operations has a medium impact.

Controlling the speed of disk scrubbing

About disk scrubbing

Disk scrubbing means periodically checking the disk blocks of all disks on the storage system for media errors and parity consistency.

Although disk scrubbing slows the storage system somewhat, network clients might not notice the change in storage system performance because disk scrubbing starts automatically at 1:00 a.m. on Sunday by default, when most storage systems are lightly loaded, and stops after six hours. You can change the start time with the `scrub sched` option, and you can change the duration time with the `scrub duration` option.

Changing disk scrub speed

To change the speed of disk scrubbing, complete the following step.

Step	Action
1	Enter the following command: **`options raid.scrub.perf_impact`** ***`impact`*** *impact* can be `high`, `medium`, or `low`. (default) High means that the storage system uses most of the available system resources—CPU time, disks, and FC loop bandwidth (on FC-based storage systems)—for disk scrubbing; this setting can heavily affect storage system performance. Low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance.

RAID operations affecting disk scrub speed

When disk scrubbing and mirror verification 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 by both operations has a medium impact.

Controlling the speed of plex resynchronization

What plex resynchronization is

Plex resynchronization refers to the process of synchronizing the data of the two plexes of a mirrored aggregate. When plexes are synchronized, the data on each plex is identical. When plexes are unsynchronized, one plex contains data that is more up to date than that of the other plex. Plex resynchronization updates the out-of-date plex until both plexes are identical.

When plex resynchronization occurs

Data ONTAP resynchronizes the two plexes of a mirrored aggregate if one of the following occurs:

- One of the plexes was taken offline and then brought online later
- You add a plex to an unmirrored aggregate

Changing plex resynchronization speed

To change the speed of plex resynchronization, complete the following step.

Step	Action
1	Enter the following command: **`options raid.resync.perf_impact` *`impact`*** *impact* can be `high`, `medium (default)`, or `low`. High means that the storage system uses most of the available system resources for plex resynchronization; this setting can heavily affect storage system performance. Low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance.

RAID operations affecting plex resynchronization speed

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.

Controlling the speed of mirror verification

What mirror verification is

You use mirror verification to ensure that the two plexes of a synchronous mirrored aggregate are identical. See the synchronous mirror volume management chapter in the *Data Protection Online Backup and Recovery Guide* for more information.

Changing mirror verification speed

To change the speed of mirror verification, complete the following step.

Step	Action
1	Enter the following command: **`options raid.verify.perf_impact impact`** *impact* can be `high`, `medium`, or `low` (default). High means that the storage system uses most of the available system resources for mirror verification; this setting can heavily affect storage system performance. Low means that the storage system uses very little of the system resources; this setting lightly affects storage system performance.

RAID operations affecting mirror verification speed

When mirror verification and disk 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.

Automatic and manual disk scrubs

About disk scrubbing

Disk scrubbing means checking the disk blocks of all disks on the storage system for media errors and parity consistency. If Data ONTAP finds media errors or inconsistencies, it fixes them by reconstructing the data from other disks and rewriting the data. Disk scrubbing reduces the chance of potential data loss as a result of media errors during reconstruction.

Data ONTAP enables block checksums to ensure data integrity. If checksums are incorrect, Data ONTAP generates an error message similar to the following:

```
Scrub found checksum error on /vol/vol0/plex0/rg0/4.0 block 436964
```

If RAID4 is enabled, Data ONTAP scrubs a RAID group only when all the group's disks are operational.

If RAID-DP is enabled, Data ONTAP can carry out a scrub even if one disk in the RAID group has failed.

This section includes the following topics:

- "Scheduling an automatic disk scrub" on page 167
- "Manually running a disk scrub" on page 170

Scheduling an automatic disk scrub

About disk scrub scheduling

By default, automatic disk scrubbing is enabled for once a week and begins at 1:00 a.m. on Sunday. However, you can modify this schedule to suit your needs.

- You can reschedule automatic disk scrubbing to take place on other days, at other times, or at multiple times during the week.
- You might want to disable automatic disk scrubbing if disk scrubbing encounters a recurring problem.
- You can specify the duration of a disk scrubbing operation.
- You can start or stop a disk scrubbing operation manually.

Rescheduling disk scrubbing

If you want to reschedule the default weekly disk scrubbing time of 1:00 a.m. on Sunday, you can specify the day, time, and duration of one or more alternative disk scrubbings for the week.

To schedule weekly disk scrubbings, complete the following steps.

Step	Action
1	Enter the following command: **options raid.scrub.schedule *duration*{h|m}@*weekday*@*start_time* [,*duration*{h|m}@*weekday*@*start_time*] ...** *duration* {h|m} is the amount of time, in hours (h) or minutes (m) that you want to allot for this operation. **Note** If no duration is specified for a given scrub, the value specified in the raid.scrub.duration option is used. For details, see "Setting the duration of automatic disk scrubbing" on page 169. *weekday* is the day of the week (sun, mon, tue, wed, thu, fri, sat) when you want the operation to start. *start_time* is the hour of the day you want the scrub to start. Acceptable values are 0-23, where 0 is midnight and 23 is 11 p.m. **Example:** The following command schedules two weekly RAID scrubs. The first scrub is for 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**
2	Verify your modification with the following command: **options raid.scrub.schedule** The duration, weekday, and start times for all your scheduled disk scrubs appear. **Note** If you want to restore the default automatic scrub schedule of Sunday at 1:00 a.m., reenter the command with an empty value string as follows: **options raid.scrub.schedule " "**.

Toggling automatic disk scrubbing

To enable and disable automatic disk scrubbing for the storage system, complete the following step.

Step	Action
1	Enter the following command: **`options raid.scrub.enable off | on`** Use `on` to enable automatic disk scrubbing. Use `off` to disable automatic disk scrubbing.

Setting the duration of automatic disk scrubbing

You can set the duration of automatic disk scrubbing. The default is to perform automatic disk scrubbing for six hours (360 minutes). If scrubbing does not finish in six hours, Data ONTAP records where it stops. The next time disk scrubbing automatically starts, scrubbing starts from the point where it stopped.

To set the duration of automatic disk scrubbing, complete the following step.

Step	Action
1	Enter the following command: **`options raid.scrub.duration`** ***`duration`*** *duration* is the length of time, in minutes, that automatic disk scrubbing runs. **Note** If you set *duration* to `-1`, all automatically started disk scrubs run to completion.

Note

If an automatic disk scrubbing is scheduled through the `options raid.scrub.schedule` command, the duration specified for the `raid.scrub.duration` option applies only if no duration was specified for disk scrubbing in the `options raid.scrub.schedule` command.

Changing disk scrub speed

To change the speed of disk scrubbing, see "Controlling the speed of disk scrubbing" on page 163.

Manually running a disk scrub

About disk scrubbing and checking RAID group parity

You can manually run disk scrubbing to check RAID group parity on RAID groups at the RAID group level, plex level, or aggregate level. The parity checking function of the disk scrub compares the data disks in a RAID group to the parity disk in a RAID group. If during the parity check Data ONTAP determines that parity is incorrect, Data ONTAP corrects the parity disk contents.

At the RAID group level, you can check only RAID groups that are in an active parity state. If the RAID group is in a degraded, reconstructing, or repairing state, Data ONTAP reports errors if you try to run a manual scrub.

If you are checking an aggregate that has some RAID groups in an active parity state and some not in an active parity state, Data ONTAP checks and corrects the RAID groups in an active parity state and reports errors for the RAID groups not in an active parity state.

Running manual disk scrubs on all aggregates

To run manual disk scrubs on all aggregates, complete the following step.

Step	Action
1	Enter the following command: **`aggr scrub start`**

You can use your UNIX or CIFS host to start a disk scrubbing operation at any time. For example, you can start disk scrubbing by putting `disk scrub start` into a remote shell command in a UNIX `cron` script.

Disk scrubs on specific RAID groups

To run a manual disk scrub on the RAID groups of a specific aggregate, plex, or RAID group, complete the following step.

Step	Action
1	Enter one of the following commands: **aggr scrub start *name*** *name* is the name of the aggregate, plex, or RAID group.

Examples:

In this example, the command starts the manual disk scrub on all the RAID groups in the aggr2 aggregate:

```
aggr scrub start aggr2
```

In this example, the command starts a manual disk scrub on all the RAID groups of plex1 of the aggr2 aggregate:

```
aggr scrub start aggr2/plex1
```

In this example, the command starts a manual disk scrub on RAID group 0 of plex1 of the aggr2 aggregate:

```
aggr scrub start aggr2/plex1/rg0
```

Stopping manual disk scrubbing

You might need to stop Data ONTAP from running a manual disk scrub. If you stop a disk scrub, you can not resume it at the same location. You must start the scrub from the beginning. To stop a manual disk scrub, complete the following step.

Step	Action
1	Enter the following command: **aggr scrub stop *aggr_name*** If *aggr_name* is not specified, Data ONTAP stops all manual disk scrubbing.

Suspending a manual disk scrub

Rather than stopping Data ONTAP from checking and correcting parity, you can suspend checking for any period of time and resume it later, at the same offset when you suspended the scrub.

To suspend manual disk scrubbing, complete the following step.

Step	Action
1	Enter the following commands: **`aggr scrub suspend`** ***`aggr_name`*** If *aggr_name* is not specified, Data ONTAP suspends all manual disk scrubbing.

Resuming a suspended disk scrub

To resume manual disk scrubbing, complete the following step.

Step	Action
1	Enter the following command: **`aggr scrub resume`** ***`aggr_name`*** If *aggr_name* is not specified, Data ONTAP resumes all suspended manual disk scrubbing.

Viewing disk scrub status

The disk scrub status tells you what percentage of the disk scrubbing has been completed. Disk scrub status also displays whether disk scrubbing of a volume, plex, or RAID group is suspended.

To view the status of a disk scrub, complete the following step.

Step	Action
1	Enter one of the following commands: **`aggr scrub status`** ***`aggr_name`*** If *aggr_name* is not specified, Data ONTAP shows the disk scrub status of all RAID groups.

Minimizing media error disruption of RAID reconstructions

About minimizing media errors

A media error encountered during RAID reconstruction for a single-disk failure might cause a storage system panic or data loss. To minimize the risk of storage system disruption due to media errors, Data ONTAP provides the following features:

- Improved handling of media errors by a WAFL repair mechanism. See "Handling of media errors during RAID reconstruction" on page 174.
- Default continuous media error scrubbing on storage system disks. See "Continuous media scrub" on page 175.
- Continuous monitoring of disk media errors and automatic failing and replacement of disks that exceed system-defined media error thresholds. See "Disk media error failure thresholds" on page 180.

Handling of media errors during RAID reconstruction

About media error handling during RAID reconstruction

By default, if Data ONTAP encounters media errors during a RAID reconstruction, it automatically invokes an advanced mode process (`wafliron`) that compensates for the media errors and enables Data ONTAP to bypass the errors.

If this process is successful, RAID reconstruction continues, and the aggregate in which the error was detected remains online.

If you configure Data ONTAP so that it does not invoke this process, or if this process fails, Data ONTAP attempts to place the affected aggregate in restricted mode. If restricted mode fails, the storage system panics, and after a reboot, Data ONTAP brings up the affected aggregate in restricted mode. In this mode, you can manually invoke the `wafliron` process in advanced mode or schedule downtime for your storage system for reconciling the error by running the `WAFL_check` command from the Boot menu.

Purpose of the raid.reconstruction.wafliron.enable option

The `raid.reconstruction.wafliron.enable` option determines whether Data ONTAP automatically starts the `wafliron` process after detecting medium errors during RAID reconstruction. By default, the option is set to On.

Recommendation: Leave the `raid.reconstruction.wafliron.enable` option at its default setting of On, which might increase data availability.

Enabling and disabling the automatic wafliron process

To enable or disable the `raid.reconstruct.wafliron.enable` option, complete the following step.

Step	Action
1	Enter the following command: **`options raid.reconstruction.wafliron.enable {on | off}`**

Continuous media scrub

About continuous media scrubbing

By default, Data ONTAP runs continuous background media scrubbing for media errors on storage system disks. The purpose of the continuous media scrub is to detect and scrub media errors in order to minimize the chance of storage system disruption due to media error while a storage system is in degraded or reconstruction mode.

Negligible performance impact: Because continuous media scrubbing searches only for media errors, the impact on system performance is negligible.

Note

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. 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.

Not a substitute for a scheduled disk scrub: Because the continuous process described in this section scrubs only media errors, you should continue to run the storage system's scheduled complete disk scrub operation, which is described in "Automatic and manual disk scrubs" on page 166. The complete disk scrub carries out parity and checksum checking and repair operations, in addition to media checking and repair operations, on a scheduled rather than a continuous basis.

Adjusting maximum time for a media scrub cycle

You can 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.

By default, one cycle of a storage system's continuous media scrub can take a maximum of 72 hours to complete. In most situations, one cycle completes in a much shorter time; however, under heavy client workload conditions, the default 72-hour maximum ensures that whatever the client load is on the storage system, enough CPU resources will be allotted to the media scrub to complete one cycle in no more than 72 hours.

If you want the media scrub operation to consume even fewer CPU resources under heavy storage system client workload, you can increase the maximum number of hours for the media scrub cycle. This uses fewer CPU resources for the media scrub in times of heavy storage system usage.

To change the maximum time for a media scrub cycle, complete the following step.

Step	Action
1	Enter the following command: **`options raid.media_scrub.rate max_hrs_per_cycle`** *max_hrs_per_cycle* is the maximum number of hours that you want to allow for one cycle of the continuous media scrub. Valid options range from 72 to 336 hours.

Disabling continuous media scrubbing

You should keep continuous media error scrubbing enabled, particularly for R100 and R200 series storage systems, but you might decide to disable your continuous media scrub if your storage system is carrying out operations with heavy performance impact and if you have alternative measures (such as aggregate SyncMirror replication or RAID-DP configuration) in place that prevent data loss due to storage system disruption or double-disk failure.

To disable continuous media scrubbing, complete the following step.

Step	Action
1	Enter the following command at the Data ONTAP command line: **`options raid.media_scrub.enable off`** **Note** If you disable continuous media scrubbing, you can restart it by entering the following command: **`options raid.media_scrub.enable on`**

Checking media scrub activity

You can confirm media scrub activity on your storage system by completing the following step.

Step	Action
1	Entering one of the following commands: **`aggr media_scrub status [/aggr[/plex][/raidgroup]] [-v]`** **`aggr media_scrub status [-s spare_disk_name] [-v]`** */aggr*[*/plex*] [*/raidgroup*] is the optional pathname to the aggregate, plex, or RAID group on which you want to confirm media scrubbing activity. `-s` *disk_name* specifies the optional name of a specific spare disk on which you want to confirm media scrubbing activity. `-v` is specifies the verbose version of the media scrubbing activity status. The verbose status information includes the percentage of the current scrub that is complete, the start time of the current scrub, and the completion time of the last scrub. **Note** If you enter `aggr media_scrub` status without specifying a pathname or a disk name, the status of the current media scrubs on all RAID groups and spare disks is displayed.

Example 1. Checking of storage system-wide media scrubbing: The following command displays media scrub status information for all the aggregates and spare disks on the storage system.

```
aggr media_scrub status
aggr media_scrub /aggr0/plex0/rg0 is 0% complete
aggr media_scrub /aggr2/plex0/rg0 is 2% complete
aggr media_scrub /aggr2/plex0/rg1 is 2% complete
aggr media_scrub /aggr3/plex0/rg0 is 30% complete
aggr media_scrub 9a.8 is 31% complete
aggr media_scrub 9a.9 is 31% complete
aggr media_scrub 9a.13 is 31% complete
aggr media_scrub 9a.2 is 31% complete
aggr media_scrub 9a.12 is 31% complete
```

Example 2. Verbose checking of storage system-wide media scrubbing: The following command displays verbose media scrub status information for all the aggregates on the storage system.

```
aggr media_scrub status -v
aggr media_scrub: status of /aggr0/plex0/rg0 :
  Current instance of media_scrub is 0% complete.
  Media scrub started at Thu Mar  4 21:26:00 GMT 2004
  Last full media_scrub completed: Thu Mar  4 21:20:12 GMT 2004

aggr media_scrub: status of 9a.8 :
  Current instance of media_scrub is 31% complete.
  Media scrub started at Thu Feb 26 23:14:00 GMT 2004
  Last full media_scrub completed: Wed Mar  3 23:22:33 GMT 2004

aggr media_scrub: status of 9a.9 :
  Current instance of media_scrub is 31% complete.
  Media scrub started at Thu Feb 26 23:14:00 GMT 2004
  Last full media_scrub completed: Wed Mar  3 23:22:33 GMT 2004

aggr media_scrub: status of 9a.13 :
  Current instance of media_scrub is 31% complete.
  Media scrub started at Thu Feb 26 23:14:00 GMT 2004
  Last full media_scrub completed: Wed Mar  3 23:22:37 GM
```

Example 3. Checking for media scrubbing on a specific aggregate:

The following command displays media scrub status information for the aggregate *aggr2*.

```
aggr media_scrub status /aggr2
  aggr media_scrub /aggr2/plex0/rg0 is 4% complete
  aggr media_scrub /aggr2/plex0/rg1 is 10% complete
```

Example 4. Checking for media scrubbing on a specific spare disk:

The following commands display media scrub status information for the spare disk 9b.12.

```
aggr media_scrub status -s 9b.12
aggr media_scrub 9b.12 is 31% complete
```

```
aggr media_scrub status -s 9b.12 -v
aggr media_scrub: status of 9b.12 :
  Current instance of media_scrub is 31% complete.
  Media scrub started at Thu Feb 26 23:14:00 GMT 2004
  Last full media_scrub completed: Wed Mar  3 23:23:33 GMT 2004
```

Enabling continuous media scrubbing on disks

Data disks: Set the following system-wide default option to On to enable a continuous media scrub on its data disks that have been assigned to an aggregate:

```
options raid.media_scrub.enable
```

Spare disks: Set the following system-wide default option to On to enable a continuous media scrub on its spare disks:

```
options raid.media_scrub.enable
```

```
options raid.media_scrub.spares.enable
```

Disk media error failure thresholds

About media error thresholds

To prevent a storage system panic or data loss that might occur if too many media errors are encountered during single-disk failure reconstruction, Data ONTAP tracks media errors on each active storage system disk and sends a disk failure request to the RAID system if system-defined media error thresholds are crossed on that disk.

Disk media error thresholds that trigger an immediate disk failure request include

- More than twenty-five media errors (that are not related to disk scrub activity) occurring on a disk within a ten-minute period
- Three or more media errors occurring on the same sector of a disk

If the aggregate is not already running in degraded mode due to single-disk failure reconstruction when the disk failure request is received, Data ONTAP fails the disk in question, swaps in a hot spare disk, and begins RAID reconstruction to replace the failed disk.

In addition, if one hundred or more media errors occur on a disk in a one-week period, Data ONTAP pre-fails the disk and causes Rapid RAID Recovery to start. For more information, see "Predictive disk failure and Rapid RAID Recovery" on page 144.

Failing disks at the thresholds listed in this section greatly decreases the likelihood of a storage system panic or double-disk failure during a single-disk failure reconstruction.

Viewing RAID status

About RAID status

You use the `aggr status` command to check the current RAID status and configuration for your aggregates.

To view RAID status for your aggregates, complete the following step.

Step	Action
1	Enter the following command: **`aggr status [aggr_name] -r`** *aggregate_name* is the name of the aggregate whose RAID status you want to view. **Note** If you omit the name of the aggregate (or the traditional volume), Data ONTAP displays the RAID status of all the aggregates on the storage system.

Possible RAID status displayed

The `aggr status -r` or `volume status -r` command displays the following possible status conditions that pertain to RAID:

- Degraded—The aggregate contains at least one degraded RAID group that is not being reconstructed after single- disk failure.
- Double degraded—The aggregate contains at least one RAID group with double-disk failure that is not being reconstructed (this state is possible if RAID-DP protection is enabled for the affected aggregate).
- Double reconstruction xx% complete—At least one RAID group in the aggregate is being reconstructed after experiencing a double-disk failure (this state is possible if RAID-DP protection is enabled for the affected aggregate).
- Mirrored—The aggregate is mirrored, and all of its RAID groups are functional.
- Mirror degraded—The aggregate is mirrored, and one of its plexes is offline or resynchronizing.
- Normal—The aggregate is unmirrored, and all of its RAID groups are functional.

- Partial—At least one disk was found for the aggregate, but two or more disks are missing.
- Reconstruction xx% complete—At least one RAID group in the aggregate is being reconstructed after experiencing a single- disk failure.
- Resyncing—The aggregate contains two plexes; one plex is resynchronizing with the aggregate.

Aggregate Management 5

About this chapter

This chapter describes how to use aggregates to manage storage system resources.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding aggregates" on page 184
- "Creating aggregates" on page 187
- "Changing the state of an aggregate" on page 193
- "Adding disks to aggregates" on page 198
- "Destroying aggregates" on page 204
- "Undestroying aggregates" on page 206
- "Physically moving aggregates" on page 208

Understanding aggregates

Aggregate management

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.

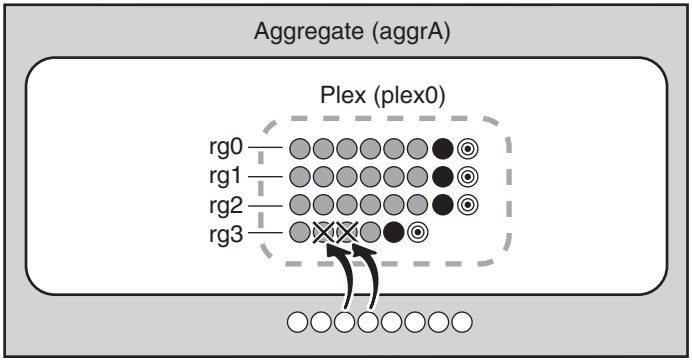
Each aggregate possesses its own RAID configuration, plex structure, and set of assigned disks. Within each aggregate you can create one or more FlexVol volumes—the logical file systems that share the physical storage resources, RAID configuration, and plex structure of that common containing aggregate.

For example, you can create a large aggregate with large numbers of disks in large RAID groups to support multiple FlexVol volumes, maximize your data resources, provide the best performance, and accommodate SnapVault backup.

You can also create a smaller aggregate to support FlexVol volumes that require special functions like SnapLock non-erasable data storage.

An unmirrored aggregate: In the following diagram, the unmirrored aggregate, arbitrarily named *aggrA* by the user, consists of one plex, which is made up of three double-parity RAID groups, automatically named *rg0*, *rg1*, and *rg2* by Data ONTAP.

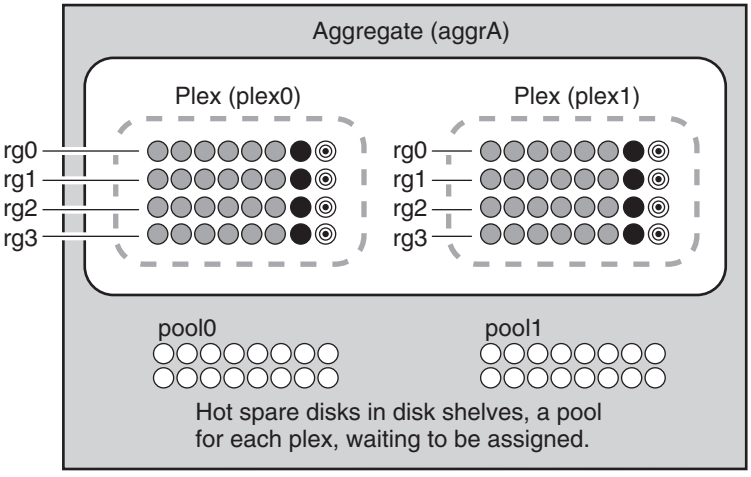
Notice that RAID-DP requires that both a parity disk and a double parity disk be in each RAID group. In addition to the disks that have been assigned to RAID groups, there are eight hot spare disks in the pool. In this diagram, two of the disks are needed to replace two failed disks, so only six will remain in the pool.



A mirrored aggregate: Consists of two plexes, which provides an even higher level of data redundancy via RAID-level mirroring. For an aggregate to be enabled for mirroring, the storage system must have a SyncMirror license for syncmirror_local or cluster_remote installed and enabled, and the storage system's disk configuration must support RAID-level mirroring.

When SyncMirror is enabled, all the disks are divided into two disk pools, and a copy of the plex is created. The plexes are physically separated (each plex has its own RAID groups and its own disk pool), and the plexes are updated simultaneously. This provides added protection against data loss if there is a double-disk failure or a loss of disk connectivity, because the unaffected plex continues to serve data while you fix the cause of the failure. Once the plex that had a problem is fixed, you can resynchronize the two plexes and reestablish the mirror relationship. For more information about snapshots, SnapMirror, and SyncMirror, see the *Data Protection Online Backup and Recovery Guide*.

In the following diagram, SyncMirror is enabled and implemented, so plex0 has been copied and automatically named *plex1* by Data ONTAP. Plex0 and plex1 contain copies of one or more file systems. In this diagram, thirty-two disks had been available prior to the SyncMirror relationship being initiated. After initiating SyncMirror, each pool has its own collection of sixteen hot spare disks.



When you create an aggregate, Data ONTAP assigns data disks and parity disks to RAID groups, depending on the options you choose, such as the size of the RAID group (based on the number of disks to be assigned to it) or the level of RAID protection.

Choosing the right size and the protection level for a RAID group depends on the kind of data that you intend to store on the disks in that RAID group. For more information about planning the size of RAID groups, see “Size of RAID groups” on page 25 and Chapter 4, “RAID Protection of Data,” on page 135.

Creating aggregates

About creating aggregates

When a single, unmirrored aggregate is first created, all the disks in the single plex must come from the same disk pool.

How Data ONTAP enforces checksum type rules

As mentioned in Chapter 3, Data ONTAP uses the disk's checksum type for RAID parity checksums. You must be aware of a disk's checksum type because Data ONTAP enforces the following rules when creating aggregates or adding disks to existing aggregates (these rules also apply to creating traditional volumes or adding disks to them):

- An aggregate can have only one checksum type, and it applies to the entire aggregate.
- When you create an aggregate:
  - Data ONTAP determines the checksum type of the aggregate, based on the type of disks available.
  - If enough block checksum disks (BCDs) are available, the aggregate uses BCDs.
  - Otherwise, the aggregate uses zoned checksum disks (ZCDs).
  - To use BCDs when you create a new aggregate, you must have at least the same number of block checksum spare disks available that you specify in the `aggr create` command.
- When you add disks to an existing aggregate:
  - You can add a BCD to either a block checksum aggregate or a zoned checksum aggregate.
  - You cannot add a ZCD to a block checksum aggregate.

If you have a system with both BCDs and ZCDs, Data ONTAP attempts to use the BCDs first. For example, if you issue the command to create an aggregate, Data ONTAP checks to see whether there are enough BCDs available.

- If there *are* enough BCDs, Data ONTAP creates a block checksum aggregate.
- If there *are not* enough BCDs, and there *are no* ZCDs available, the command to create an aggregate fails.
- If there *are not* enough BCDs, and there *are* ZCDs available, Data ONTAP checks to see whether there are enough of them to create the aggregate.

- If there *are not* enough ZCDs, Data ONTAP checks to see whether there are enough mixed disks to create the aggregate.
- If there *are* enough mixed disks, Data ONTAP mixes block and zoned checksum disks to create a zoned checksum aggregate.
- If there *are not* enough mixed disks, the command to create an aggregate fails.

Once an aggregate is created on a storage system, you cannot change the format of a disk. However, on NetApp V-Series systems, you can convert a disk from one checksum type to the other with the `disk assign -c block | zoned` command. For more information, see the V-Series *Software Setup, Installation, and Management Guide*.

Data ONTAP automatically creates Snapshot copies of aggregates to support commands related to the SnapMirror software, which provides volume-level mirroring. For example, Data ONTAP uses Snapshot copies when data in two plexes of a mirrored aggregate need to be resynchronized.

You can accept or modify the default Snapshot copy schedule. You can also create one or more Snapshot copies at any time. For information about aggregate Snapshot copies, see the *System Administration Guide*. For information about Snapshot copies, plexes, and SyncMirror, see the *Data Protection Online Backup and Recovery Guide*.

Creating an aggregate

When you create an aggregate, you must provide the following information:

A name for the aggregate: The names must follow these naming conventions:

- Begin with either a letter or an underscore (_)
- Contain only letters, digits, and underscores
- Contain no more than 255 characters

Disks to include in the aggregate: You specify disks by using the `-d` option and their IDs or by the number of disks of a specified size.

All of the disks in an aggregate must follow these rules:

- Disks must be of the same type (FC-AL, ATA, or SCSI).
- Disks must have the same RPM.

If disks with different speeds are present on a storage system (for example, both 10,000 RPM and 15,000 RPM disks), Data ONTAP avoids mixing them within one aggregate. By default, Data ONTAP selects disks

- With the same speed when creating an aggregate in response to the following commands:
  - `aggr create`
  - `vol create`
- That match the speed of existing disks in the aggregate that requires expansion or mirroring in response to the following commands:
  - `aggr add`
  - `aggr mirror`
  - `vol add`
  - `vol mirror`

If you use the `-d` option to specify a list of disks for commands that add disks, the operation will fail if the speeds of the disks differ from each other or differ from the speed of disks already included in the aggregate. The commands for which the `-d` option will fail in this case are `aggr create`, `aggr add`, `aggr mirror`, `vol create`, `vol add`, and `vol mirror`. For example, if you enter `aggr create vol4 -d 9b.25 9b.26 9b.27` and two of the disks are of different speeds, the operation fails.

When using the `aggr create` or `vol create` commands, you can use the `-R` *rpm* option to specify the type of disk to used based on speed. You only need to use this option on storage systems that have different disks with different speeds. Typical values for *rpm* are: 5400, 7200, 10000, and 15000. The `-R` option cannot be used with the `-d` option.

If you have any question concerning the speed of a disk that you are planning to specify, use the `sysconfig -r` command to ascertain the speed of the disks that you want to specify.

Attention

It is possible to override the RPM check with option `-f`, but doing this might have a negative impact on the performance of the resulting aggregate.

Data ONTAP periodically checks if adequate spares are available for the storage system. In those checks, only disks with matching or higher speeds are considered as adequate spares. However, if a disk fails and a spare with matching speed is not available, Data ONTAP may use a spare with a different (higher or lower) speed for RAID reconstruction.

Note

If an aggregate happens to include disks with different speeds and adequate spares are present, you can use the `disk replace` command to replace mismatched disks. Data ONTAP will use Rapid RAID Recovery to copy such disks to more appropriate replacements.

Note

If you are setting up aggregates on an FAS270c storage system with two internal system heads or a system licensed for SnapMover, you might have to assign the disks to one of the storage systems before creating aggregates on those systems. For more information, see "Software-based disk ownership" on page 58.

For information about creating aggregates, see the na_aggr man page.

To create an aggregate, complete the following steps.

Step	Action
1	View a list of the spare disks on your storage system. These disks are available for you to assign to the aggregate that you want to create. Enter the following command: `aggr status -s` **Result:** The output of `aggr status -s` lists all the spare disks that you can select for inclusion in the aggregate and their capacities.

Step	Action
2	Enter the following command: **`aggr create aggr_name [-f] [-m] [-n] [-t { raid4 | raid_dp} ] [-r raidsize] [-T disk-type][-R rpm] disk-list`** *aggr_name* is the name for the new aggregate. `-f` overrides the default behavior that does not permit disks in a plex to span disk pools. This option also allows you to mix disks with different RPM speeds. `-m` specifies the optional creation of a SyncMirror-replicated volume if you want to supplement RAID protection with SyncMirror protection. A SyncMirror license is required for this feature. `-t {raid4 | raid_dp}` specifies the level of RAID protection you want to provide for this aggregate. If no RAID level is specified, the default value (`raid_dp`) is applied. `-r` *raidsize* is the maximum number of disks that you want RAID groups created in this aggregate to consist of. If the last RAID group created contains fewer disks than the value specified, any new disks that are added to this aggregate are added to this RAID group until that RAID group reaches the number of disks specified. When that point is reached, a new RAID group will be created for any additional disks added to the aggregate.

Step	Action
	`-T` *disk-type* specifies one of the following types of disk to be used: ATA, EATA, FCAL, LUN, and SCSI. This option is only needed when creating aggregates on systems that have mixed disks. Mixing disks of different types in one aggregate is not allowed. You cannot use the `-T` option in combination with the `-d` option. `-R` *rpm* specifies the type of disk to used based on its speed. Use this option only on storage systems having different disks with different speeds. Typical values for *rpm* are: 5400, 7200, 10000, and 15000. The `-R` option cannot be used with the `-d` option.
	disk-list is one of the following: ◆ *ndisks*[@*disk-size*] *ndisks* is the number of disks to use. It must be at least 2 (3 if RAID-DP is configured). *disk-size* is the disk size to use, in gigabytes. You must have at least *ndisks* available disks of the size you specify. ◆ `-d` *disk_name1 disk_name2... disk_nameN* *disk_name1*, *disk_name2,* and *disk_nameN* are disk IDs of one or more available disks; use a space to separate multiple disks.
3	Enter the following command to verify that the aggregate exists as you specified: **`aggr status aggr_name -r`** *aggr_name* is the name of the aggregate whose existence you want to verify. **Result:** The system displays the RAID groups and disks of the specified aggregate on your storage system.

Aggregate creation example: The following command creates an aggregate called newaggr, with no more than eight disks in a RAID group consisting of the disks with disk IDs 8.1, 8.2, 8.3, and 8.4.

```
aggr create newaggr -r 8 -d 8.1 8.2 8.3 8.4.
```

Changing the state of an aggregate

Aggregate states

An aggregate can be in one of the following three states:

- Online—Read and write access to volumes hosted on this aggregate is allowed. An online aggregate can be further described as follows:
  - `copying`—The aggregate is currently the target aggregate of an active `aggr copy` operation.
  - `degraded`—The aggregate contains at least one degraded RAID group that is not being reconstructed after single disk failure.
  - `double degraded`—The aggregate contains at least one RAID group with double disk failure that is not being reconstructed (this state is possible if RAID-DP protection is enabled for the affected aggregate).
  - `double reconstruction xx% complete`—At least one RAID group in the aggregate is being reconstructed after experiencing double disk failure (this state is possible if RAID-DP protection is enabled for the affected aggregate).
  - `foreign`—Disks that the aggregate contains were moved to the current storage system from another storage system.
  - `growing`—Disks are in the process of being added to the aggregate.
  - `initializing`—The aggregate is in the process of being initialized.
  - `invalid`—The aggregate does not contain a valid file system.
  - `ironing`—A WAFL consistency check is being performed on the aggregate.
  - `mirrored`—The aggregate is mirrored and all of its RAID groups are functional.
  - `mirror degraded`—The aggregate is a mirrored aggregate and one of its plexes is offline or resynchronizing.
  - `needs check`—WAFL consistency check needs to be performed on the aggregate.
  - `normal`—The aggregate is unmirrored and all of its RAID groups are functional.
  - `partial`—At least one disk was found for the aggregate, but two or more disks are missing.
  - `reconstruction xx% complete`—At least one RAID group in the aggregate is being reconstructed after experiencing single disk failure.

- ❖ `resyncing`—The aggregate contains two plexes; one plex is resynchronizing with the aggregate.
- ❖ `verifying`—A mirror verification operation is currently running on the aggregate.
- ❖ `wafl inconsistent`—The aggregate has been marked corrupted; contact technical support.

- Restricted—Some operations, such as parity reconstruction, are allowed, but data access is not allowed (aggregates cannot be restricted if they still contain FlexVol volumes).
- Offline—Read or write access is not allowed (aggregates cannot be taken offline if they still contain any FlexVol volumes).

Determining the state of aggregates

To determine what state an aggregate is in, complete the following step.

Step	Action

1	Enter the following command:
	`aggr status`
	This command displays a concise summary of all the aggregates and traditional volumes in the storage system.
	Example: In the following example, the *State* column displays whether the aggregate is online, offline, or restricted. The *Status* column displays the RAID level and, lists any status other than normal (in the case of volA, below, the status is mirrored).
	`> aggr status`
	`Aggr Type  State      Status    Options`
	`vol0 AGGR  online     raid4     root,`
	`volA TRAD  online     raid_dp`
	`                       mirrored`

When to take an aggregate offline

You can take an aggregate offline and make it unavailable to the storage system. You do this for the following reasons:

- To perform maintenance on the aggregate
- To destroy an aggregate
- To undestroy an aggregate

Taking an aggregate offline

There are two ways to take an aggregate offline, depending on whether Data ONTAP is running in normal or maintenance mode. In normal mode, you must first offline and destroy all of the FlexVol volumes in the aggregate. In maintenance mode, the FlexVol volumes are preserved.

To take an aggregate offline while Data ONTAP is running in normal mode, complete the following steps.

Step	Action
1	Ensure that all FlexVol volumes in the aggregate have been taken offline and destroyed.
2	Enter the following command: **`aggr offline aggr_name`** *aggr_name* is the name of the aggregate to be taken offline.

To enter into maintenance mode and take an aggregate offline, complete the following steps.

Step	Action
1	Turn on or reboot the system. When prompted to do so, press Ctrl-C to display the boot menu.
2	Enter the choice for booting in maintenance mode.
3	Enter the following command: **`aggr offline aggr_name`** *aggr_name* is the name of the aggregate to be taken offline.
4	Halt the system to exit maintenance mode by entering the following command: **`halt`**
5	Reboot the system. The system will reboot in normal mode.

Restricting an aggregate

You only restrict an aggregate if you want it to be the target of an aggregate copy operation. For information about the aggregate copy operation, see the *Data Protection Online Backup and Recovery Guide*.

To restrict an aggregate, complete the following step.

Step	Action
1	Enter the following command: **`aggr restrict aggr_name`** *aggr_name* is the name of the aggregate to be made restricted.

Bringing an aggregate online

You bring an aggregate online to make it available to the storage system after you have taken it offline and are ready to put it back in service.

To bring an aggregate online, complete the following step.

Step	Action
1	Enter the following command: **`aggr online aggr_name`** *aggr_name* is the name of the aggregate to reactivate. **Attention** If you bring an inconsistent aggregate online, it might suffer further file system corruption. If the aggregate is inconsistent, the command prompts you for confirmation.

Renaming an aggregate

Generally, you might want to rename aggregates to give them descriptive names.

To rename an aggregate, complete the following step.

Step	Action
1	Enter the following command: **`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. **Result:** The aggregate is renamed.

Adding disks to aggregates

Rules for adding disks to an aggregate

You can add disks of various sizes in an aggregate, using the following rules:

- You can add only hot spare disks to an aggregate.
- You must specify the aggregate to which you are adding the disks.
- If you are using mirrored aggregates, the disks must come from the same spare disk pool.
- If the added disk replaces a failed data disk, its capacity is limited to that of the failed disk.
- If the added disk is not replacing a failed data disk and it is not larger than the parity disk, its full capacity (subject to rounding) is available as a data disk.
- If the added disk is larger than an existing parity disk, see "Adding disks larger than the parity disk" on page 199.

If you want to add disks of different speeds, follow the guidelines described in the section about "Disks must have the same RPM." on page 188.

Checksum type rules for creating or expanding aggregates

You must use disks of the appropriate checksum type to create or expand aggregates, as described in the following rules.

- You can add a BCD to a block checksum aggregate or a zoned checksum aggregate.
- You cannot add a ZCD to a block checksum aggregate. For information, see "How Data ONTAP enforces checksum type rules" on page 187.
- To use block checksums when you create a new aggregate, you must have at least the number of block checksum spare disks available that you specified in the `aggr create` command.

The following table shows the types of disks that you can add to an existing aggregate of each type.

Disk type	Block checksum aggregate	Zoned checksum aggregate
Block checksum	OK to add	OK to add
Zoned checksum	Not OK to add	OK to add

Hot spare disk planning for aggregates

To fully support an aggregate's RAID disk failure protection, at least one hot spare disk is required for that aggregate. As a result, the storage system should contain spare disks of sufficient number and capacity to

- Support the size of the aggregate that you want to create
- Serve as replacement disks should disk failure occur in any aggregate

Note

The size of the spare disks should be equal to or greater than the size of the aggregate disks that the spare disks might replace.

To avoid possible data corruption with a single disk failure, always install at least one spare disk matching the size and speed of each aggregate disk.

Adding disks larger than the parity disk

If an added disk is larger than an existing parity disk, the added disk replaces the smaller disk as the parity disk, and the smaller disk becomes a data disk. This enforces a Data ONTAP rule that the parity disk must be at least as large as the largest data disk in a RAID group.

Note

In aggregates configured with RAID-DP, the larger added disk replaces any smaller regular parity disk, but its capacity is reduced, if necessary, to avoid exceeding the capacity of the smaller-sized dParity disk.

Adding disks to an aggregate

To add new disks to an aggregate or a traditional volume, complete the following steps.

Step	Action
1	Verify that hot spare disks are available for you to add by entering the following command: `aggr status -s`

Step	Action
2	Add the disks by entering the following command: **`aggr add aggr_name [-f] [-n] {ndisks[@disk-size]	[-d disk1 [disk2 ...] [disk1 [disk2 ...]] }`** *aggr_name* is the name of the aggregate to which you are adding the disks. `-f` overrides the default behavior that does not permit disks in a plex to span disk pools (only applicable if SyncMirror is licensed). This option also allows you to mix disks with different speeds. `-n` causes the command that Data ONTAP will execute to be displayed without actually executing the command. This is useful for displaying the disks that would be automatically selected prior to executing the command. *ndisks* is the number of disks to use. *disk-size* is the disk size, in gigabytes, to use. You must have at least *ndisks* available disks of the size you specify. `-d` specifies that the *disk-name* will follow. If the aggregate is mirrored, then the `-d` argument must be used twice (if you are specifying *disk-names*). *disk-name* is the disk number of a spare disk; use a space to separate disk numbers. The disk number is under the Device column in the `aggr status -s` display. **Note** If you want to use block checksum disks in a zoned checksum aggregate even though there are still zoned checksum hot spare disks, use the `-d` option to select the disks. **Examples:** The following command adds four 72-GB disks to the thisaggr aggregate: `aggr add thisaggr 4@72` The following command adds the disks 7.17 and 7.26 to the thisaggr aggregate: `aggr add thisaggr -d 7.17 7.26`

Adding disks to a specific RAID group in an aggregate

If an aggregate has more than one RAID group, you can specify the RAID group to which you are adding disks. To add disks to a specific RAID group of an aggregate, complete the following step.

Step	Action
1	Enter the following command: `aggr add` *`aggr_name`* `-g` *`raidgroup ndisks`*`[@`*`disk-size`*`] | -d` *`disk-name...`* *raidgroup* is a RAID group in the aggregate specified by *aggr_name* **Example:** The following command adds two disks to RAID group 0 of the vol0 volume: `aggr add aggr0 -g rg0 2`

The number of disks you can add to a specific RAID group is limited by the `raidsize` setting of the aggregate to which that group belongs. For more information, see Chapter 4, "Changing the size of existing RAID groups," on page 160

Forcibly adding disks to aggregates

If you try to add disks to an aggregate (or traditional volume) under the following situations, the operation will fail:

- The disks specified in the `aggr add` (or `vol add`) command would cause the disks on a mirrored aggregate to consist of disks from two spare pools.
- The disks specified in the `aggr add` (or `vol add`) command have a different speed in revolutions per minute (RPM) than that of existing disks in the aggregate.

If you add disks to an aggregate (or traditional volume) under the following situation, the operation will prompt you for confirmation, and then succeed or abort, depending on your response.

- The disks specified in the `aggr add` command would add disks to a RAID group other than the last RAID group, thereby making it impossible for the file system to revert to an earlier version than Data ONTAP 6.2.

To force Data ONTAP to add disks in these situations, complete the following step.

Step	Action
1	Enter the following command: `aggr add aggr-name -f [-g raidgroup] -d disk-name...` **Note** You must use the `-g` *raidgroup* option to specify a RAID group other than the last RAID group in the aggregate.

Displaying disk space usage on an aggregate

You use the `aggr show_space` command to display how much disk space FlexVol volumes use in an aggregate. The command shows the following categories of information. If you specify the name of an aggregate, the command only displays information about that aggregate. Otherwise, the command displays information about all of the aggregates in the storage system.

- WAFL reserve—the amount of space used to store the metadata that Data ONTAP uses to maintain the volume.
- Snapshot copy reserve—the amount of space reserved for aggregate Snapshot copies.
- Usable space—the amount of total usable space (total disk space less the amount of space reserved for WAFL metadata and Snapshot copies).
- Allocated space—the amount of space that was reserved for the volume when the aggregate was created, and the space used by non-reserved data.

  For guaranteed volumes, this is the same amount of space as the size of the volume, since no data is unreserved.

  For non-guaranteed volumes, this is the same amount of space as the used space, since all of the data is unreserved.
- Used space—the amount of space that occupies disk blocks. It includes the metadata required to maintain the FlexVol volume. It can be greater than the Allocated value.

  Note
  This value is not the same as the value displayed for "used space" by the `df` command.

- Available space—the amount of free space in the aggregate. You can also use the `df` command to display the amount of available space.
- Total disk space—the amount of total disk space available to the aggregate

All of the values are displayed in 1024-byte blocks, unless you specify one of the following sizing options:

- `-h` displays the output of the values in the appropriate size, automatically scaled by Data ONTAP
- `-k` displays the output in kilobytes
- `-m` displays the output in megabytes
- `-g` displays the output in gigabytes
- `-t` displays the output in terabytes

To display the disk usage of an aggregate, complete the following step.

Step	Action

1	Enter the following command:
	`aggr show_space` ***`aggr_name`***
	Example:
	`aggr show_space -h aggr1`
	
	`Aggregate 'aggr1'`
	`Volume Reserved Used Guarantee`
	`vol1 100MB 80MB volume`
	`vol2 50MB 40MB volume`
	`vol3 21MB 21MB none`
	
	`Aggregate Reserved Used Avail`
	`Total space 171MB 142MB 83MB`
	`Snap reserve 13MB 2788KB 10MB`
	`WAFL reserve 30MB 1476KB 28MB`

After adding disks for LUNs, you run reallocation jobs

After you add disks to an aggregate, run a full reallocation job on each FlexVol volume contained in that aggregate. For information on how to perform this task, see your *Block Access Management Guide*.

Destroying aggregates

About destroying aggregates

When you destroy an aggregate, Data ONTAP converts its parity disk and all its data disks back into hot spares. You can then use the spares in other aggregates and other storage systems. Before you can destroy an aggregate, you must destroy all of the FlexVol volumes contained by that aggregate.

There are two reasons to destroy an aggregate:

- You no longer need the data it contains.
- You copied its data to reside elsewhere.

Attention

If you destroy an aggregate, all the data in the aggregate is destroyed and no longer accessible, unless you undestroy it before any of the data is corrupted.

Note

You can destroy a SnapLock Enterprise aggregate at any time; however, you cannot destroy a SnapLock Compliance aggregate until the retention periods for all data contained in it have expired.

Destroying an aggregate

To destroy an aggregate, complete the following steps.

Step	Action
1	Take all FlexVol volumes offline and destroy them by entering the following commands for each volume: **`vol offline vol_name`** **`vol destroy vol_name`**
2	Take the aggregate offline by entering the following command: **`aggr offline aggr_name`** *aggr_name* is the name of the aggregate that you intend to destroy. **Example:** `system> aggr offline aggr_A` **Result:** The following message is displayed. `Aggregate 'aggr_A' is now offline.`

Step	Action

3	Destroy the aggregate by entering the following command:
	`aggr destroy aggr_name`
	aggr_name is the name of the aggregate that you are destroying and whose disks will be converted to hot spares.
	Example: `system> aggr destroy aggr_A`
	Result: The following message is displayed.
	`Are you sure you want to destroy this aggregate ?`
	After typing `y`, the following message is displayed.
	`Aggregate 'aggr_A' destroyed.`

Undestroying aggregates

About undestroying aggregates

You can undestroy a partially intact or previously destroyed aggregate or traditional volume, as long as the aggregate or volume is not Snaplock-compliant.

You must know the name of the aggregate you want to undestroy, because there is no Data ONTAP command available to display destroyed aggregates, nor do they appear in FilerView.

Attention

After undestroying an aggregate or traditional volume, you must run the `wafliron` program with the privilege level set to advanced. If you need assistance, contact your local NetApp sales representative, PSE, or PSC.

Undestroying an aggregate or a traditional volume

To undestroy an aggregate or a traditional volume, complete the following steps.

Step	Action
1	Ensure the `raid.aggr.undestroy.enable` option is set to On by entering the following command: **`options raid.aggr.undestroy.enable on`** **Note** The default for this option is On for Data ONTAP 7.0.1 and later. For earlier releases, the default is Off.
2	If you want to display the disks that are contained by the destroyed aggregate you want to undestroy, enter the following command: **`aggr undestroy -n`** ***`aggr_name`*** *aggr_name* is the name of a previously destroyed aggregate or traditional volume that you want to recover.

Step	Action

3	Undestroy the aggregate or traditional volume by entering the following command:
	`aggr undestroy aggr_name`
	aggr_name is the name of a previously destroyed aggregate or traditional volume that you want to recover.
	Example: `system> aggr undestroy aggr1`
	Result: 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 with a date and time stamp appears for each RAID disk that is restored to the aggregate and has its label edited. The last line of the message says:
	`Aggregate 'aggr1' undestroyed. Run wafliron to bring the aggregate online.`

4	Set the privilege level to advanced by entering the following command: **`priv set advanced`**
5	Run the `wafliron` program by entering the following command: **`aggr wafliron start aggr_name`**

Physically moving aggregates

About physically moving aggregates

You can physically move aggregates from one storage system to another. You might want to move an aggregate to a different storage system to perform one of the following tasks:

- Replace a disk shelf with one that has a greater storage capacity
- Replace current disks with larger disks
- Gain access to the files on disks belonging to a malfunctioning storage system

You can physically move disks, disk shelves, or loops to move an aggregate from one storage system to another.

When performing either of these types of move, the following terms are used:

- The *source storage system* is the storage system from which you are moving the aggregate.
- The *destination storage system* is the storage system to which you are moving the aggregate.
- The aggregate you are moving is a *foreign aggregate* to the destination storage system.

You should only move disks from a source storage system to a destination storage system if the destination storage system has higher NVRAM capacity.

Note

The procedure described here does not apply to V-Series systems. For information about how to physically move aggregates in V-Series systems, see the V-Series *Software Setup, Installation, and Management Guide*.

Physically moving an aggregate

To physically move an aggregate, complete the following steps.

Step	Action
1	In normal mode, enter the following command at the source storage system to locate the disks that contain the aggregate: **`aggr status aggr_name -r`** **Result:** The locations of the data and parity disks in the aggregate appear under the aggregate name on the same line as the labels Data and Parity.
2	Reboot the source storage system into maintenance mode.
3	In maintenance mode, take the aggregate that you want to move offline. **`aggr offline aggr_name`** Then follow the instructions in the disk shelf hardware guide to remove the disks from the source storage system.
4	Halt and turn off the destination storage system.
5	Install the disks in a disk shelf connected to the destination storage system.
6	Reboot the destination storage system in maintenance mode. **Result:** When the destination storage system boots, it takes the foreign aggregate offline. If the foreign aggregate has the same name as an existing aggregate on the storage system, the storage system renames it *aggr_name(1)*, where *aggr_name* is the original name of the aggregate. **Attention** If the foreign aggregate is incomplete, repeat Step 5 to add the missing disks. Do not try to add missing disks while the aggregate is online—doing so causes them to become hot spare disks.

Step	Action
7	If the storage system renamed the foreign aggregate because of a name conflict, 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`
8	Enter the following command to bring the aggregate online in the destination storage system: **`aggr online aggr_name`** *aggr_name* is the name of the aggregate. **Result:** The aggregate is online in its new location in the destination storage system.
9	Enter the following command to confirm that the added aggregate came online: **`aggr status aggr_name`** *aggr_name* is the name of the aggregate.
10	Power up and reboot the source storage system.
11	Reboot the destination storage system out of maintenance mode.

Volume Management 6

About this chapter

This chapter describes how to use volumes to contain and manage user data.

Topics in this chapter

This chapter discusses the following topics:

- "Traditional and FlexVol volumes" on page 212
- "Traditional volume operations" on page 215
- "FlexVol volume operations" on page 224
- "General volume operations" on page 243
- "Managing FlexCache volumes" on page 268
- "Space management for volumes and files" on page 283

Traditional and FlexVol volumes

About traditional and FlexVol volumes

Volumes are file systems that hold user data that is accessible via one or more of the access protocols supported by Data ONTAP, including NFS, CIFS, HTTP, WebDAV, FTP, FCP and iSCSI. You can create one or more snapshots of the data in a volume so that multiple, space-efficient, point-in-time images of the data can be maintained for such purposes as backup and error recovery.

Each volume depends on its containing aggregate for all its physical storage, that is, for all storage in the aggregate's disks and RAID groups. A volume is associated with its containing aggregate in one of the two following ways:

- A *traditional volume* is a volume that is contained by a single, dedicated, aggregate; it is tightly coupled with its containing aggregate. The only way to grow a traditional volume is to add entire disks to its containing aggregate. It is impossible to decrease the size of a traditional volume. The smallest possible traditional volume must occupy all of two disks (for RAID4) or three disks (for RAID-DP).

  No other volumes can get their storage from this containing aggregate.

  All volumes created with a version of Data ONTAP earlier than 7.0 are traditional volumes. If you upgrade to Data ONTAP 7.0 or later, your volumes and data remain unchanged, and the commands you used to manage your volumes and data are still supported.
- A *FlexVol volume* (sometimes called a flexible volume) is a volume that is loosely coupled to its containing aggregate. Because the 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.

  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.

Data ONTAP automatically creates and deletes Snapshot copies of data in volumes to support commands related to Snapshot technology. You can accept or modify the default Snapshot copy schedule. For more information about Snapshot copy, see the *Data Protection Online Backup and Recovery Guide*.

Note

FlexVol volumes have different best practices, optimal configurations, and performance characteristics compared to traditional volumes. Make sure you understand these differences and deploy the configuration that is optimal for your environment.

For information about deploying a storage solution with FlexVol volumes, including migration and performance considerations, see the technical report *Introduction to Data ONTAP Release 7G* (available from the NetApp Library at http://www.netapp.com/tech_library/ftp/3356.pdf).

Limits on how many volumes you can have

You can create up to 200 FlexVol and traditional volumes on a single storage system. In addition, the following limits apply.

Traditional volumes: You can have up to 100 traditional volumes and aggregates combined on a single storage system.

FlexVol volumes: The only limit imposed on FlexVol volumes is the overall system limit of 200 for all volumes.

For clusters, these limits apply to each node individually, so the overall limits for the pair are doubled.

Types of volume operations

The volume operations described in this chapter fall into three types:

- "Traditional volume operations" on page 215

  These are RAID and disk management operations that pertain only to traditional volumes.

  - "Creating traditional volumes" on page 216
  - "Physically transporting traditional volumes" on page 221
- "FlexVol volume operations" on page 224

  These are operations that use the advantages of FlexVol volumes, so they pertain only to FlexVol volumes.

  - "Creating FlexVol volumes" on page 225
  - "Resizing FlexVol volumes" on page 229
  - "Cloning FlexVol volumes" on page 231
  - "Displaying a FlexVol volume's containing aggregate" on page 242

- "General volume operations" on page 243

  These are operations that apply to both FlexVol and traditional volumes.

  - "Migrating between traditional volumes and FlexVol volumes" on page 244
  - "Managing volume languages" on page 253
  - "Determining volume status and state" on page 256
  - "Renaming volumes" on page 262
  - "General volume operations" on page 263
  - "Destroying volumes" on page 263
  - "Increasing the maximum number of files in a volume" on page 265
  - "Reallocating file and volume layout" on page 267

Traditional volume operations

About traditional volume operations

Operations that apply exclusively to traditional volumes generally involve management of the disks assigned to those volumes and the RAID groups to which those disks belong.

Traditional volume operations described in this section include:

- "Creating traditional volumes" on page 216
- "Physically transporting traditional volumes" on page 221

Additional traditional volume operations that are described in other chapters or other guides include:

- Configuring and managing RAID protection of volume data

  See "RAID Protection of Data" on page 135.
- Configuring and managing SyncMirror replication of volume data

  See the *Data Protection Online Backup And Recovery Guide*.
- Increasing the size of a traditional volume

  To increase the size of a traditional volume, you increase the size of its containing aggregate. For more information about increasing the size of an aggregate, see "Adding disks to aggregates" on page 198.
- Configuring and managing SnapLock volumes

  See the *Data Protection Online Recovery and Backup Guide*.

Creating traditional volumes

About creating traditional volumes

When you create a traditional volume, you provide the following information:

- A name for the volume

  For more information about volume naming conventions, see "Volume naming conventions" on page 216.
- An optional language for the volume

  The default value is the language of the root volume.

  For more information about choosing a volume language, see "Managing volume languages" on page 253.
- The RAID-related parameters for the aggregate that contains the new volume

  For a complete description of RAID-related options for volume creation see "Setting RAID level and group size" on page 149.

Volume naming conventions

You choose the volume names. The names must follow these naming conventions:

- Begin with either a letter or an underscore (_)
- Contain only letters, digits, and underscores
- Contain no more than 255 characters

Creating a traditional volume

To create a traditional volume, complete the following steps.

Step	Action
1	At the system prompt, enter the following command: `aggr status -s` **Result:** The output of `aggr status -s` lists all the hot-swapable spare disks that you can assign to the traditional volume and their capacities. **Note** If you are setting up traditional volumes on a FAS270c system with two internal heads, or a storage system that has SnapMover licensed, you might have to assign the disks before creating volumes on those storage systems. For more information, see "Software-based disk ownership" on page 58.

Step	Action
2	At the system prompt, enter the following command: **`aggr create vol_name -v [-l language_code] [-f] [-n] [-m] [-t raid-type] [-r raid-size] [-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. See "Viewing the language list online" on page 254. The `-L` flag is used only when creating SnapLock volumes. For more information about SnapLock volumes, see the *Data Protection Online Recovery and Backup Guide*. **Note** For a complete description of the all the options for the aggr command, see "Creating an aggregate" on page 188. For information about RAID related options for `aggr create`, see "Setting RAID level and group size" on page 149 or the na_aggr(1) man page. For backward compatibility, you can also use the `vol create` command to create a traditional volume. However, not all of the RAID related options are available for the `vol` command. For more information, see the na_vol(1) man page. **Result:** The new volume is created and, if NFS is in use, an entry for the new volume is added to the /etc/export file. **Example:** The following command creates a traditional volume called newvol, with no more than eight disks in a RAID group, using the French character set, and consisting of the disks with disk IDs 8.1, 8.2, 8.3, and 8.4. `aggr create newvol -v -r 8 -l fr -d 8.1 8.2 8.3 8.4`

Step	Action
3	Enter the following command to verify that the volume exists as you specified: **`aggr status vol_name -r`** *vol_name* is the name of the volume whose existence you want to verify. **Result:** The system displays the RAID groups and disks of the specified volume on your storage system.
4	If you access the storage system using CIFS, update your CIFS shares as necessary.
5	If you access the storage system using NFS, complete the following steps: 1. Verify that the line added to the /etc/exports file for the new volume is correct for your security model. 2. Add the appropriate mount point information to the /etc/fstab or /etc/vfstab file on clients that mount volumes from the storage system.

Parameters to accept or change after volume creation

After you create a volume, you can accept the defaults for CIFS oplocks and security style settings or you can change the values. You should decide what to do as soon as possible after creating the volume. If you change the parameters 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.

CIFS oplocks setting: The CIFS oplocks setting determines whether the volume uses CIFS oplocks. The default is to use CIFS oplocks.

For more information about CIFS oplocks, see "Changing the CIFS oplocks setting" on page 308.

Security style: The security style determines whether the files in a volume use NTFS security, UNIX security, or both.

For more information about file security styles, see "Understanding security styles" on page 303.

When you have a new storage system, the default depends on what protocols you licensed, as shown in the following table.

Protocol licenses	Default volume security style
CIFS only	NTFS
NFS only	UNIX
CIFS and NFS	UNIX

When you change the configuration of a storage system from one protocol to another (by licensing or unlicensing protocols), the default security style for new volumes changes as shown in the following table.

From	To	Default for new volumes	Note
NTFS	Multiprotocol	UNIX	The security styles of volumes are not changed.
Multiprotocol	NTFS	NTFS	The security style of all volumes is changed to NTFS.

Checksum type usage

A checksum type applies to an entire aggregate. An aggregate can have only one checksum type. For more information about checksum types, see "How Data ONTAP enforces checksum type rules" on page 187.

Physically transporting traditional volumes

About physically moving traditional volumes

You can physically move traditional volumes from one storage system to another. You might want to move a traditional volume to a different storage system to perform one of the following tasks:

- Replace a disk shelf with one that has a greater storage capacity
- Replace current disks with larger disks
- Gain access to the files on disks on a malfunctioning storage system

You can physically move disks, disk shelves, or loops to move a volume from one storage system to another. You need the manual for your disk shelf to move a traditional volume.

The following terms are used:

- The *source system* is the storage system from which you are moving the volume.
- The *destination system* is the storage system to which you are moving the volume.
- The volume you are moving is a *foreign volume* to the destination storage system.

Note

If MultiStore® and SnapMover licenses are installed, you might be able to move traditional volumes without moving the drives on which they are located. For more information, see the *MultiStore Management Guide*.

Moving a traditional volume

To physically move a traditional volume, perform the following steps.

Step	Action
1	Enter the following command at the source system to locate the disks that contain the volume *vol_name*: **`aggr status vol_name -r`** **Result:** The locations of the data and parity disks in the volume are displayed.

Step	Action
2	Enter the following command on the source system to take the volume and its containing aggregate offline: **`aggr offline vol_name`**
3	Follow the instructions in the disk shelf hardware guide to remove the data and parity disks for the specified volume from the source system.
4	Follow the instructions in the disk shelf hardware guide to install the disks in a disk shelf connected to the destination system. **Result:** When the destination system sees the disks, it places the foreign volume offline. If the foreign volume has the same name as an existing volume on the system, the system renames it *vol_name(d)*, where *vol_name* is the original name of the volume and *d* is a digit that makes the name unique.
5	Enter the following command to make sure that the newly moved volume is complete: **`aggr status new_vol_name`** *new_vol_name* is the (possibly new) name of the volume you just moved. **Attention** If the foreign volume is incomplete (it has a status of `partial`), add all missing disks before proceeding. Do not try to add missing disks after the volume comes online—doing so causes them to become hot spare disks. You can identify the disks currently used by the volume using the `aggr status -r` command.
6	If the storage system renamed the foreign volume because of a name conflict, enter the following command on the target system to rename the volume: **`aggr rename new_vol_name vol_name`** *new_vol_name* is the name of the volume you want to rename. *vol_name* is the new name of the volume.

Step	Action
7	Enter the following command on the target system to bring the volume and its containing aggregate online: **`aggr online vol_name`** *vol_name* is the name of the newly moved volume. **Result:** The volume is brought online on the target system.
8	Enter the following command to confirm that the added volume came online: **`aggr status vol_name`** *vol_name* is the name of the newly moved volume.
9	If you access the storage systems using CIFS, update your CIFS shares as necessary.
10	If you access the storage systems using NFS, complete the following steps for both the source and the destination system: **1.** Update the system /etc/exports file. **2.** Run `exportfs -a`. **3.** Update the appropriate mount point information to the /etc/fstab or /etc/vfstab file on clients that mount volumes from the storage system.

FlexVol volume operations

About FlexVol volume operations

These operations apply exclusively to FlexVol volumes because they take advantage of the virtual nature of FlexVol volumes.

FlexVol volume operations described in this section include:

- "Creating FlexVol volumes" on page 225
- "Resizing FlexVol volumes" on page 229
- "Cloning FlexVol volumes" on page 231
- "Configuring FlexVol volumes to grow automatically" on page 240
- "Displaying a FlexVol volume's containing aggregate" on page 242

Creating FlexVol volumes

About creating FlexVol volumes

When you create a FlexVol volume, you must provide the following information:

- A name for the volume
- The name of the containing aggregate
- The size of the volume

  The size of a FlexVol volume must be at least 20 MB. The maximum size is 16 TB, or what your storage system configuration can support.

In addition, you can provide the following optional values:

- The language used for file names

  The default language is the language of the root volume.
- The space guarantee setting for the new volume

  For more information, see "Space guarantees" on page 286.

Volume naming conventions

You choose the volume names. The names must follow these naming conventions:

- Begin with either a letter or an underscore (_)
- Contain only letters, digits, and underscores
- Contain no more than 255 characters

Creating a FlexVol volume

To create a FlexVol volume, complete the following steps.

Step	Action
1	If you have not already done so, create one or more aggregates to contain the FlexVol volumes that you want to create. To view a list of the aggregates that you have already created, and the volumes that they contain, enter the following command: `aggr status -v`

Step	Action
2	At the system prompt, enter the following command: **`vol create f_vol_name [-l language_code] [-s {volume|file|none}] aggr_name size{k|m|g|t}`** *f_vol_name* is the name for the new FlexVol volume (without the /vol/ prefix). This name must be different from all other volume names on the storage system. *language_code* specifies a language other than that of the root volume. See "Viewing the language list online" on page 254. `-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`. For more information, see "Space guarantees" on page 286. *aggr_name* is the name of the containing aggregate for this 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 twenty megabytes. If you do not specify a unit, *size* is taken as bytes and rounded up to the nearest multiple of 4 KB. **Result:** The new volume is created and, if NFS is in use, an entry is added to the /etc/export file for the new volume. **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`
3	Enter the following command to verify that the volume exists as you specified: **`vol status f_vol_name`** *f_vol_name* is the name of the FlexVol volume whose existence you want to verify.
4	If you access the storage system using CIFS, update the share information for the new volume.

Step	Action
5	If you access the storage system using NFS, complete the following steps: 1. Verify that the line added to the /etc/exports file for the new volume is correct for your security model. 2. Add the appropriate mount point information to the /etc/fstab or /etc/vfstab file on clients that mount volumes from the storage system.

Parameters to accept or change after volume creation

After you create a volume, you can accept the defaults for CIFS oplocks and security style settings or you can change the values. You should decide what to do as soon as possible after creating the volume. If you change the parameters 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.

CIFS oplocks setting: The CIFS oplocks setting determines whether the volume uses CIFS oplocks. The default is to use CIFS oplocks.

For more information about CIFS oplocks, see "Changing the CIFS oplocks setting" on page 308.

Security style: The security style determines whether the files in a volume use NTFS security, UNIX security, or both.

For more information about file security styles, see "Understanding security styles" on page 303.

When you have a new storage system, the default depends on what protocols you licensed, as shown in the following table.

Protocol licenses	Default volume security style
CIFS only	NTFS
NFS only	UNIX
CIFS and NFS	UNIX

When you change the configuration of a storage system from one protocol to another, the default security style for new volumes changes as shown in the following table.

From	To	Default for new volumes	Note
NTFS	Multiprotocol	UNIX	The security styles of volumes are not changed.
Multiprotocol	NTFS	NTFS	The security style of all volumes is changed to NTFS.

Resizing FlexVol volumes

About resizing FlexVol volumes

You can increase or decrease the amount of space that an existing FlexVol volume can occupy on 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.

Resizing a FlexVol volume

To resize a FlexVol volume, complete the following steps.

Step	Action
1	Check the available space of the containing aggregate by entering the following command: **`df -A aggr_name`** *aggr_name* is the name of the containing aggregate for the FlexVol volume whose size you want to change.
2	If you want to determine the current size of the volume, enter one of the following commands: **`vol size f_vol_name`** **`df f_vol_name`** *f_vol_name* is the name of the FlexVol volume that you intend to resize.

Step	Action
3	Enter the following command to resize the volume: **vol size *f_vol_name* [+|-] *n*{k|m|g|t}** *f_vol_name* is the name of the FlexVol volume that you intend to resize. 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.
4	Verify the success of the resize operation by entering the following command: **vol size *f_vol_name***

Cloning FlexVol volumes

About cloning FlexVol volumes

Data ONTAP provides the ability to clone FlexVol volumes, creating FlexClone volumes. The following list outlines some key facts about FlexClone volumes that you should know:

- You must install the license for the FlexClone feature before you can create FlexClone volumes.
- FlexClone volumes are a point-in-time, writable copy of the parent volume. Changes made to the parent volume after the FlexClone volume is created are not reflected in the FlexClone volume.
- FlexClone volumes are fully functional volumes; you manage them using the `vol` command, just as you do the parent volume.
- FlexClone volumes always exist in the same aggregate as their parent volumes.
- FlexClone volumes can themselves be cloned.
- FlexClone volumes and their parent volumes share the same disk space for any data common to the clone and parent. This means that creating a FlexClone volume is instantaneous and requires no additional disk space (until changes are made to the clone or parent).
- Because creating a FlexClone volume does not involve copying data, FlexClone volume creation is very fast.
- A FlexClone volume is created with the same space guarantee as its parent.

  For more information, see "Space guarantees" on page 286.
- While a FlexClone volume exists, some operations on its parent are not allowed.

  For more information about these restrictions, see "Limitations of volume cloning" on page 232.
- If, at a later time, you decide you want to sever the connection between the parent and the clone, you can split the FlexClone volume. This removes all

restrictions on the parent volume and enables the space guarantee on the FlexClone volume.

Attention

Splitting a FlexClone volume from its parent volume deletes all existing snapshots of the FlexClone volume.

For more information, see "FlexClone volumes and space guarantees" on page 235.

- When a FlexClone volume is created, quotas are reset on the FlexClone volume, and any LUNs present in the parent volume are present in the FlexClone volume but are unmapped.

  For more information about using volume cloning with LUNs, see your *Block Access Management Guide*.

- Only FlexVol volumes can be cloned. To create a copy of a traditional volume, you must use the `vol copy` command, which creates a distinct copy with its own storage.

Uses of volume cloning

You can use volume cloning whenever you need a writable, point-in-time copy of an existing FlexVol volume, including the following scenarios:

- You need to create a temporary copy of a volume for testing purposes.
- You need to make a copy of your data available to additional users without giving them access to the production data.
- You want to create a clone of a database for manipulation and projection operations, while preserving the original data in unaltered form.

Benefits of volume cloning versus volume copying

Volume cloning provides similar results to volume copying, but cloning offers some important advantages over volume copying:

- Volume cloning is instantaneous, whereas volume copying can be time consuming.
- If the original and cloned volumes share a large amount of identical data, considerable space is saved because the shared data is not duplicated between the volume and the clone.

Limitations of volume cloning

The following operations are not allowed on parent volumes or their clones.

- You cannot delete the base snapshot in a parent volume while a FlexClone volume using that snapshot exists. The base snapshot is the snapshot that

was used to create the FlexClone volume, and is marked `busy, vclone` in the parent volume.

- You cannot perform a volume SnapRestore® operation on the parent volume using a snapshot that was taken before the base snapshot was taken.
- You cannot destroy a parent volume if any clone of that volume exists.
- You cannot clone a volume that has been taken offline, although you can take the parent volume offline after it has been cloned.
- You cannot create a volume SnapMirror relationship or perform a `vol copy` command using a FlexClone volume or its parent as the destination volume.

  For more information about using SnapMirror with FlexClone volumes, see "Using volume SnapMirror replication with FlexClone volumes" on page 236.

Cloning a FlexVol volume

To create a FlexClone volume by cloning a FlexVol volume, complete the following steps.

Step	Action
1	Ensure that you have the flex_clone license installed.

Step	Action
2	Enter the following command to clone the volume: **`vol clone create cl_vol_name [-s {volume|file|none}] -b f_p_vol_name [parent_snap]`** *cl_vol_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. For more information, see "Space guarantees" on page 286. *f_p_vol_name* is the name of the FlexVol volume that you intend to clone. *parent_snap* is the name of the base snapshot of the parent FlexVol volume. If no name is specified, Data ONTAP creates a base snapshot with the name clone_*cl_name_prefix.id,* where *cl_name_prefix* is the name of the new FlexClone volume (up to 16 characters) and *id* is a unique digit identifier (for example 1,2, etc.). The base snapshot cannot be deleted as long as the parent volume or any of its clones exists. **Result:** The FlexClone volume is created and, if NFS is in use, an entry is added to the /etc/export file for every entry found for the parent volume.
	Example snapshot name: To create a FlexClone volume "newclone" from the parent "flexvol1", the following command is entered: `vol clone create newclone -b flexvol1` The snapshot created by Data ONTAP is named "clone_newclone.1".
3	Verify the success of the FlexClone volume creation by entering the following command: **`vol status -v cl_vol_name`**

FlexClone volumes and space guarantees

A FlexClone volume inherits its initial space guarantee setting 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` as well. You can change the FlexClone volume's space guarantee at any time.

FlexClone volumes and shared snapshots

A newly created FlexClone volume uses the snapshots it shares with its parent volume to optimize its space requirements from the aggregate when applying space guarantees. This is important because you can delete the shared snapshots from the FlexClone volume, but doing so may increase the space requirements of the FlexClone volume in the aggregate if space guarantees are being used.

Example: Suppose 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 only needs to allocate 100 MB minus 70 MB or 30 MB of free space to the clone.

Now, suppose that you delete the base snapshot from the FlexClone volume. Now, the FlexClone volume can no longer optimize its space requirements, and the full 100 MB is required from the containing aggregate.

What to do if you are prevented from deleting a snapshot from a FlexClone volume

If you are prevented from deleting a snapshot from a FlexClone volume due to "insufficient space in the aggregate," it is because deleting that snapshot requires the allocation of more data than the aggregate currently has available. You can either increase the size of the aggregate, or change the space guarantee of the FlexClone volume.

Note

Before changing the space guarantee of the FlexClone volume, make sure you understand the effects of making such a change. For more information, see "Space guarantees" on page 286.

Identifying shared snapshots in FlexClone volumes

Snapshots that are shared between a FlexClone volume and its parent are not identified as such in the FlexClone volume. However, you can identify a shared snapshot by listing the snapshots *in the parent volume*. Any snapshot that appears as `busy, vclone` in the parent volume and is also present in the FlexClone volume is a shared snapshot.

Using volume SnapMirror replication with FlexClone volumes

Because both volume SnapMirror replication and FlexClone volumes rely on snapshots, there are some restrictions on how the two features can be used together.

Creating a volume SnapMirror relationship using an existing FlexClone volume or its parent: 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.

Creating a FlexClone volume from volumes currently in a SnapMirror relationship: You can create a FlexClone volume from a volume that is currently either the source or destination in an existing volume SnapMirror relationship. For example, you might want to create a FlexClone volume to create a writable copy of a SnapMirror destination volume without affecting the data in the SnapMirror source volume.

However, when you create the FlexClone volume, you might lock a snapshot that is used by SnapMirror. If that happens, SnapMirror stops replicating to the destination volume until the FlexClone volume is destroyed or split from its parent. You have two options for addressing this issue:

- If your need for the FlexClone volume is temporary, and you can accept the temporary cessation of SnapMirror replication, you can create the FlexClone volume and either delete it or split it from its parent when possible. At that time, the SnapMirror replication will continue normally.
- If you cannot accept the temporary cessation of SnapMirror replication, you can create a snapshot in the SnapMirror source volume, and then use that snapshot to create the FlexClone volume. (If you are creating the FlexClone volume from the destination volume, you must wait until that snapshot replicates to the SnapMirror destination volume.) This method allows you to create the clone without locking down a snapshot that is in use by SnapMirror.

About splitting a FlexClone volume from its parent volume

You might want to split your FlexClone volume and its parent into two independent volumes that occupy their own disk space.

Attention

When you split a FlexClone volume from its parent, all existing snapshots of the FlexClone volume are deleted.

Splitting a FlexClone volume from its parent will remove any space optimizations currently employed by the FlexClone volume. After the split, both the FlexClone volume and the parent volume will require the full space allocation determined by their space guarantees.

Because the clone-splitting operation is a copy operation that might take considerable time to carry out, Data ONTAP also provides 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.

If you take the FlexClone volume offline while the splitting operation is in progress, the operation is suspended; when you bring the FlexClone volume back online, the splitting operation resumes.

Once a FlexClone volume and its parent volume have been split, they cannot be rejoined.

Splitting a FlexClone volume

To split a FlexClone volume from its parent volume, complete the following steps.

Step	Action
1	Determine the approximate amount of free space required to split a FlexClone volume from its parent, enter the following command: **`vol clone split estimate`** ***`clone_name`***

Step	Action
2	Verify that enough free space exists in the containing aggregate to support storing the data of both the FlexClone volume and its parent volume by entering the following command: **df -A *aggr_name*** *aggr_name* is the name of the containing aggregate of the FlexClone volume that you want to split. The `avail` column tells you how much available space you have in your aggregate. **Note** When a FlexClone volume is split from its parent, the resulting two FlexVol volumes occupy completely different blocks within the same aggregate.
3	Enter the following command to split the volume: **vol clone split start *cl_vol_name*** *cl_vol_name* is the name of the FlexClone volume that you want to split from its parent. **Result:** The original volume and its clone begin to split apart, no longer sharing the blocks that they formerly shared. All existing snapshots of the FlexClone volume are deleted.
4	If you want to check the status of a clone-splitting operation, enter the following command: **vol clone status *cl_vol_name***
5	If you want to stop the progress of an ongoing clone-splitting operation, enter the following command: **vol clone stop *cl_vol_name*** **Result:** The clone-splitting operation halts; the original and FlexClone volumes remain clone partners, but the disk space that was duplicated up to that point remains duplicated. All existing snapshots of the FlexClone volume are deleted.

Step	Action
6	To display status for the newly split FlexVol volume and verify the success of the clone-splitting operation, enter the following command: **`vol status -v cl_vol_name`**

Configuring FlexVol volumes to grow automatically

About configuring FlexVol volumes to grow automatically

The capability to configure FlexVol volumes to grow automatically when they are close to running out of space is part of the space management policy. You use this capability, along with the capability to automatically delete snapshots, to ensure that the free space in your aggregates is used efficiently, and to reduce the likelihood of your FlexVol volumes having no free space available.

For more information on space management policies, see "Space management policies" on page 295, and your *Block Access Management Guide*.

Configuring a volume to grow automatically

To configure a volume to grow automatically, complete the following step.

Step	Action
1	Enter the following command: **`vol autosize vol-name [-m size] [-i size] [on|off|reset]`** *vol-name* is the name of the flexible volume. You cannot use this command on traditional volumes. `-m` *size* is the maximum size to which you can grow the volume. Specify a size in `k` (KB), `m` (MB), `g` (GB) or `t` (TB). The volume does not grow if its size is equal to or greater than the maximum size. `-i` *size* is the increment by which you grow the volume. Specify a size in `k` (KB), `m` (MB), `g` (GB) or `t` (TB). `on` enables the volume to automatically grow. `off` disables automatically growing the volume. By default, the `vol autosize` command is set to Off. `reset` restores the `autosize` settings of the volume to the default setting, which is Off.

Defining how you apply space management policies

You can configure Data ONTAP to apply space management policies in one of the following ways:

- Automatically grow the volume first, then automatically delete snapshots.

  This approach is useful if you create smaller flexible volumes and leave enough space in the aggregate to increase the size of these volumes as needed. If you provision your data based on aggregates, you might want to automatically grow the volume when it is nearly full before you begin automatically deleting snapshots.

- Automatically delete snapshots first, then grow the volume

  You might want to automatically delete snapshots if you are maintaining a large number of snapshots in your volume or you maintain older snapshots that are no longer needed.

To determine how Data ONTAP applies space management policies, complete the following step:

Step	Action
1	Enter the following command. **`vol options vol-name try_first [volume_grow|snap_delete]`** *vol-name* is the name of the flexible volume `volume_grow`—Automatically increase the size of the flexible volume before automatically deleting snapshots. `snap_delete`—Automatically delete snapshots before automatically increasing the size of the volume. For more information about this option, see your *Block Access Management Guide*.

Displaying a FlexVol volume's containing aggregate

Showing a FlexVol volume's containing aggregate

To display the name of a FlexVol volume's containing aggregate, complete the following step.

Step	Action
1	Enter the following command: **`vol container vol_name`** *vol_name* is the name of the volume whose containing aggregate you want to display.

General volume operations

About general volume operations

General volume operations apply to both traditional volumes and FlexVol volumes.

General volume operations described in this section include:

- "Migrating between traditional volumes and FlexVol volumes" on page 244
- "Managing duplicate volume names" on page 252
- "Managing volume languages" on page 253
- "Determining volume status and state" on page 256
- "Renaming volumes" on page 262
- "Destroying volumes" on page 263
- "Increasing the maximum number of files in a volume" on page 265
- "Reallocating file and volume layout" on page 267

Additional general volume operations that are described in other chapters or other guides include:

- Making a volume available

  For more information on making volumes available to users who are attempting access through NFS, CIFS, FTP, WebDAV, or HTTP protocols, see the *File Access and Protocols Management Guide*.
- Copying volumes

  For more information about copying volumes see the *Data Protection Online Backup and Recovery Guide*.
- Changing the root volume

  For more information about changing the root volume from one volume to another, see the section on the root volume in the *System Administration Guide*.

Migrating between traditional volumes and FlexVol volumes

About migrating between traditional and FlexVol volumes

FlexVol volumes have different best practices, optimal configurations, and performance characteristics compared to traditional volumes. Make sure you understand these differences by referring to the available documentation on FlexVol volumes and deploy the configuration that is optimal for your environment.

For information about deploying a storage solution with FlexVol volumes, including migration and performance considerations, see the technical report *Introduction to Data ONTAP Release 7G* (available from the NetApp Library at http://www.netapp.com/tech_library/ftp/3356.pdf). For information about configuring FlexVol volumes, see "FlexVol volume operations" on page 224. For information about configuring aggregates, see "Aggregate Management" on page 183.

The following list outlines some facts about migrating between traditional and FlexVol volumes that you should know:

- You cannot convert directly from a traditional volume to a FlexVol volume, or from a FlexVol volume to a traditional volume. You must create a new volume of the desired type and then move the data to the new volume using `ndmpcopy`.
- If you move the data to another volume on the same storage system, remember that this requires the system to have enough storage to contain both copies of the volume.
- Snapshots on the original volume are unaffected by the migration, but they are not valid for the new volume.

NetApp offers assistance

NetApp Professional Services staff, including Professional Services Engineers (PSEs) and Professional Services Consultants (PSCs) are trained to assist customers with converting volume types and migrating data, among other services. For more information, contact your local NetApp Sales representative, PSE, or PSC.

Migrating a traditional volume to a FlexVol volume

The following procedure describes how to migrate from a traditional volume to a FlexVol volume. If you are migrating your root volume, you can use the same procedure, including the steps that are specific to migrating a root volume.

To migrate a traditional volume to a FlexVol volume, complete the following steps.

Step	Action

1	Determine the size requirements for the new FlexVol volume. Enter the following command to determine the amount of space your current volume uses:
	`df -Ah [vol_name]`
	Example: `df -Ah vol0`
	Result: The following output is displayed.
	`Aggregate total used avail capacity`
	`vol0 24GB 1434MB 22GB 7%`
	`vol0/.snapshot 6220MB 4864MB 6215MB 0%`
	Root volume: If the new FlexVol volume is going to 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 "Understanding the Root Volume" chapter in the *System Administration Guide*.

2	You can use an existing aggregate or you can create a new one to contain the new FlexVol volume. To determine if an existing aggregate is large enough to contain the new FlexVol volume, enter the following command: **`df -Ah`** **Result:** All of the existing aggregates are displayed.

Step	Action
3	If needed, create a new aggregate by entering the following command: **`aggr create aggr_name disk-list`** **Example:** `aggr create aggrA 8@144` **Result:** An aggregate called *aggrA* is created with eight 144-GB disks. The default RAID level is RAID-DP, so two disks will be used for parity (one parity disk and one dParity disk). The aggregate size will be 1,128 GB. If you want to use RAID4, and use one less parity disk, enter the following command: **`aggr create aggrA -t raid4 8@144`**
4	If you want to use the new FlexVol volume to have the same name as the old traditional volume, you must rename the existing traditional root volume before creating the new FlexVol volume. Do this by entering the following command: **`aggr rename vol_name new_vol_name`** **Example:** `aggr rename vol0 vol0trad`
5	Create the new FlexVol volume in the containing aggregate. For more information about creating FlexVol volumes, see "Creating FlexVol volumes" on page 225. **`vol create vol_name aggr_name [-s {volume | file | none}] size`** **Example:** `vol create vol0 aggrA 90g` **Root volume:** You must use the (default) `volume` space guarantee for root volumes, because it ensures that writes to the volume do not fail due to a lack of available space in the containing aggregate.
6	Confirm the size of the new FlexVol volume by entering the following command: **`df -h vol_name`**

Step	Action
7	Shut down any applications that use the data to be migrated. Make sure that all data is unavailable to clients and that all files to be migrated are closed.
8	Enable the `ndmpd.enable` option by entering the following command: **`options ndmpd.enable on`**
9	Migrate the data by entering the following command: **`ndmpcopy old_vol_name new_vol_name`** **Example:** `ndmpcopy /vol/vol0trad /vol/vol0` For more information about using `ndmpcopy`, see the *Data Protection Tape Backup and Recovery Guide*.
10	Verify that the `ndmpcopy` operation completed successfully by verifying that the data was replicated correctly.
11	If you are migrating your root volume, make the new FlexVol volume the root volume by entering the following command: **`vol options vol_name root`** **Example:** `vol options vol0 root`
12	If you are migrating your root volume, reboot the storage system.
13	Update the clients to point to the new FlexVol volume. In a CIFS environment, follow these steps: 1. Point CIFS shares to the new FlexVol volume. 2. Update the CIFS maps on the client machines so that they point to the new FlexVol volume. In an NFS environment, follow these steps: 1. Point NFS exports to the new FlexVol volume. 2. Update the NFS mounts on the client machines so that they point to the new FlexVol volume.

Step	Action
14	Make sure all clients can see the new FlexVol volume and read and write data. To test whether data can be written, complete the following steps: **1.** Create a new folder. **2.** Verify that the new folder exists. **3.** Delete the new folder.
15	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.
16	If quotas were used with the traditional volume, configure the quotas on the new FlexVol volume.
17	Take a snapshot of the target volume and create a new snapshot schedule as needed. For more information about taking snapshots, see the *Data Protection Online Backup and Recovery Guide*.
18	When you are confident the volume migration was successful, you can take the original volume offline or destroy it. **Attention** You should preserve the original volume and its snapshots until the new FlexVol volume has been stable for some time.

Migrating a FlexVol volume to a traditional volume

To migrate a FlexVol volume to a traditional volume, complete the following steps.

Step	Action

1	Determine the size requirements for the new traditional volume. Enter the following command to determine the amount of space your current volume uses:
	df -Ah [*vol_name*]
	Example: df -Ah vol_users
	Result: The following output is displayed.
	`Aggregate total used avail capacity`
	`users 94GB 1434GB 22GB 6%`
	`users/.snapshot 76220MB 74864MB 6215MB 0%`

2	Create the traditional volume that will replace the FlexVol volume by entering the following command: **aggr create *vol_name disk-list*** **Example:** aggr create users 3@144
3	Confirm the size of the new traditional volume by entering the following command: **df -h *vol_name***
4	Shut down the applications that use the data to be migrated. Make sure that all data is unavailable to clients and that all files to be migrated are closed.
5	Enable the ndmpd.enable option by entering the following command: **options ndmpd.enable on**
6	Migrate the data using the ndmpcopy command. For more information about using ndmpcopy, see the *Data Protection Tape Backup and Recovery Guide*.
7	Verify that the ndmpcopy operation completed successfully by checking that the data has been replicated correctly.

Step	Action
8	Update the clients to point to the new volume. In a CIFS environment, follow these steps: **1.** Point CIFS shares to the new volume. **2.** Update the CIFS maps on the client machines so that they point to the new volume. **3.** Repeat steps 1 and 2 for each new volume. In an NFS environment, follow these steps: **1.** Point NFS exports to the new volume. **2.** Update the NFS mounts on the client machines so that they point to the new volume. **3.** Repeat steps 1 and 2 for each new volume.
9	Make sure all clients can see the new traditional volume and read and write data. To test whether data can be written, complete the following steps: **1.** Create a new folder. **2.** Verify that the new folder exists. **3.** Delete the new folder. **4.** Repeat steps 1 through 3 for each new volume.
10	If quotas were used with the FlexVol volume, configure the quotas on the new volume.
11	Take a snapshot of the target volume and create a new snapshot schedule as needed. For more information about taking snapshots, see the *Data Protection Online Backup and Recovery Guide*.

Step	Action
12	When you are confident the volume migration was successful, you can take the source volume offline or destroy it. **Attention** You should preserve the original volume and its snapshots until the new volume has been stable for some time.

Managing duplicate volume names

How duplicate volume names can occur

Data ONTAP does not support having two volumes with the same name on the same storage system. However, certain events can cause this to happen, as outlined in the following list:

- You copy an aggregate using the `aggr copy` command, and when you bring the target aggregate online, there are one or more volumes on the destination system with the duplicated names.
- You move an aggregate from one storage system to another by moving its associated disks, and there is another volume on the destination system with the same name as a volume contained by the aggregate you moved.
- You move a traditional volume from one storage system to another by moving its associated disks, and there is another volume on the destination system with the same name.
- Using SnapMover, you migrate a vFiler unit that contains a volume with the same name as a volume on the destination system.

How Data ONTAP handles duplicate volume names

When Data ONTAP senses 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, the newly copied volume might be named vol1(1).

Duplicate volumes should be renamed as soon as possible

You might consider a volume name such as vol1(1) to be acceptable. However, it is important that you 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 will prevent 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.

Managing volume languages

About volumes and languages

Every volume has a language. The storage system uses a character set appropriate to the language for the following items on that volume:

- File names
- File access

The language of the root volume is used for the following items:

- System name
- CIFS share names
- NFS user and group names
- CIFS user account names
- Domain name
- Console commands and command output
- Access from CIFS clients that don't support Unicode
- Reading the following files:
  - /etc/quotas
  - /etc/usermap.cfg
  - the home directory definition file

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.

Note

Names of the following objects must be in ASCII characters:

- Qtrees
- Snapshots
- Volumes

Viewing the language list online

It might be useful to view the list of languages before you choose one for a volume. To view the list of languages, complete the following step.

Step	Action
1	Enter the following command: `vol lang`

Choosing a language for a volume

To choose a language for a volume, complete the following step.

Step	Action	
1	**If the volume is accessed using...**	**Then...**
	NFS Classic (v2 or v3) only	Do nothing; the language does not matter.
	NFS Classic (v2 or v3) and CIFS	Set the language of the volume to the language of the clients.
	NFS v4, with or without CIFS	Set the language of the volume to *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.

Displaying volume language use

You can display a list of volumes with the language each volume is configured to use. This is useful for the following kinds of decisions:

- How to match the language of a volume to the language of clients
- Whether to create a volume to accommodate clients that use a language for which you don't have a volume
- Whether to change the language of a volume (usually from the default language)

To display which language a volume is configured to use, complete the following step.

Step	Action

1	Enter the following command:
	`vol status [vol_name] -l`
	vol_name is the name of the volume about which you want information. Leave out *vol_name* to get information about every volume on the storage system.
	Result: Each row of the list displays the name of the volume, the language code, and the language, as shown in the following sample output.
	`Volume Language`
	`vol0  ja (Japanese euc-j)`

Changing the language for a volume

Before changing the language that a volume uses, be sure you read and understand the section titled "About volumes and languages" on page 253.

To change the language that a volume uses to store file names, complete the following steps.

Step	Action
1	Enter the following command: **`vol lang vol_name language`** *vol_name* is the name of the volume about which you want information. *language* is the code for the language you want the volume to use.
2	Enter the following command to verify that the change has successfully taken place: **`vol status vol_name -l`** *vol_name* is the name of the volume whose language you changed.

Determining volume status and state

Volume states

A volume can be in one of the following three states, sometimes called mount states:

- `online`—Read and write access is allowed.
- `offline`—Read or write access is not allowed.
- `restricted`—Some operations, such as copying volumes and parity reconstruction, are allowed, but data access is not allowed.

Volume status

A volume can have one or more of the following statuses:

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.

- `copying`

  The volume is currently the target volume of active `vol copy` or `snapmirror` operations.
- `degraded`

  The volume's containing aggregate has at least one degraded RAID group that is not being reconstructed.
- `flex`

  The volume is a FlexVol volume.
- `flexcache`

  The volume is a FlexCache volume. For more information about FlexCache volumes, see "Managing FlexCache volumes" on page 268.
- `foreign`

  Disks used by the volume's containing aggregate were moved to the current storage system from another storage system.
- `growing`

  Disks are in the process of being added to the volume's containing aggregate.

- `initializing`

  The volume or its containing aggregate are in the process of being initialized.
- `invalid`

  The volume does not contain a valid file system. This typically happens only after an aborted `vol copy` operation.
- `ironing`

  A WAFL consistency check is being performed on the volume's containing aggregate.
- `mirror degraded`

  The volume's containing aggregate is a mirrored aggregate, and one of its plexes is offline or resyncing.
- `mirrored`

  The volume's containing aggregate is mirrored and all of its RAID groups are functional.
- `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 RAID-0 (no parity) RAID groups (V-Series and NetCache® systems only).
- `raid4`

  The volume's containing aggregate consists of RAID4 RAID groups.
- `raid_dp`

  The volume's containing aggregate consists of RAID-DP (Double Parity) RAID groups.
- `reconstruct`

  At least one RAID group in the volume's containing aggregate is being reconstructed.
- `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 a volume appears in this state, contact technical support.
- `verifying`

  A RAID mirror verification operation is currently being run on the volume's containing aggregate.
- `wafl inconsistent`

  The volume or its containing aggregate has been marked corrupted. If a volume appears in this state, contact technical support.

Determining the state and status of volumes

To determine what state a volume is in, and what status currently applies to it, complete the following step.

Step	Action

1	Enter the following command:
	`vol status`
	Result: Data ONTAP displays a concise summary of all the volumes in the storage system. The *State* column displays whether the volume is online, offline, or restricted. The *Status* column displays the volume's RAID level, whether the volume is a FlexVol or traditional volume, and any status other than normal (such as partial or degraded).
	Example:
	`> vol status`
	`Volume State Status Options`
	`vol0 online raid4, flex root,guarantee=volume`
	`volA online raid_dp, trad`
	`mirrored`
	Note
	To see a complete list of all options, including any that are off or not set for this volume, use the `-v` flag with the `vol status` command.

When to take a volume offline

You can take a volume offline and make it unavailable to the storage system. You do this for the following reasons:

- To perform maintenance on the volume
- To move a volume to another storage system
- To destroy a volume

Note

You cannot take the root volume offline.

Taking a volume offline

To take a volume offline, complete the following step.

Step	Action
1	Enter the following command: **`vol offline vol_name`** *vol_name* is the name of the volume to be taken offline. **Note** 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. For more information, see "Bringing a volume online in an overcommitted aggregate" on page 290.

When to make a volume restricted

When you make a volume restricted, it is available for only a few operations. You do this for the following reasons:

- To copy a volume to another volume

  For more information about volume copy, see the *Data Protection Online Backup and Recovery Guide*.
- To perform a level-0 SnapMirror operation

  For more information about SnapMirror, see the *Data Protection Online Backup and Recovery Guide*.

Note

When you restrict a FlexVol volume, it relinquishes any unused space that is 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.

For more information, see "Bringing a volume online in an overcommitted aggregate" on page 290.

Restricting a volume

To restrict a volume, complete the following step.

Step	Action
1	Enter the following command: **`vol restrict vol_name`**

Bringing a volume online

You bring a volume back online to make it available to the storage system after you deactivated that volume.

Note

If you bring a FlexVol volume online into an aggregate that does not have sufficient free space in the aggregate to fulfill the space guarantee for that volume, this command fails.

For more information, see "Bringing a volume online in an overcommitted aggregate" on page 290.

To bring a volume back online, complete the following step.

Step	Action
1	Enter the following command: **`vol online vol_name`** *vol_name* is the name of the volume to reactivate. **Attention** If the volume is inconsistent, the command prompts you for confirmation. If you bring an inconsistent volume online, it might suffer further file system corruption.

Renaming volumes

Renaming a volume

To rename a volume, complete the following steps.

Step	Action
1	Enter the following command: **`vol rename vol_name new-name`** *vol_name* is the name of the volume you want to rename. *new-name* is the new name of the volume. **Result:** The following events occur: ◆ The volume is renamed. ◆ If NFS is in use and the `nfs.exports.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.
2	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 volumes

About destroying volumes

There are two reasons to destroy a volume:

- You no longer need the data it contains.
- You copied the data it contains elsewhere.

When you destroy a traditional volume: You also destroy the traditional volume's dedicated containing aggregate. This converts its parity disk and all its data disks back into hot spares. You can then use them in other aggregates, traditional volumes, or storage systems.

When you destroy a FlexVol volume: All the disks included in its containing aggregate remain assigned to that containing aggregate.

Attention

If you destroy a volume, all the data in the volume is destroyed and no longer accessible.

Destroying a volume

To destroy a volume, complete the following steps.

Step	Action
1	Take the volume offline by entering the following command: **`vol offline vol_name`** *vol_name* is the name of the volume that you intend to destroy.

Step	Action
2	Enter the following command to destroy the volume: **`vol destroy vol_name`** *vol_name* is the name of the volume that you intend to destroy. **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-swapable spare disks.
3	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.

Increasing the maximum number of files in a volume

About increasing the maximum number of files

The storage system automatically sets the maximum number of files for a newly created volume based on the amount of disk space in the volume. The storage system increases the maximum number of files when you add a disk to a volume. The number set by the storage system never exceeds 33,554,432 unless you set a higher number with the `maxfiles` command. This prevents a storage system with terabytes of storage from creating a larger than necessary inode file.

If you get an error message telling you that you are out of inodes (data structures containing information about files), you can use the `maxfiles` command to increase the number. This should only be necessary if you are using an unusually large number of small files, or if your volume is extremely large.

Attention

Use caution when increasing the maximum number of files, because after you increase this number, you can never decrease it. As new files are created, the file system consumes the additional disk space required to hold the inodes for the additional files; there is no way for the storage system to release that disk space.

Increasing the maximum number of files allowed on a volume

To increase the maximum number of files allowed on a volume, complete the following step.

Step	Action
1	Enter the following command: **`maxfiles vol_name max`** *vol_name* is the volume whose maximum number of files you are increasing. *max* is the maximum number of files. **Note** Inodes are added in blocks, and five percent of the total number of inodes is reserved for internal use. If the requested increase in the number of files is too small to require a full inode block to be added, the `maxfiles` value is not increased. If this happens, repeat the command with a larger value for *max*.

Displaying the number of files in a volume

To see how many files are in a volume and the maximum number of files allowed on the volume, complete the following step.

Step	Action
1	Enter the following command: **`maxfiles vol_name`** *vol_name* is the volume whose maximum number of files you are increasing. **Result:** A display like the following appears: `Volume home: maximum number of files is currently 120962 (2872 used)` **Note** The value returned reflects only the number of files that can be created by users; the inodes reserved for internal use are not included in this number.

Reallocating file and volume layout

About reallocation

If your volumes contain large files or LUNs that store information that is frequently accessed and revised (such as databases), the layout of your data can become suboptimal. Additionally, when you add disks to an aggregate, your data is no longer evenly distributed across all of the disks. The Data ONTAP `reallocate` commands allow you to reallocate the layout of files, LUNs or entire volumes for better data access.

For more information

For more information about the reallocation commands, see your *Block Access Management Guide*, keeping in mind that for reallocation, files are managed exactly the same as LUNs.

Managing FlexCache volumes

About FlexCache volumes

A FlexCache volume is a sparsely populated volume on a local (caching) storage system that is backed by a volume on a different, possibly remote, (origin) storage system. A sparsely populated volume, sometimes called a sparse volume, provides access to all data in the origin volume without requiring that the data be physically in the sparse volume.

You use FlexCache volumes to speed up access to remote data, or to offload traffic from heavily accessed volumes. Because the cached data must be ejected when the data is changed, FlexCache volumes work best for data that does not change often.

About this section

This section contains the following topics:

- "How FlexCache volumes work" on page 269
- "Sample FlexCache deployments" on page 275
- "Creating FlexCache volumes" on page 277
- "Sizing FlexCache volumes" on page 279
- "Administering FlexCache volumes" on page 281

How FlexCache volumes work

Direct access to cached data

When a client requests data from the FlexCache volume, the data is read through the network from the origin system and cached on the FlexCache volume. Subsequent requests for that data are then served directly from the FlexCache volume. In this way, clients in remote locations are provided with direct access to cached data. This improves performance when data is accessed repeatedly, because after the first request, the data no longer has to travel across the network.

FlexCache license requirement

You must have the `flex_cache` license installed on the caching system before you can create FlexCache volumes. For more information about licensing, see the *System Administration Guide*.

Types of volumes you can use

A FlexCache volume must always be a FlexVol volume. FlexCache volumes can be created in the same aggregate as regular FlexVol volumes.

The origin volume can be a FlexVol volume or a traditional volume; it can also be a SnapLock volume. The origin volume cannot be a FlexCache volume itself, nor can it be a qtree.

Cache objects

The following objects can be cached in a FlexCache volume:

- Files
- Directories
- Symbolic links

Note

In this document, the term *file* is used to refer to all of these object types.

File attributes are cached

When a data block from a specific file is requested from a FlexCache volume, then the attributes of that file are cached, and that file is considered to be cached. This is true even if not all of the data blocks that make up that file are present in the cache.

Cache consistency

Cache consistency for FlexCache volumes is achieved using three primary techniques: *delegations, attribute cache timeouts,* and *write operation proxy.*

Delegations: When data from a particular file is retrieved from the origin volume, the origin volume can give a delegation for that file to the caching volume. If that file is changed on the origin volume, whether from another caching volume or through direct client access, then the origin volume revokes the delegation for that file with all caching volumes that have that delegation. You can think of a delegation as a contract between the origin volume and the caching volume; as long as the caching volume has the delegation, the file has not changed.

Note

Delegations can cause a small performance decrease for writes to the origin volume, depending on the number of caching volumes holding delegations for the file being modified.

Delegations are not always used. The following list outlines situations when delegations cannot be used to guarantee that an object has not changed:

- Objects other than regular files do not use delegations

  Delegations are not used for any objects other than regular files. Directories, symbolic links, and other objects have no delegations.

- When connectivity is lost

  If connectivity is lost between the caching and origin systems, then delegations cannot be honored and must be considered to be revoked.

- When the maximum number of delegations has been reached

  If the origin volume cannot store all of its delegations, it might revoke an existing delegation to make room for a new one.

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. However, if no delegation for the file exists, then it is considered valid for a specified length of time, called the attribute cache timeout. As long as a file is considered valid, if a client reads from that file and the requested data blocks are cached, the read request is fulfilled without any access to the origin volume.

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 verifies that the attributes of the file have not changed on the origin system. Then one of the following actions is taken:

- If the attributes of the file have not changed since the file was cached, then the requested data is either 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 attributes of the file have changed, the file is marked as invalid in the cache. Then the requested data blocks are read from the origin system, 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 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

To prevent clients from ever getting stale data, you can set the attribute cache timeout to zero. However, this will negatively affect your caching performance, because then 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 volume option names and default values are outlined in the following table.

Volume option name	Description	Default Value
`acdirmax`	Attribute cache timeout for directories	`30s`
`acregmax`	Attribute cache timeout for regular files	`30s`
`acsymmax`	Attribute cache timeout for symbolic links	`30s`
`actimeo`	Attribute cache timeout for all objects	`30s`

For more information about modifying these options, see the na_vol(1) man page.

Write operation proxy: If the client modifies the file, that operation is proxied through to the origin system, and the file is ejected from the cache. This also changes the attributes of the file on the origin volume, so any other FlexCache volume that has that data cached will re-request the data once the attribute cache timeout is reached and a client requests that data.

Cache hits and misses

When a client makes 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 cached and no verify is required; the request is fulfilled locally and no access to the origin system is made.
- Hit-Verify

  The requested data is cached but the verification timeout has been exceeded, so the file attributes are verified against the origin system. No data is requested from the origin system.

If data is requested that is not currently on the FlexCache volume, or if that data has changed since it was cached, the caching system loads the data from the origin system 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 in the cache; it is read from the origin system and cached.
- Miss-Verify

  The requested data is cached, but the file attributes have changed since the file was cached; the file is ejected from the cache and the requested data is read from the origin system and cached.

Limitations of FlexCache volumes

There are certain limitations of the FlexCache feature, for both the caching volume and for the origin volume.

Limitations of FlexCache caching volumes: You cannot use the following capabilities on FlexCache volumes (these limitations do *not* apply to the origin volumes):

- Client access using any protocol other than NFSv2 or NFSv3
- Snapshot creation
- SnapRestore
- SnapMirror (qtree or volume)
- SnapVault
- FlexClone volume creation
- `ndmp`
- Quotas
- Qtrees
- `vol copy`
- Creation of FlexCache volumes in any vFiler unit other than vFiler0

Limitations of FlexCache origin volumes: You cannot perform the following operations on a FlexCache origin volume or storage system without rendering all FlexCache volumes backed by that origin volume unusable:

- You cannot move an origin volume between vFiler units or to vFiler0 using any of the following commands:
  - `vfiler move`
  - `vfiler add`
  - `vfiler remove`
  - `vfiler destroy`

  If you want to perform these operations on the origin volume, you can delete all FlexCache volumes backed by that volume, perform the operation, and then recreate the FlexCache volumes.

  Note

  You can use SnapMover (`vfiler migrate`) to migrate an origin volume without having to recreate any FlexCache volumes backed by that volume.

- You cannot use a FlexCache origin volume as the destination of a `snapmirror migrate` command.

  If you want to perform a `snapmirror migrate` operation to a FlexCache origin volume, you must delete and recreate all FlexCache volumes backed by that volume after the migrate operation completes.

- You cannot change the IP address of the origin NetApp system.

  If you must change the IP address of the origin system, you can delete all FlexCache volumes backed by the volumes on that system, change the IP address, then recreate the FlexCache volumes.

What happens when connectivity to origin system is lost

If connectivity between the caching and origin systems is lost after a FlexCache volume is created, any data access that does not require access to the origin system succeeds. However, any operation that requires access to the origin volume, either because the requested data is not cached or because its attribute cache timeout has been exceeded, hangs until connectivity is reestablished.

Sample FlexCache deployments

WAN or LAN deployment

A FlexCache volume can be deployed in a WAN configuration or a LAN configuration.

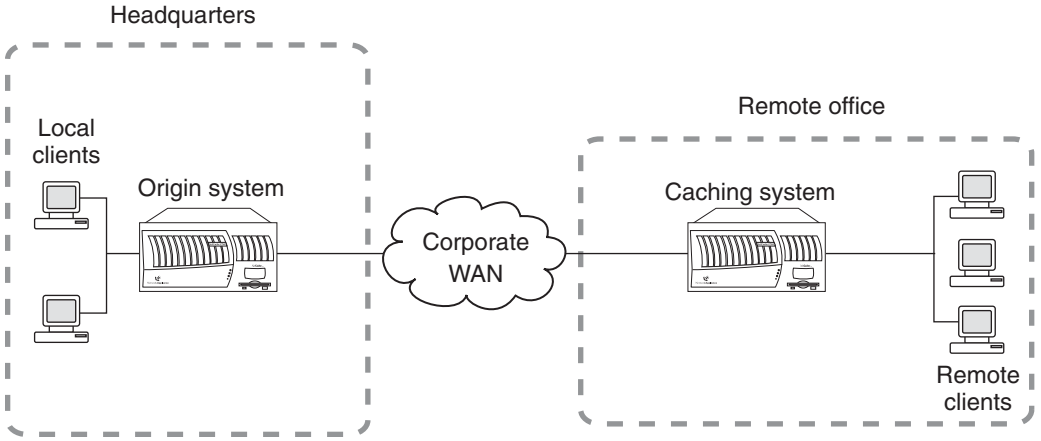
WAN deployment: In a WAN deployment, the FlexCache volume is remote from the data center. As clients request data, the FlexCache volume caches popular data, giving the end user faster access to information.

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.

WAN deployment

In a WAN deployment, the FlexCache volume is placed as close as possible to the remote office. Client requests are then explicitly directed to the storage system. 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.

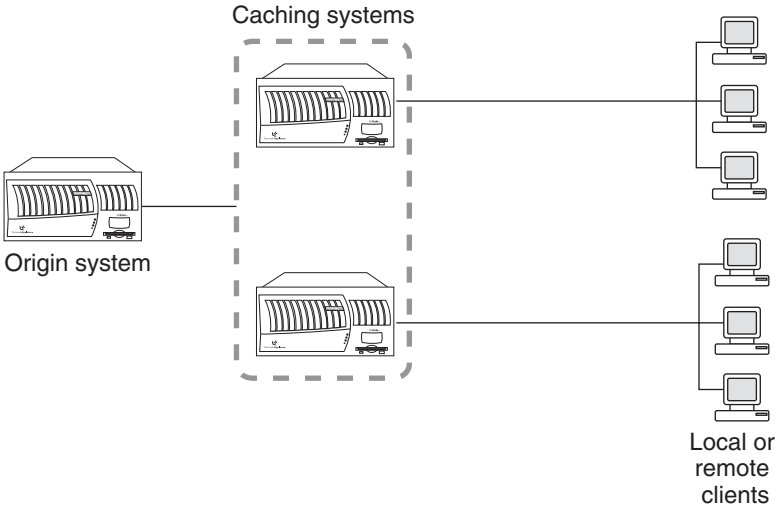
The following diagram shows a typical FlexCache WAN deployment.



LAN deployment

In a LAN deployment, a FlexCache volume is used to offload busy data servers. Frequently accessed data, or “hot objects” are replicated and cached by the FlexCache volume. This saves network bandwidth, reduces latency, and improves storage use, because only the most frequently used data is moved and stored.

The following example illustrates a typical LAN deployment.



Creating FlexCache volumes

Before creating a FlexCache volume

Before creating a FlexCache volume, ensure that you have the following configuration options set correctly:

- `flex_cache` license installed on the caching system
- `flexcache.access` option on origin system set to allow access from caching system

  Note

  If the origin volume is in a vFiler unit, set this option for the vFiler context.

  For more information about this option, see the na_protocolaccess(8) man page.

- `flexcache.enable` option on the origin system set to `on`

  Note

  If the origin volume is in a vFiler unit, set this option for the vFiler context.

- NFS licensed and enabled for the caching system

  Note

  FlexCache volumes function correctly without an NFS license on the origin system. However, for maximum caching performance, you should install a license for NFS on the origin system also.

- Both the caching and origin systems running Data ONTAP 7.0.1 or later

Creating a FlexCache volume

To create a FlexCache volume, complete the following steps.

Step	Action
1	Ensure that your options are set correctly as outlined in “Before creating a FlexCache volume” on page 277.

Step	Action
2	Enter the following command: **vol create *cache_vol aggr size*{k|m|g|t} -S *origin:source_vol*** *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. For example, you would enter 20m to indicate twenty megabytes. If you do not specify a unit, size is taken as bytes and rounded up to the nearest multiple of 4 KB. **Note** Because FlexCache volumes are sparsely populated, you can make the FlexCache volume smaller than the source volume. However, the larger the FlexCache volume is, the better caching performance it provides. For more information about sizing FlexCache volumes, see "Sizing FlexCache volumes" on page 279. *origin* is the name of the origin storage system *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 100-MB FlexCache volume called newcachevol, in the aggregate called aggr1, with a source volume vol1 on storage system corp_toaster. `vol create newcachevol aggr1 100M -S corp_toaster:vol1`

Sizing FlexCache volumes

About sizing FlexCache volumes

FlexCache volumes can be smaller than their origin volumes. However, making your FlexCache volume too small can negatively impact your caching performance. When the FlexCache volume begins to fill up, it flushes old data to make room for newly requested data. When that old data is requested again, it must be retrieved from the origin volume.

For best performance, set all FlexCache volumes to the size of their containing aggregate. For example, if you have two FlexCache volumes sharing a single 2TB aggregate, you should set the size of both FlexCache volumes to 2TB. This approach provides the maximum caching performance for both volumes, because the FlexCache volumes manage the shared space to accelerate the client workload on both volumes. The aggregate should be large enough to hold all of the clients' working sets.

FlexCache volumes and space management

FlexCache volumes do not use space management in the same manner as regular FlexVol volumes. When you create a FlexCache volume of a certain size, that volume will not grow larger than that size. However, only a certain amount of space is preallocated for the volume. The amount of disk space allocated for a FlexCache volume is determined by the value of the `flexcache_min_reserved` volume option.

Note

The default value for the `flexcache_min_reserved volume` option is 100 MB. You should not need to 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 can now be used by other volumes in the aggregate; this is true for 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.

Space allocation for multiple volumes in the same aggregate

You can have multiple FlexCache volumes in the same aggregate; you can also have regular FlexVol volumes in the same aggregate as your FlexCache volumes.

Multiple FlexCache volumes in the same aggregate: When you put multiple FlexCache volumes in the same aggregate, they can each be sized to be as large as the aggregate permits. This is because only the amount of space specified by the `flexcache_min_reserved` volume option is actually reserved for each one. The rest of the space is allocated as needed. This means that a "hot" FlexCache volume, or one that is receiving more data accesses, is permitted to take up more space, while a FlexCache volume that is not being accessed as often will gradually be reduced in size.

FlexVol volumes and FlexCache volumes in the same aggregate: If you have regular FlexVol volumes in the same aggregate as your FlexCache volumes, and you start to fill up the aggregate, the FlexCache volumes can lose some of their unreserved space (only if they are not currently using it). In this case, when the FlexCache volume needs to fetch a new data block and it does not have enough free space to accommodate it, a data block must be ejected from one of the FlexCache volumes to make room for the new data block.

If this situation causes too many cache misses (as shown by the `flexcache stats` command), you can add more space to your aggregate or move some of your data to another aggregate.

Using the df command with 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.

Administering FlexCache volumes

Viewing FlexCache statistics

Data ONTAP provides statistics about FlexCache volumes to help you understand the access patterns and administer the FlexCache volumes effectively. You can get statistics for your FlexCache volumes using the following commands:

- `flexcache stats` (client and server statistics)
- `nfsstat` (client statistics only)

For more information about these commands, see the na_flexcache(1) and nfsstat(1) man pages.

Client (caching system) statistics: You can use client statistics to see how how many operations are being served by the FlexCache rather than the origin system. A large number of cache misses after the FlexCache volume has had time to become populated may indicate that the FlexCache volume is too small and data is being discarded and fetched again later.

To view client FlexCache statistics, you use the `-C` option of the `flexcache stats` command on the *caching* system.

You can also view the `nfs` statistics for your FlexCache volumes using the `-C` option for the `nfsstat` command.

Server (origin system) statistics: You can use server statistics to see how much load is hitting the origin volume and which clients are causing that load. This can be useful if you are using the LAN deployment to offload an overloaded volume, and you want to make sure that the load is evenly distributed among the caching volumes.

To view server statistics, you use the `-S` option of the `flexcache stats` command on the *origin* system.

Note

You can also view the server statistics by client, using the `-c` option of the `flexcache stats` command. The `flexcache.per_client_stats` option must be set to On.

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.

LUNs in FlexCache volumes

Although you cannot use SAN access protocols to access FlexCache volumes, you might want to cache a volume that contains LUNs along with other data. When you attempt to access a directory in a FlexCache volume that contains a LUN file, the command sometimes returns "stale NFS file handle" for the LUN file. If you get that error message, repeat the command. In addition, if you use the `fstat` command on a LUN file, `fstat` always indicates that the file is not cached. This is expected behavior.

Space management for volumes and files

What space management is

The space management capabilities of Data ONTAP allow you to configure your storage systems to provide the storage availability required by the users and applications accessing the system, while using your available storage as effectively as possible.

Data ONTAP provides space management using the following capabilities:

- Space guarantees

  This capability is available only for FlexVol volumes.

  For more information, see "Space guarantees" on page 286.
- Space reservations

  For more information, see "Space reservations" on page 292 and your *Block Access Management Guide*.
- Fractional reserve

  This capability is an extension of space reservations that is new for Data ONTAP 7.0.

  For more information, see "Fractional reserve" on page 294 and your *Block Access Management Guide*.
- Space management policy

  This capability allows you to automatically reclaim space for a FlexVol volume when that FlexVol volume is nearly full.

  For more information, see "Space management policies" on page 295 and your *Block Access Management Guide*.

Space management and files

Space reservations, fractional reserve, and the space management policy are designed primarily for use with LUNs. Therefore, they are explained in greater detail in your *Block Access Management Guide*. If you want to use these space management capabilities for files, consult those guides, keeping in mind that files are managed by Data ONTAP exactly the same as LUNs, except that space reservations are enabled for LUNs by default, whereas space reservations must be explicitly enabled for files.

What kind of space management to use

The following table can help you determine which space management capabilities best suit your requirements.

If…	Then use…	Typical usage	Notes
◆ You want management simplicity ◆ You have been using a version of Data ONTAP earlier than 7.0 and want to continue to manage your space the same way	◆ FlexVol volumes with space guarantee = `volume` ◆ Traditional volumes	NAS file systems	This is the easiest option to administer. As long as you have sufficient free space in the volume, writes to any file in this volume will always succeed. For more information about space guarantees, see "Space guarantees" on page 286.
◆ Writes to certain files must always succeed ◆ You want to overcommit your aggregate	◆ FlexVol volumes with space guarantee = `file` OR ◆ Traditional volume AND space reservation enabled for files that require writes to succeed	◆ LUNs ◆ Databases	This option enables you to guarantee writes to specific files. For more information about space guarantees, see "Space guarantees" on page 286. For more information about space reservations, see "Space reservations" on page 292 and your *Block Access Management Guide*.

If…	Then use…	Typical usage	Notes
◆ You need even more effective storage usage than file space reservation provides ◆ You actively monitor available space on your volume and can take corrective action when needed ◆ Snapshots are short-lived ◆ Your rate of data overwrite is relatively predictable and low	◆ FlexVol volumes with space guarantee = `volume` OR ◆ Traditional volume AND Space reservation on for files that require writes to succeed AND Fractional reserve < 100%	◆ LUNs (with active space monitoring) ◆ Databases (with active space monitoring)	With fractional reserve <100%, it is possible to use up all available space, even with space reservations on. Before enabling this option, be sure either that you can accept failed writes or that you have correctly calculated and anticipated storage and snapshot usage. For more information, see "Fractional reserve" on page 294 and your *Block Access Management Guide*.
◆ You want to overcommit your aggregate ◆ You actively monitor available space on your aggregate and can take corrective action when needed	FlexVol volumes with space guarantee = `none`	◆ Storage providers who need to provide storage that they know will not immediately be used ◆ Storage providers who need to allow available space to be dynamically shared between volumes	With an overcommitted aggregate, writes can fail due to insufficient space. For more information about aggregate overcommitment, see "Aggregate overcommitment" on page 289.

Space guarantees

What space guarantees are

Space guarantees on a FlexVol volume ensure that writes to a specified FlexVol volume or writes to files with space reservations enabled do not fail because of lack of available space in the containing aggregate.

Other operations such as creation of snapshots or new volumes in the containing aggregate can occur only if there is enough available uncommitted space in that aggregate; other operations are restricted from using space already committed to another volume.

When the uncommitted space in an aggregate is exhausted, only writes to volumes or files in that aggregate with space guarantees are guaranteed to succeed.

- A space guarantee of `volume` preallocates space in the aggregate for the volume. The preallocated space cannot be allocated to any other volume in that aggregate.

  The space management for a FlexVol volume with space guarantee of `volume` is equivalent to a traditional volume, or all volumes in versions of Data ONTAP earlier than 7.0.
- A space guarantee of `file` preallocates space in the volume so that any file in the volume with space reservation enabled can be completely rewritten, even if its blocks are pinned for a snapshot.

  For more information on file space reservation see "Space reservations" on page 292.
- A FlexVol volume with a space guarantee of `none` reserves no extra space; writes to LUNs or files contained by that volume could fail if the containing aggregate does not have enough available space to accommodate the write.

  Note

  Because out-of-space errors are unexpected in a CIFS environment, do not set space guarantee to `none` for volumes accessed using CIFS.

Space guarantee is an attribute of the volume. It is persistent across storage system reboots, takeovers, and givebacks, but it does not persist through reversions to versions of Data ONTAP earlier than 7.0.

Space guarantees and volume status

Space guarantees are honored only for online volumes. If you take a volume offline, any committed but unused space for that volume becomes available for other volumes in that aggregate. When you bring that volume back online, if there is not sufficient available space in the aggregate to fulfill its space guarantees, you must use the force (`-f`) option, and the volume's space guarantees are disabled.

For more information, see "Bringing a volume online in an overcommitted aggregate" on page 290.

Traditional volumes and space management

Traditional volumes provide the same space guarantee as FlexVol volumes with space guarantee of `volume`. To guarantee that writes to a specific file in a traditional volume will always succeed, you need to enable space reservations for that file. (LUNs have space reservations enabled by default.)

For more information about space reservations, see "Space reservations" on page 292.

Specifying space guarantee at FlexVol volume creation time

To specify the space guarantee for a volume at creation time, complete the following steps.

Note

To create a FlexVol volume with space guarantee of `volume`, you can ignore the `guarantee` parameter, because `volume` is the default.

Step	Action
1	Enter the following command: **`vol create f_vol_name aggr_name -s {volume|file|none} size{k|m|g|t}`** *f_vol_name* is the name for the new FlexVol volume (without the /vol/ prefix). This name must be different from all other volume names on the storage system. *aggr_name* is the containing aggregate for this FlexVol volume. `-s` specifies the space guarantee to be used for this volume. The possible values are `{volume|file|none}`. The default value is `volume`. *size* `{k|m|g|t}` specifies the maximum volume size in kilobytes, megabytes, gigabytes, or terabytes. For example, you would enter `4m` to indicate four megabytes. If you do not specify a unit, size is considered to be in bytes and rounded up to the nearest multiple of 4 KB.
2	To confirm that the space guarantee is set, enter the following command: **`vol options f_vol_name`**

Changing space guarantee for existing volumes

To change the space guarantee for an existing FlexVol volume, complete the following steps.

Step	Action
1	Enter the following command: **`vol options f_vol_name guarantee guarantee_value`** *f_vol_name* is the name of the FlexVol volume whose space guarantee you want to change. *guarantee_value* is the space guarantee you want to assign to this volume. The possible values are `volume`, `file`, and `none`. **Note** If there is insufficient space in the aggregate to honor the space guarantee you want to change to, the command succeeds, but a warning message is printed and the space guarantee for that volume is disabled.
2	To confirm that the space guarantee is set, enter the following command: **`vol options f_vol_name`**

Aggregate overcommitment

Aggregate overcommitment provides flexibility to the storage provider. Using aggregate overcommitment, you can appear to provide more storage than is actually available from a given aggregate. This could be useful if you are asked to provide greater amounts of storage than you know will be used immediately. Alternatively, if you have several volumes that sometimes need to grow temporarily, the volumes can dynamically share the available space with each other.

To use aggregate overcommitment, you create FlexVol volumes with a space guarantee of `none` or `file`. With a space guarantee of `none` or `file`, 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 LUNs are created or data is appended to files in the volumes.

Of course, when the aggregate is overcommitted, it is possible for these types of writes to fail due to lack of available space:

- Writes to any volume with space guarantee of `none`
- Writes to any file that does not have space reservations enabled and that is in a volume with space guarantee of `file`

Therefore, 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.

Note

Because out-of-space errors are unexpected in a CIFS environment, do not set space guarantee to `none` for volumes accessed using CIFS.

Bringing a volume online in an overcommitted aggregate

When you take a FlexVol volume offline, it releases its allocation of storage space in its containing aggregate. Storage allocation for other volumes in that aggregate while that volume is offline can result in that storage being used. When you bring the volume back online, if there is insufficient space in the aggregate to fulfill the space guarantee of that volume, the normal `online` command fails unless you force the volume online by using the `-f` flag.

Attention

When you force a FlexVol volume to come online due to insufficient space, the space guarantees for that volume are disabled. That means that attempts to write to that volume could fail due to insufficient available space. In environments that are sensitive to that error, such as CIFS or LUNs, forcing a volume online should be avoided if possible.

When you make sufficient space available to the aggregate, the space guarantees for the volume are automatically re-enabled.

Note

FlexCache volumes cannot be brought online if there is insufficient space in the aggregate to fulfill their space guarantee.

For more information about FlexCache volumes, see “Managing FlexCache volumes” on page 268.

To bring a FlexVol volume online when there is insufficient storage space to fulfill its space guarantees, complete the following step.

Step	Action
1	Enter the following command: **`vol online vol_name -f`** *vol_name* is the name of the volume you want to force online.

Space reservations

What space reservations are

When space reservation is enabled for one or more files, Data ONTAP reserves enough space in the volume (traditional or FlexVol) so that writes to those files do not fail because of a lack of disk space. Other operations, such as snapshots or the creation of new files, can occur only if there is enough available unreserved space; these operations are restricted from using reserved space.

Writes to new or existing unreserved space in the volume fail when the total amount of available space in the volume is less than the amount set aside by the current file reserve values. Once available space in a volume goes below this value, only writes to files with reserved space are guaranteed to succeed.

File space reservation is an attribute of the file; it is persistent across storage system reboots, takeovers, and givebacks.

There is no way to automatically enable space reservations for every file in a given volume, as you could with versions of Data ONTAP earlier than 7.0 using the `create_reserved` option. In Data ONTAP 7.0, to guarantee that writes to a specific file will always succeed, you need to enable space reservations for that file. (LUNs have space reservations enabled by default.)

Note

For more information about using space reservation for files, see your *Block Access Management Guide*, keeping in mind that Data ONTAP manages files exactly the same as LUNs, except that space reservations are enabled automatically for LUNs, whereas for files, you must explicitly enable space reservations.

Enabling space reservation for a specific file

To enable space reservation for a file, complete the following step.

Step	Action
1	Enter the following command: **`file reservation file_name [enable|disable]`** *file_name* is the file in which file space reservation is set. `enable` turns space reservation on for the file *file_name.* `disable` turns space reservation off for the file *file_name*. **Example:** `file reservation myfile enable` **Note** In FlexVol volumes, the volume option `guarantee` must be set to `file` or `volume` for file space reservations to work. For more information, see "Space guarantees" on page 286.

Turning on space reservation for a file fails if there is not enough available space for the new reservation.

Querying space reservation for files

To find out the status of space reservation for files in a volume, complete the following step.

Step	Action
1	Enter the following command: **`file reservation file_name`** *file_name* is the file you want to query the space reservation status for. **Example:** `file reservation myfile` **Result:** The space reservation status for the specified file is displayed: `space reservations for file /vol/flex1/1gfile: off`

Fractional reserve

Fractional reserve

If you have enabled space reservation for a file or files, you can reduce the space that you preallocate for those reservations using fractional reserve. Fractional reserve is an option on the volume, and it can be used with either traditional or FlexVol volumes. Setting fractional reserve to less than 100 causes the space reservation held for all space-reserved files in that volume to be reduced to that percentage. Writes to the space-reserved files are no longer unequivocally guaranteed; you must monitor your reserved space and take action if your free space becomes scarce.

Fractional reserve is generally used for volumes that hold LUNs with a small percentage of data overwrite.

Note

If you are using fractional reserve in environments where write errors due to lack of available space are unexpected, you must monitor your free space and take corrective action to avoid write errors.

For more information about fractional reserve, see your *Block Access Management Guide*.

Space management policies

What space management policies are

Space management policies enable you to automatically reclaim space for a flexible volume when that volume is nearly full. You can configure a flexible volume to automatically reclaim space by using the following policies:

- Grow a flexible volume automatically when it is nearly full.

  This policy is useful if the containing aggregate has enough space to grow the flexible volume. You can grow a volume in increments and set a maximum size for the volume.

- Automatically delete snapshots when the flexible volume is nearly full.

  For example, you can automatically delete snapshots that are not linked to snapshots in cloned volumes or LUNs, or you can define which snapshots you want to delete first—your oldest or newest snapshots. You can also determine when to begin deleting snapshots—for example, when the volume is nearly full or when the volume's snapshot reserve is nearly full.

You can define the order in which you want to apply these policies when a flexible volume is running out of space. For example, you can automatically grow the volume first, and then begin deleting snapshots, or you can reclaim space by first automatically deleting snapshots, and then growing the volume.

For more information about determining and configuring your space management policy, see your *Block Access Management Guide*.

Qtree Management 7

About this chapter

This chapter describes how to use qtrees to manage user data. Read this chapter if you plan to organize user data into smaller units (qtrees) for flexibility or in order to use tree quotas.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding qtrees" on page 298
- "Understanding qtree creation" on page 300
- "Creating qtrees" on page 302
- "Understanding security styles" on page 303
- "Changing security styles" on page 306
- "Changing the CIFS oplocks setting" on page 308
- "Displaying qtree status" on page 311
- "Displaying qtree access statistics" on page 312
- "Converting a directory to a qtree" on page 313
- "Renaming or deleting qtrees" on page 316

Additional qtree operations are described in other chapters or other guides:

- For information about setting usage quotas for users, groups, or qtrees, see the chapter titled "Quota Management" on page 319.
- For information about configuring and managing qtree-based SnapMirror replication, see the *Data Protection Online Backup and Recovery Guide*.

Understanding qtrees

What qtrees are

A qtree is a logically defined file system that can exist as a special subdirectory of the root directory within either a traditional volume or a FlexVol volume.

Note

You can have a maximum of 4,995 qtrees on any volume.

When creating qtrees is appropriate

You might create a qtree for either or both of the following reasons:

- You can easily create qtrees for managing and partitioning your data within the volume.
- You can create a qtree to assign user- or workgroup-based soft or hard usage quotas to limit the amount of storage space that a specified user or group of users can consume on the qtree to which they have access.

Qtrees and volumes comparison

In general, qtrees are similar to volumes. However, they have the following key differences:

- Snapshots can be enabled or disabled for individual volumes, but not for individual qtrees.
- Qtrees do not support space reservations or space guarantees.

Qtrees, traditional volumes, and FlexVol volumes have other differences and similarities as shown in the following table.

Function	Traditional volume	FlexVol volume	Qtree
Enables organizing user data	Yes	Yes	Yes
Enables grouping users with similar needs	Yes	Yes	Yes
Can assign a security style to determine whether files use UNIX or Windows NT permissions.	Yes	Yes	Yes

Function	Traditional volume	FlexVol volume	Qtree
Can configure the oplocks setting to determine whether files and directories use CIFS opportunistic locks.	Yes	Yes	Yes
Can be used as units of SnapMirror backup and restore operations	Yes	Yes	Yes
Can be used as units of SnapVault backup and restore operations	No	No	Yes
Easily expandable and shrinkable	No (expandable but not shrinkable)	Yes	Yes
Snapshots	Yes	Yes	No (qtree replication extractable from volume snapshots)
Manage user based quotas	Yes	Yes	Yes
Cloneable	No	Yes	No (but can be part of a FlexClone volume)

Understanding qtree creation

Qtree grouping criteria

You create qtrees when you want to group files without creating a volume. You can group files by any combination of the following criteria:

- Security style
- Oplocks setting
- Quota limit
- Backup unit

Using qtrees for projects

One way to group files is to set up a qtree for a project, such as one maintaining a database. Setting up a qtree for a project provides you with the following capabilities:

- Set the security style of the project without affecting the security style of other projects.

  For example, you use NTFS-style security if the members of the project use Windows files and applications. Another project in another qtree can use UNIX files and applications, and a third project can use Windows as well as UNIX files.
- If the project uses Windows, set CIFS oplocks (opportunistic locks) as appropriate to the project, without affecting other projects.

  For example, if one project uses a database that requires no CIFS oplocks, you can set CIFS oplocks to Off on that project qtree. If another project uses CIFS oplocks, it can be in another qtree that has oplocks set to On.
- Use quotas to limit the disk space and number of files available to a project qtree so that the project does not use up resources that other projects and users need. For instructions about managing disk space by using quotas, see Chapter 8, "Quota Management," on page 319.
- Back up and restore all the project files as a unit.

Using qtrees for backups

You can back up individual qtrees to

- Add flexibility to backup schedules
- Modularize backups by backing up only one set of qtrees at a time
- Limit the size of each backup to one tape

Detailed information

Creating a qtree involves the activities described in the following topics:

- "Creating qtrees" on page 302
- "Understanding security styles" on page 303

If you do not want to accept the default security style of a volume or a qtree, you can change it, as described in "Changing security styles" on page 306.

If you do not want to accept the default CIFS oplocks setting of a volume or a qtree, you can change it, as described in "Changing the CIFS oplocks setting" on page 308.

Creating qtrees

Creating a qtree

To create a qtree, complete the following step.

Step	Action
1	Enter the following command: **`qtree create [-m mode] path`** *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. ◆ If you want to create the qtree in a volume other than the root volume, include the volume in the name. ◆ If *path* does not begin with a slash (/), the qtree is created in the root volume. **Note** Qtree names can be up to 64 characters long.

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 -m 770 /vol/users/news
```

The following command creates the news qtree in the root volume:

```
qtree create news
```

Understanding security styles

About security styles

Every qtree and volume has a security style setting. This setting determines whether files in that qtree or volume can use Windows NT or UNIX security.

Note

Although security styles can be applied to both qtrees and volumes, they are not shown as a volume attribute, and are managed for both volumes and qtrees using the `qtree` command.

Security styles

Three security styles apply to qtrees and volumes. They are described in the following table.

Security style	Description	Effect of changing to the style
NTFS	For CIFS clients, security is handled using Windows NTFS ACLs. For NFS clients, the NFS UID (user id) is mapped to a Windows SID (security identifier) and its associated groups. Those mapped credentials are used to determine file access, based on the NFTS ACL. **Note** To use NTFS security, the storage system must be licensed for CIFS. You cannot use an NFS client to change file or directory permissions on qtrees with the NTFS security style.	If the change is from a mixed qtree, Windows NT permissions determine file access for a file that had Windows NT permissions. Otherwise, UNIX-style (NFS) permission bits determine file access for files created before the change. **Note** If the change is from a CIFS storage system to a multiprotocol storage system, and the /etc directory is a qtree, its security style changes to NTFS.
UNIX	Exactly like UNIX; files and directories have UNIX permissions.	The storage system disregards any Windows NT permissions established previously and uses the UNIX permissions exclusively.

Security style	Description	Effect of changing to the style
Mixed	Both NTFS and UNIX security are allowed: A file or directory can have either Windows NT permissions or UNIX permissions. The default security style of a file is the style most recently used to set permissions on that file.	If NTFS permissions on a file are changed, the storage system recomputes UNIX permissions on that file. If UNIX permissions or ownership on a file are changed, the storage system deletes any NTFS permissions on that file.

Note

When you create an NTFS qtree or change a qtree to NTFS, every Windows user is given full access to the qtree, by default. You must change the permissions if you want to restrict access to the qtree for some users. If you do not set NTFS file security on a file, UNIX permissions are enforced.

For more information about file access and permissions, see the *File Access and Protocols Management Guide*.

Changing security styles

When to change the security style of a qtree or volume

There are many circumstances in which you might want to change qtree or volume security style. Two examples are as follows:

- You might want to change the security style of a qtree after creating it to match the needs of the users of the qtree.
- You might want to change the security style to accommodate other users or files. For example, if you start with an NTFS qtree and subsequently want to include UNIX files and users, you might want to change the qtree from an NTFS qtree to a mixed qtree.

Effects of changing the security style on quotas

Changing the security style of a qtree or volume requires quota reinitialization if quotas are in effect. For information about how changing the security style affects quota calculation, see "Turning quota message logging on or off" on page 358.

Changing the security style of a qtree

To change the security style of a qtree or volume, complete the following steps.

Step	Action
1	Enter the following command: **`qtree security [`*`path`*` {unix | ntfs | mixed}]`** *path* is the path name of the qtree or volume. Use `unix` for a UNIX qtree. Use `ntfs` for an NTFS qtree. Use `mixed` for a qtree with both UNIX and NTFS files.

Step	Action
2	If you have quotas in effect on the qtree whose security style you just changed, reinitialize quotas on the volume containing this qtree. **Result:** This allows Data ONTAP to recalculate the quota usage for users who own files with ACL or UNIX security on this qtree. For information about reinitializing quotas, see "Activating or reinitializing quotas" on page 350.

Attention

There are two changes to the security style of a qtree that you cannot perform while CIFS is running and users are connected to shares on that qtree: You cannot change UNIX security style to mixed or NTFS, and you cannot change NTFS or mixed security style to UNIX.

Example with a qtree: To change the security style of /vol/users/docs to be the same as that of Windows NT, use the following command:

```
qtree security /vol/users/docs ntfs
```

Example with a volume: To change the security style of the root directory of the users volume to mixed, so that, outside a qtree in the volume, one file can have NTFS security and another file can have UNIX security, use the following command:

```
qtree security /vol/users/ mixed
```

Changing the CIFS oplocks setting

What CIFS oplocks do

CIFS oplocks (opportunistic locks) enable the redirector on 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 work with a file (read or write it) without regularly reminding the server that it needs access to the file in question. This improves performance by reducing network traffic.

For more information on CIFS oplocks, see the CIFS section of the *File Access and Protocols Management Guide*.

When to turn CIFS oplocks off

CIFS oplocks on the storage system are on by default.

You might turn CIFS oplocks off on a volume or a qtree under either of the following circumstances:

- You are using a database application whose documentation recommends that CIFS oplocks be turned off.
- You are handling critical data and cannot afford even the slightest data loss.

Otherwise, you can leave CIFS oplocks on.

Effect of the cifs.oplocks.enable option

The `cifs.oplocks.enable` option enables and disables CIFS oplocks for the entire storage system.

Setting the `cifs.oplocks.enable` option has the following effects:

- If you set the `cifs.oplocks.enable` option to Off, all CIFS oplocks on all volumes and qtrees on the storage system are turned off.
- If you set the `cifs.oplocks.enable` option back to On, CIFS oplocks are enabled for the storage system, and the individual setting for each qtree and volume takes effect.

Enabling CIFS oplocks for a specific volume or qtree

To enable CIFS opslocks for a specific volume or a qtree, complete the following steps.

Step	Action
1	Make sure the global `cifs.oplocks.enable` option is set to On.
2	Enter the following command: **`qtree oplocks path enable`** *path* is the path name of the volume or the qtree.
3	To verify that CIFS oplocks were updated as expected, enter the following command: **`qtree status vol_name`** *vol_name* is the name of the specified volume, or the volume that contains the specified qtree.

Example: To enable CIFS oplocks on the proj1 qtree in vol2, use the following commands:

```
system1> options cifs.oplocks.enable on
system1> qtree oplocks /vol/vol2/proj enable
```

Disabling CIFS oplocks for a specific volume or qtree

To disable CIFS opslocks for a specific volume or a qtree, complete the following steps.

Attention

If you disable the CIFS oplocks feature on a volume or a qtree, any existing CIFS oplocks in the qtree will be broken.

Step	Action
1	Enter the following command: **`qtree oplocks path disable`** *path* is the path name of the volume or the qtree.

Step	Action
2	To verify that CIFS oplocks were updated as expected, enter the following command: **`qtree status vol_name`** *vol_name* is the name of the specified volume, or the volume that contains the specified qtree.

Example: To disable CIFS oplocks on the proj1 qtree in vol2, use the following command:

```
qtree oplocks /vol/vol2/proj disable
```

Displaying qtree status

Determining the status of qtrees

To find the security style, oplocks attribute, and SnapMirror status for all volumes and qtrees on the storage system or for a specified volume, complete the following step.

Step	Action
1	Enter the following command: **`qtree status [-i] [-v] [path]`** 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.

Example 1:

```
system> qtree status
  Volume       Tree        Style     Oplocks  Status
  --------     --------    -----     -------- ---------
  vol0                     unix      enabled  normal
  vol0         marketing   ntfs      enabled  normal
  vol1                     unix      enabled  normal
  vol1         engr        ntfs      disabled normal
  vol1         backup      unix      enabled  snapmirrored
```

Example 2:

```
system> qtree status -v vol1
Volume    Tree    Style  Oplocks  Status        Owning vfiler
--------  -----   -----  -------- ------        -------------
vol1              unix   enabled  normal        vfiler0
vol1      engr    ntfs   disabled normal        vfiler0
vol1      backup  unix   enabled  snapmirrored  vfiler0
```

Example 3:

```
system> qtree status -i vol1
Volume    Tree     Style    Oplocks    Status        ID
------    ----     -----    --------   ------------  ----
vol1               unix     enabled    normal        0
vol1      engr     ntfs     disabled   normal        1
vol1      backup   unix     enabled    snapmirrored  2
```

Displaying qtree access statistics

About qtree stats

The `qtree stats` command enables you to display statistics on user accesses to files in qtrees on your system. This can help you determine what qtrees are incurring the most traffic. Determining traffic patterns helps with qtree-based load balancing.

How the qtree stats command works

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.

Using qtree stats

To use the `qtree stats` command, complete the following step.

Step	Action
1	Enter the following command: **`qtree stats [-z] [`*`path`*`]`** The `-z` option clears the counter for the designated qtree, or clears all counters if no qtree is specified.

Example:

```
system> qtree stats vol1
Volume        Tree         NFS ops       CIFS ops
--------      --------     -------       --------
vol1          proj1           1232             23
vol1          proj2             55            312
```

Example with -z option:

```
system> qtree stats -z vol1
Volume        Tree         NFS ops       CIFS ops
--------      --------     -------       --------
vol1          proj1              0              0
vol1          proj2              0              0
```

Converting a directory to a qtree

Converting a rooted directory to a qtree

A rooted directory is a directory at the root of a volume. If you have a rooted directory 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. The following process outlines the tasks you need to complete to convert a rooted directory to a qtree:

Stage	Task
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.

Following are procedures showing how to complete this process on Windows clients and on UNIX clients.

Note
These procedures are not supported in the Windows command-line interface or at the DOS prompt.

Converting a rooted directory to a qtree using a Windows client

To convert a rooted directory to a qtree using a Windows client, complete the following steps.

Step	Action
1	Open Windows Explorer.
2	Click the folder representation of the directory you want to change.

Step	Action
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.
5	In Windows Explorer, open the renamed 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 a folder that you are moving across qtrees, the longer the move operation for that folder will take.
7	From the File menu, select Delete to delete the renamed, now-empty directory folder.

Converting a rooted directory to a qtree using a UNIX client

To convert a rooted directory to a qtree using a UNIX client, complete the following steps.

Step	Action
1	Open a UNIX 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`

Step	Action
4	From the client, use the `mv` command to move the contents of the old directory into the qtree. **Example:** `client: mv /n/joel/vol1/olddir/* /n/joel/vol1/dir1` **Note** Depending on how your UNIX client implements the `mv` command, storage system ownership and permissions may not be preserved. If this is the case for your UNIX client, you may need to update file owners and permissions after the mv command completes. The more subdirectories contained in a directory that you are moving across qtrees, the longer the move operation for that directory will take.
5	Use the `rmdir` command to delete the old, now-empty directory. **Example:** `client: rmdir /n/joel/vol1/olddir`

Renaming or deleting qtrees

Before renaming or deleting a qtree

Before you rename or delete a qtree, ensure that the following conditions are true:

- The volume that contains the qtree you want to rename or delete is mounted (for NFS) or mapped (for CIFS).
- The qtree you are renaming or deleting is not directly mounted and does not have a CIFS share directly associated with it.
- The qtree permissions allow you to modify the qtree.

Renaming a qtree

To rename a qtree, complete the following steps.

Step	Action
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, rename a qtree by using Windows Explorer. If you have quotas on the renamed qtree, update the /etc/quotas file to use the new qtree name.

Deleting a qtree

To delete a qtree, complete the following steps.

Step	Action
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, delete a qtree by using Windows Explorer. If you have quotas on the deleted qtree, remove the qtree from the /etc/quotas file.

Quota Management 8

About this chapter

This chapter describes how to restrict and track the disk space and number of files used by a user, group, or qtree.

Topics in this chapter

This chapter discusses the following topics:

- "Understanding quotas" on page 320
- "When quotas take effect" on page 323
- "Understanding default quotas" on page 324
- "Understanding derived quotas" on page 325
- "How Data ONTAP identifies users for quotas" on page 328
- "Notification when quotas are exceeded" on page 331
- "Understanding the /etc/quotas file" on page 332
- "Activating or reinitializing quotas" on page 350
- "Modifying quotas" on page 353
- "Deleting quotas" on page 356
- "Turning quota message logging on or off" on page 358
- "Effects of qtree changes on quotas" on page 360
- "Understanding quota reports" on page 362

For information about quotas and their effect in a client environment, see the *File Access and Protocols Management Guide*.

Understanding quotas

Reasons for specifying quotas

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 quota target
- To track the amount of disk space or the number of files used by a quota target, without imposing a limit
- To warn users when their disk space or file usage is high

Quota targets

A quota target can be

- A user, as represented by a UNIX ID or a Windows ID.
- A group, as represented by a UNIX group name or GID.

Note

Data ONTAP does not apply group quotas based on Windows IDs.

- A qtree, as represented by the path name to the qtree.

The quota target determines the quota type, as shown in the following table.

Quota target	Quota type
user	user quota
group	group quota
qtree	tree quota

Tree quotas

If you apply a tree quota to a qtree, the qtree is similar to a disk partition, except that you can change its size at any time. When applying a tree quota, Data ONTAP limits the disk space and number of files regardless of the owner of the disk space or files in the qtree. No users, including root and members of the BUILTIN\Administrators group, can write to the qtree if the write causes the tree quota to be exceeded.

Quota specifications

Quota specifications are stored in the /etc/quotas file, which you can edit at any time.

User and group quotas are applied on a per-volume or per-qtree basis. You cannot specify a single quota for an aggregate or for multiple volumes. For example, you can specify that a user named jsmith can use up to 10 GB of disk space in the cad volume, or that a group named engineering can use up to 50 GB of disk space in the /vol/cad/projects qtree.

Explicit quotas

If the quota specification references the name or ID of the quota target, the quota is an explicit quota. For example, if you specify a user name, jsmith, as the quota target, the quota is an explicit user quota. If you specify the path name of a qtree, /vol/cad/engineering, as the quota target, the quota is an explicit tree quota.

For examples of explicit quotas, see "Explicit quota examples" on page 342.

Default quotas and derived quotas

Even if you do not specify an explicit quota for it in the /etc/quotas file, you can create default quotas to restrict or track the disk space used by a quota target,. If a quota is applied to a target and the name or ID of the target does not appear in an /etc/quotas entry, the quota is called a derived quota.

For more information about default quotas, see "Understanding default quotas" on page 324. For more information about derived quotas, see "Understanding derived quotas" on page 325. For examples, see "Default quota examples" on page 342.

Hard quotas, soft quotas, and threshold quotas

A hard quota is a limit that cannot be exceeded. If an operation, such as a write, causes a quota target to exceed a hard quota, the operation fails. When this happens, a warning message is logged to the storage system console and an SNMP trap is issued.

A soft quota is a limit that can be exceeded. When a soft quota is exceeded, a warning message is logged to the storage system console and an SNMP trap is issued. When the soft quota limit is no longer being exceeded, another syslog message and SNMP trap are generated. You can specify both hard and soft quota limits for the amount of disk space used and the number of files created.

A threshold quota is similar to a soft quota. When a threshold quota is exceeded, a warning message is logged to the storage system console and an SNMP trap is issued.

A single type of SNMP trap is generated for all types of quota events. You can find details on SNMP traps in the storage system's /etc/mib/netapp.mib file.

Syslog messages about quotas contain qtree ID numbers rather than qtree names. You can correlate qtree names to the qtree ID numbers in syslog messages by using the `qtree status -i` command.

Tracking quotas

You can use tracking quotas to track, but not limit, the resources used by a particular user, group, or qtree. To see the resources used by that user, group, or qtree, you can use quota reports.

For examples of tracking quotas, see "Tracking quota examples" on page 342.

When quotas take effect

Prerequisite for quotas to take effect

You must activate quotas on a per-volume basis before Data ONTAP applies quotas to quota targets. For more information about activating quotas, see "Activating or reinitializing quotas" on page 350.

Note

Quota activation persists across halts and reboots. You should not activate quotas in the /etc/rc file.

About quota initialization

After you turn on quotas, Data ONTAP performs quota initialization. This involves scanning the entire file system in a volume and reading from the /etc/quotas file to compute the disk usage for each quota target.

Quota initialization is necessary under the following circumstances:

- You add an entry to the /etc/quotas file, but the quota target for that entry is not currently tracked by the storage system.
- You change user mapping in the /etc/usermap.cfg file and you use the QUOTA_PERFORM_USER_MAPPING entry in the /etc/quotas file. For more information about QUOTA_PERFORM_USER_MAPPING, see "Special entries for mapping users" on page 345.
- 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.

Quota initialization can take a few minutes. The amount of time required depends on the size of the file system. During quota initialization, data access is not affected. However, quotas are not enforced until initialization completes.

For more information about quota initialization, see "Activating or reinitializing quotas" on page 350.

About changing a quota size

You can change the size of a quota that is being enforced. Resizing an existing quota, whether it is an explicit quota specified in the /etc/quotas file or a derived quota, does not require quota initialization. For more information about changing the size of a quota, see "Modifying quotas" on page 353.

Understanding default quotas

About default quotas

You can create a default quota for users, groups, or qtrees. A default quota applies to quota targets that are not explicitly referenced in the /etc/quotas file. You create default quotas by using an asterisk (*) in the Quota Target field in the /etc/quota file. For more information about creating default quotas, see "Fields of the /etc/quotas file" on page 336 and "Tracking quota examples" on page 342.

How to override a default quota

If you do not want Data ONTAP to apply a default quota to a particular target, you can create an entry in the /etc/quotas file for that target. The explicit quota for that target overrides the default quota.

Where default quotas are applied

You apply a default user or group quota on a per-volume or per-qtree basis.

You apply a default tree quota on a per-volume basis. For example, you can specify that a default tree quota be applied to the cad volume, which means that all qtrees created in the cad volume are subject to this quota but that qtrees in other volumes are unaffected.

Typical default quota usage

As an example, suppose you want a user quota to be applied to most users of your system. Rather than applying that quota individually to every user, you can create a default user quota that will be automatically applied to every user. If you want to change that quota for a particular user, you can override the default quota for that user by creating an entry for that user in the /etc/quotas file.

For an example of a default quota, see "Tracking quota examples" on page 342.

About default tracking quotas

If you do not want to specify a default user, group or tree quota limit, you can specify default tracking quotas. These special default quotas do not enforce any resource limits, but they enable you to resize rather than reinitialize quotas after adding or deleting quota file entries.

Understanding derived quotas

About derived quotas

Data ONTAP derives the quota information from the default quota entry in the /etc/quotas file and applies it if a write request affects the disk space or number of files used by the quota target. A quota applied due to a default quota, not due to an explicit entry in the /etc/quotas file, is referred to as a derived quota.

Derived user quotas from a default user quota

When a default user quota is in effect, Data ONTAP applies derived quotas to all users in the volume or qtree to which the default quota applies, except those users who have explicit entries in the /etc/quotas file. Data ONTAP also tracks disk usage for the root user and BUILTIN\Administrators in that volume or qtree.

Example: A default user quota entry specifies that users in the cad volume are limited to 10 GB of disk space. Suppose a user named jsmith creates a file in that volume. Data ONTAP applies a derived quota to jsmith to limit that user's disk usage in the cad volume to 10 GB.

Derived group quotas from a default group quota

When a default group quota is in effect, Data ONTAP applies derived quotas for all UNIX groups in the volume or qtree to which the quota applies, except those groups that have explicit entries in the /etc/quotas file. Data ONTAP also tracks disk usage for the group with GID 0 in that volume or qtree.

Example: A default group quota entry specifies that groups in the cad volume are limited to 10 GB of disk space.m Suppose a file is created that is owned by a group named writers. Data ONTAP applies a derived quota to the writers group to limit its disk usage in the cad volume to 10 GB.

Derived tree quotas from a default tree quota

When a default tree quota is in effect, derived quotas apply to all qtrees in the volume to which the quota applies, except those qtrees that have explicit entries in the /etc/quotas file.

Example: A default tree quota entry specifies that qtrees in the cad volume are limited to 10 GB of disk space. Suppose a qtree named projects is created in the cad volume. Data ONTAP applies a derived quota to the cad projects qtree to limit its disk usage to 10 GB.

Default user or group quotas derived from default tree quotas

When a qtree is created in a volume that has a default tree quota defined in the /etc/quotas file, and that default quota is applied as a derived quota to the qtree just created, Data ONTAP also applies derived default user and group quotas to that qtree.

- If a default user quota or group quota is already defined for the volume containing the newly created qtree, Data ONTAP automatically applies that quota as the derived default user quota or group quota for that qtree.
- If no default user quota or group quota is defined for the volume containing the newly created qtree, then the effective derived user or group quota for that qtree is unlimited. In theory, a single user with no explicit user quota defined can use up the newly defined qtree's entire qtree quota allotment.
- You can replace the initial derived default user quotas or group quotas that Data ONTAP applies to the newly created qtree. To do so, you add explicit or default user or group quotas for the qtree just created to the /etc/quotas file.

Example of a default user quota for a volume applied to a qtree:

Suppose the default user quota in the cad volume specifies that each user is limited to 10 GB of disk space, and the default tree quota in the cad volume specifies that each qtree is limited to 100 GB of disk space. If you create a qtree named projects in the cad volume, a default tree quota limits the projects qtree to 100 GB. Data ONTAP also applies a derived default user quota, which limits to 10 GB the amount of space used by each user who does not have an explicit user quota defined in the /vol/cad/projects qtree.

You can change the limits on the default user quota for the /vol/cad/projects qtree or add an explicit quota for a user in the /vol/cad/projects qtree by using the `quota resize` command.

Example of no default user quota for a volume applied to a qtree:

If no default user quota is defined for the cad volume, and the default tree quota for the cad volume specifies that all qtrees are limited to 100 GB of disk space, and if you create a qtree named projects, Data ONTAP does not apply a derived default user quota that limits the amount of disk space that users can use on the /vol/cad/projects tree quota. In theory, a single user with no explicit user quota defined can use all 100 GB of a qtree's quota if no other user writes to disk space on the new qtree first.

In addition, UID 0, BUILTIN\Administrators, and GID 0 have derived quotas. These derived quotas do not limit the disk space and the number of files. They only track the disk space and the number of files owned by these IDs.

Even with no default user quota defined, no user with files on a qtree can use more disk space in that qtree than is allotted to that qtree as a whole.

Advantages of specifying default quotas

Specifying default quotas offers the following advantages:

- You can automatically apply a limit to a large set of quota targets without typing multiple entries in the /etc/quotas file. For example, if you want no user to use more than 10 GB of disk space, you can specify a default user quota of 10 GB of disk space instead of creating an entry in the /etc/quotas file for each user.
- You can be flexible in changing quota specifications. Because Data ONTAP already tracks disk and file usage for quota targets of derived quotas, you can change the specifications of these derived quotas without having to perform a full quota reinitialization.

  For example, you can create a default user quota for the vol1 volume that limits each user to 10 GB of disk space, and default tracking group and tree quotas for the cad volume. After quota initialization, these default quotas and their derived quotas go into effect.

  If you later decide that a user named jsmith should have a larger quota, you can add an /etc/quotas entry that limits jsmith to 20 GB of disk space, overriding the default 10-GB limit. After making the change to the /etc/quotas file, to make the jsmith entry effective, you can simply resize the quota, which takes less time than quota reinitialization.

  Without the default user, group and tree quotas, the newly created jsmith entry requires a full quota reinitialization to be effective.

How Data ONTAP identifies users for quotas

Two types of user IDs

When applying a user quota, Data ONTAP distinguishes one user from another based on the ID, which can be a UNIX ID or a Windows ID.

Format of a UNIX ID

If you want to apply user quotas to UNIX users, specify the UNIX ID of each user in one of the following formats:

- The user name, as defined in the /etc/passwd file or the NIS password map, such as jsmith.
- The UID, such as 20.
- A file or directory whose UID matches the user. In this case, you should choose a path name that will last as long as the user account remains on the system.

Note

Specifying a file or directory name only enables Data ONTAP to obtain the UID. Data ONTAP does not apply quotas to the file or directory, or to the volume in which the file or directory resides.

Restrictions on UNIX user names: A UNIX user name must not include a backslash (\) or an @ sign, because Data ONTAP treats names containing these characters as Windows names.

Special UID: You cannot impose restrictions on a user whose UID is 0. You can specify a quota only to track the disk space and number of files used by this UID.

Format of a Windows ID

If you want to apply user quotas to Windows users, specify the Windows ID of each user in one of the following formats:

- A Windows name specified in pre-Windows 2000 format. For details, see the section on specifying a Windows name in the CIFS chapter of the *File Access and Protocols Management Guide*.

  If the domain name or user name contains spaces or special characters, the entire Windows name must be in quotation marks, such as “tech support\john#smith”.
- A security ID (SID), as displayed by Windows in text form, such as S-1-5-32-544.

- A file or directory that has an ACL owned by the SID of the user. In this case, you should choose a path name that will last as long as the user account remains on the system.

  Note

  For Data ONTAP to obtain the SID from the ACL, the ACL must be valid.

  If a 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 matches that of the file or directory, not to the SID.

How Windows group IDs are treated

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 treated like a user quota.

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.

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 entry 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.

Example: 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

Quota targets listed in different quota entries are considered separate targets, even though the IDs belong to the same user.

Example: 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.

If the user has another Windows ID, engineering\jsmith, and there is no applicable quota entry (including a default quota), files owned by engineering\jsmith are not subject to restrictions, even though quota entries are in effect for UID 20 and corp\john_smith.

Root users and quotas

A root user is subject to tree quotas, but not user quotas or group quotas.

When root carries out a file or directory ownership change or other operation (such as the UNIX `chown` command) on behalf of a nonroot user, Data ONTAP checks the quotas based on the new owner but does not report errors or stop the operation even if the nonroot user's hard quota restrictions are exceeded. The root user can therefore carry out operations for a nonroot user (such as recovering data), even if those operations temporarily result in that nonroot user's quotas being exceeded.

Once the ownership transfer is carried out, however, a client system will report a disk space error for the nonroot user who is attempting to allocate more disk space while the quota is still exceeded.

Notification when quotas are exceeded

Console messages

When Data ONTAP receives a write request, it first determines whether the file to be written is in a qtree. If it is, and the write would exceed any hard quota, the write fails and a message is written to the console describing the type of quota exceeded and the volume. If the write would exceed any soft quota, the write succeeds, but a message is still written to the console.

SNMP notification

SNMP traps can be used to arrange e-mail notification when hard or soft quotas are exceeded. You can access and adapt a sample quota notification script on the NOW site at http://now.netapp.com/ under Software Downloads, in the Tools and Utilities section.

Understanding the /etc/quotas file

About this section

This section provides information about the /etc/quotas file so that you can specify user, group, or tree quotas.

Detailed information

This section discusses the following topics:

- "Overview of the /etc/quotas file" on page 333
- "Fields of the /etc/quotas file" on page 336
- "Sample quota entries" on page 342
- "Special entries for mapping users" on page 345
- "How disk space owned by default users is counted" on page 349

Overview of the /etc/quotas file

Contents of the /etc/quotas file

The /etc/quotas file consists of one or more entries, each entry specifying a default or explicit space or file quota limit for a qtree, group, or user.

The fields of a quota entry in the /etc/quotas file are

```
quota_target type[@/vol/dir/qtree_path] disk [files] [threshold]
[soft_disk] [soft_files]
```

The fields of an /etc/quotas file entry specify the following:

- *quota_target* specifies an explicit qtree, group, or user to which this quota is being applied. An asterisk (*) applies this quota as a default to all members of the type specified in this entry that do not have an explicit quota.
- *type* [*@/vol/dir/qtree_path*] 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 this user or group quota to a specific volume, directory, or qtree.
- *disk* is the disk space limit that this quota imposes on the qtree, group, user, or type in question.
- *files* (optional) is the limit on the number of files that this quota imposes on the qtree, group, or user in question.
- *threshold* (optional) is the disk space usage point at which warnings of approaching quota limits are issued.
- *soft_disk* (optional) is a soft quota space limit that, if exceeded, issues warnings rather than rejecting space requests.
- *soft_files* (optional) is a soft quota file limit that, if exceeded, issues warnings rather than rejecting file creation requests.

Note

For a detailed description of the above fields, see "Fields of the /etc/quotas file" on page 336.

Sample /etc/quotas file entries

The following sample quota entry assigns to user jsmith explicit limitations of 500 MB of disk space and 10,240 files in the rls volume and directory.

```
#Quota target  type           disk  files  thold  sdisk  sfile
#------------  ----           ----  -----  -----  -----  -----
jsmith         user@/vol/rls  500m  10k
```

The following sample quota entry assigns to groups in the cad volume a default quota of 750 megabytes of disk space and 85,000 files per group. This quota applies to any group in the cad volume that does not have an explicit quota defined.

```
#Quota target  type             disk  files  thold  sdisk  sfile
#------------  ----             ---   -----  ----   -----  -----
*              group@/vol/cad   750M  85K
```

Note

A line beginning with a pound sign (#) is considered a comment.

Each entry in the /etc/quotas file can extend to multiple lines, but the Files, Threshold, Soft Disk, and Soft Files fields must be on the same line as the Disk field. If they are not on the same line as the Disk field, they are ignored.

Order of entries

Entries in the /etc/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 /etc/quotas entries. If a quota target is affected by several /etc/quotas entries, the most restrictive entry applies.

Rules for a user or group quota

The following rules apply to a user or group quota:

- If you do not specify a path name to a volume or qtree to which the quota is applied, the quota takes effect in the root volume.
- You cannot impose restrictions on certain quota targets. For the following targets, you can specify quotas entries for tracking purposes only:
  - User with UID 0
  - Group with GID 0
  - BUILTIN\Administrators

- A file created by a member of the BUILTIN\Administrators group is owned by the BUILTIN\Administrators group, not by the member. When determining the amount of disk space or the number of files used by that user, Data ONTAP does not count the files that are owned by the BUILTIN\Administrators group.

Character coding of the /etc/quotas file

For information about character coding of the /etc/quotas file, see the *System Administration Guide*.

Fields of the /etc/quotas file

Quota Target field

The quota target specifies the user, group, or qtree to which you apply the quota. If the quota is a user or group quota, the same quota target can be in multiple /etc/quotas entries. If the quota is a tree quota, the quota target can be specified only once.

For a user quota: Data ONTAP applies a user quota to the user whose ID is specified in any format described in "How Data ONTAP identifies users for quotas" on page 328.

For a group quota: Data ONTAP applies a group quota to a GID, which you specify in the Quota Target field in any of these formats:

- The group name, such as publications
- The GID, such as 30
- A file or subdirectory whose GID matches the group, such as /vol/vol1/archive

Note

Specifying a file or directory name only enables Data ONTAP to obtain the GID. Data ONTAP does not apply quotas to that file or directory, or to the volume in which the file or directory resides.

For a tree quota: The quota target is the complete path name to an existing qtree (for example, /vol/vol0/home).

For default quotas: Use an asterisk (*) in the Quota Target field to specify a default quota. The quota is applied to the following users, groups, or qtrees:

- New users or groups that are created after the default entry takes effect. For example, if the maximum disk space for a default user quota is 500 MB, any new user can use up to 500 MB of disk space.
- Users or groups that are not explicitly mentioned in the /etc/quotas file. For example, if the maximum disk space for a default user quota is 500 MB, users for whom you have not specified a user quota in the /etc/quotas file can use up to 500 MB of disk space.

Type field

The Type field specifies the quota type, which can be

- 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.

For a user or group quota: The following table lists the possible values you can specify in the Type field, depending on the volume or the qtree to which the user or group quota is applied.

Quota type	Value in the Type field	Sample entry in the Type field
User quota in a volume	`user@/vol/`*`volume`*	`user@/vol/vol1`
User quota in a qtree	`user@/vol/`*`volume/qtree`*	`user@/vol/vol0/home`
Group quota in a volume	`group@/vol/`*`volume`*	`group@/vol/vol1`
Group quota in a qtree	`group@/vol/`*`volume/qtree`*	`group@/vol/vol0/home`

For a tree quota: The following table lists the values you can specify in the Type field, depending on whether the entry is an explicit tree quota or a default tree quota.

Entry	Value in the Type field
Explicit tree quota	`tree`
Default tree quota	`tree@/vol/`*`volume`* **Example:** `tree@/vol/vol0`

Disk field

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:

- K is equivalent to 1,024 bytes, M means 2^{20} bytes, and G means 2^{30} 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 16 TB, or
  - 16,383G
  - 16,777,215M
  - 17,179,869,180K

  Note

  If you omit the K, M, or G, Data ONTAP assumes a default value of K.

- 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.
- The value cannot be specified in decimal notation.
- If you want to track the disk usage but do not want to impose a hard limit on disk usage, type a hyphen (-).
- Do not 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 quota file entries as equivalent:

```
#Quota Target         type           disk      files
 /export              tree           75K
 /export              tree                     75K
```

Note

If you do not specify disk space limits as a multiple of 4 KB, disk space fields can appear incorrect in quota reports. This happens because disk space fields are always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.

Files field

The Files field specifies the maximum number of files 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:

- K is equivalent to 1,024, M means 2^{20}, and G means 2^{30}. 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 3GB, or
  - 4,294,967,295
  - 4,194,303K
  - 4,095M
  - 3G
- The value cannot be specified in decimal notation.
- If you want to track the number of files but do not want to impose a hard limit on the number of files that the quota target can use, type a hyphen (-). If the quota target is root, or if you specify 0 as the UID or GID, you must type a hyphen.
- A blank in this field means there is no restriction on the number of files that the quota target can use. If you leave this 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.

Threshold field

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. Use the Threshold field to specify disk space threshold limits for CIFS.

The following list describes the rules for specifying a value in this field:

- The use of K, M, and G for the Threshold field is the same as for the Disk field.

- The maximum value you can enter in the Threshold field is 16 TB, or
  - 16,383G
  - 16,777,215M
  - 17,179,869,180K

  Note

  If you omit the K, M, or G, Data ONTAP assumes the default value of K.

- The value cannot be specified in decimal notation.
- 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 limit on the amount of disk space the quota target can use, enter a hyphen (-) in this field or leave blank.

Note

Threshold fields can appear incorrect in quota reports if you do not specify threshold limits as multiples of 4 KB. This happens because threshold fields are always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.

Soft Disk field

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. When the soft disk limit is no longer being exceeded, another syslog message and SNMP trap are generated.

The following list describes the rules for specifying a value in this field:

- The use of K, M, and G for the Threshold field is the same as for the Disk field.
- The maximum value you can enter in the Soft Disk field is 16 TB, or
  - 16,383G
  - 16,777,215M
  - 17,179,869,180K
- The value cannot be specified in decimal notation.
- If you do not want to specify a soft limit on the amount of disk space that the quota target can use, type a hyphen (-) in this field (or leave this field blank if no value for the Soft Files field follows).

- The Soft Disk field must be on the same line as the Disk field. Otherwise, the Soft Disk field is ignored.

Note

Disk space fields can appear incorrect in quota reports if you do not specify disk space limits as multiples of 4 KB. This happens because disk space fields are always rounded up to the nearest multiple of 4 KB to match disk space limits, which are translated into 4-KB disk blocks.

Soft Files field

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. When the soft files limit is no longer being exceeded, another syslog message and SNMP trap are generated.

The following list describes the rules for specifying a value in this field.

- The format of the Soft Files field is the same as the format of the Files field.
- The maximum value you can enter in the Soft Files field is 4,294,967,295.
- The value cannot be specified in decimal notation.
- If you do not want to specify a soft limit on the number of files that the quota target can use, type a hyphen (-) in this field or leave the field blank.
- The Soft Files field must be on the same line as the Disk field. Otherwise, the Soft Files field is ignored.

Sample quota entries

Explicit quota examples

The following list contains examples of explicit quotas:

- `jsmith  user@/vol/rls  500M  10K`

  The user named jsmith can use 500 MB of disk space and 10,240 files in the rls volume.

- `jsmith,corp\jsmith,engineering\"john smith",`
  `S-1-5-32-544  user@/vol/rls  500M  10K`

  This user, represented by four IDs, can use 500 MB of disk space and 10,240 files in the rls volume.

- `writers  group@/vol/cad/proj1  150M`

  The writers group can use 150 MB of disk space and an unlimited number of files in the /vol/cad/proj1 qtree.

- `/vol/cad/proj1  tree  750M  75K`

  The proj1 qtree in the cad volume can use 750 MB of disk space and 76,800 files.

Tracking quota examples

The following list contains examples of tracking quotas:

- `root  user@/vol/rls  -  -`

  Data ONTAP tracks but does not limit the amount of disk space and the number of files in the rls volume owned by root.

- `builtin\administrators  user@/vol/rls  -  -`

  Data ONTAP tracks but does not limit the amount of disk space and the number of files in the rls volume owned by or created by members of BUILTIN\Administrators.

- `/vol/cad/proj1  tree  -  -`

  Data ONTAP tracks but does not limit the amount of disk space and the number of files for the proj1 qtree in the cad volume.

Default quota examples

The following list contains examples of default quotas:

- `*  user@/vol/cad  50M  15K`

  Any user not explicitly listed in the quota file can use 50 MB of disk space and 15,360 files in the cad volume.

- `*  group@/vol/cad  750M  85K`

  Any group not explicitly listed in the quota file can use 750 MB of disk space and 87,040 files in the cad volume.
- `*  tree@vol/cad  75M`

  Any qtree in the cad volume that is not explicitly listed in the quota file can use 75 MB of disk space and an unlimited number of files.

Default tracking quota example

Default tracking quotas enable you to create default quotas that do not enforce any resource limits. This is helpful when you want to use the `quota resize` command when you modify your /etc/quotas file, but you do not want to apply resource limits with your default quotas. Default tracking quotas are created per-volume, as shown in the following example:

```
#Quota Target    type            disk  files  thold  sdisk  sfile
#------------    ----            ----  -----  -----  -----  -----
*                user@/vol/vol1   -      -
*                group@/vol/vol1  -      -
*                tree@/vol/vol1   -      -
```

Sample quota file and explanation

The following sample /etc/quotas file contains default quotas and explicit quotas:

```
#Quota Target type                disk  files  thold  sdisk  sfile
#------------ ----                ----  -----  -----  -----  -----
*             user@/vol/cad       50M   15K
*             group@/vol/cad      750M  85K
*             tree@/vol/cad       100M  75K
jdoe          user@/vol/cad/proj1 100M  75K
msmith        user@/vol/cad       75M   75K
msmith        user@/vol/cad/proj1 75M   75K
```

The following list explains the effects of these /etc/quotas entries:

- Any user not otherwise mentioned in this file can use 50 MB of disk space and 15,360 files in the cad volume.
- Any group not otherwise mentioned in this file can use 750 MB of disk space and 87,040 files in the cad volume.
- Any qtree in the cad 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 cad volume (for example, a qtree named /vol/cad/proj2), Data ONTAP enforces a derived default user quota and a derived default group quota that have the same effect as these quota entries:

```
*                 user@/vol/cad/proj2     50M      15K
*                 group@/vol/cad/proj2    750M     85K
```

- If a qtree is created in the cad volume (for example, a qtree named /vol/cad/proj2), Data ONTAP tracks the disk space and number of files owned by UID 0 and GID 0 in the /vol/cad/proj2 qtree. This is due to this quota file entry:

```
*                 tree@/vol/cad           100M     75K
```

- A user named msmith can use 75 MB of disk space and 76,800 files in the cad 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.

  Quota oversubscription is supported; however, a warning is printed alerting you to the oversubscription.

How conflicting quotas are resolved

When more than one quota is in effect, the most restrictive quota is applied. Consider the following example /etc/quota file:

```
*                   tree@/vol/cad          100M     75K
jdoe                user@/vol/cad/proj1    750M     75K
```

Because the jdoe user has a disk quota of 750 MB in the proj1 qtree, you might expect that to be the limit applied in that qtree. But the proj1 qtree has a tree quota of 100 MB, because of the first line in the quota file. So jdoe will not be able to write more than 100 MB to the qtree. If other users have already written to the proj1 qtree, the limit would be reached even sooner.

To remedy this situation, you can create an explicit tree quota for the proj1 qtree, as shown in this example:

```
*                   tree@/vol/cad          100M     75K
/vol/cad/proj1      tree                   800M     75K
jdoe                user@/vol/cad/proj1    750M     75K
```

Now the jdoe user is no longer restricted by the default tree quota and can use the entire 750 MB of the user quota in the proj1 qtree.

Special entries for mapping users

Special entries in the /etc/quotas file

The /etc/quotas file supports two special entries whose formats are different from the entries described in "Fields of the /etc/quotas file" on page 336. These special entries enable you to quickly add Windows IDs to the /etc/quotas file. If you use these entries, you can avoid typing individual Windows IDs.

These special entries are

- QUOTA_TARGET_DOMAIN
- QUOTA_PERFORM_USER_MAPPING

Note

If you add or remove these entries from the /etc/quotas file, you must perform a full quota reinitialization for your changes to take effect. A `quota resize` command is not sufficient. For more information about quota reinitialization, see "Modifying quotas" on page 353.

Special entry for changing UNIX names to Windows names

The QUOTA_TARGET_DOMAIN entry enables you to change UNIX names to Windows names in the Quota Target field. Use this entry if both of the following conditions apply:

- The /etc/quotas file contains user quotas with UNIX names.
- The quota targets you want to change have identical UNIX and Windows names. For example, a user whose UNIX name is jsmith also has a Windows name of jsmith.

Format: The following is the format of the QUOTA_TARGET_DOMAIN entry:

```
QUOTA_TARGET_DOMAIN domain_name
```

Effect: For each user quota, Data ONTAP adds the specified domain name as a prefix to the user name. Data ONTAP stops adding the prefix when it reaches the end of the /etc/quotas file or another QUOTA_TARGET_DOMAIN entry without a domain name.

Example: The following example illustrates the use of the QUOTA_TARGET_DOMAIN entry:

```
QUOTA_TARGET_DOMAIN corp
roberts     user@/vol/rls          900M      30K
smith       user@/vol/rls          900M      30K
QUOTA_TARGET_DOMAIN engineering
daly        user@/vol/rls          900M      30K
thomas      user@/vol/rls          900M      30K
QUOTA_TARGET_DOMAIN
stevens     user@/vol/rls          900M      30K
```

Explanation of example: 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. The following entries produce the same effects:

```
corp\roberts        user@/vol/rls          900M      30K
corp\smith          user@/vol/rls          900M      30K
engineering\daly    user@/vol/rls          900M      30K
engineering\thomas  user@/vol/rls          900M      30K
stevens             user@/vol/rls          900M      30K
```

Special entry for mapping names

The QUOTA_PERFORM_USER_MAPPING entry enables you to map UNIX names to Windows names or vice versa. Use this entry if both of the following conditions apply:

- There is a one-to-one correspondence between UNIX names and Windows names.
- You want to apply the same quota to the user whether the user uses the UNIX name or the Windows name.

Note

The QUOTA_PERFORM_USER_MAPPING entry does not work if the QUOTA_TARGET_DOMAIN entry is present.

How names are mapped: Data ONTAP consults the /etc/usermap.cfg file to map the user names. For more information about how Data ONTAP uses the usermap.cfg file, see the *File Access and Protocols Management Guide*.

Format: The QUOTA_PERFORM_USER_MAPPING entry has the following format:

```
QUOTA_PERFORM_USER_MAPPING [on | off]
```

Data ONTAP maps the user names in the Quota Target fields of all entries following the `QUOTA_PERFORM_USER_MAPPING on` entry. It stops mapping when it reaches the end of the /etc/quotas file or when it reaches a `QUOTA_PERFORM_USER_MAPPING off` entry.

Note

If a default user quota entry is encountered after the QUOTA_PERFORM_USER_MAPPING directive, any user quotas derived from that default quota are also mapped.

Example: The following example illustrates the use of the QUOTA_PERFORM_USER_MAPPING entry:

```
QUOTA_PERFORM_USER_MAPPING on
roberts         user@/vol/rls          900M      30K
corp\stevens    user@/vol/rls          900M      30K
QUOTA_PERFORM_USER_MAPPING off
```

Explanation of example: If the /etc/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 a user with 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 a user with either user name is subject to the restriction of this quota entry.

The following entries produce the same effects:

```
roberts,corp\jroberts        user@/vol/rls          900M      30K
corp\stevens,cws             user@/vol/rls          900M      30K
```

Importance of one-to-one mapping: If the name mapping is not one-to-one, the QUOTA_PERFORM_USER_MAPPING entry produces confusing results, as illustrated in the following examples.

Example of multiple Windows names for one UNIX name: Suppose the /etc/usermap.cfg file contains the following entries:

```
domain1\user1 => unixuser1
domain2\user2 => unixuser1
```

Data ONTAP displays a warning message if the /etc/quotas file contains the following entries:

```
QUOTA_PERFORM_USER_MAPPING on
domain1\user1 user 1M
domain2\user2 user 1M
```

The /etc/quotas file effectively contains two entries for unixuser1. Therefore, the second entry is treated as a duplicate entry and is ignored.

Example of wildcard entries in usermap.cfg: Confusion can result if the following conditions exist:

- The /etc/usermap.cfg file contains the following entry:

  ```
  *\* *
  ```

- The /etc/quotas file contains the following entries:

  ```
  QUOTA_PERFORM_USER_MAPPING on
  unixuser2 user 1M
  ```

Problems arise because Data ONTAP tries to locate unixuser2 in one of the trusted domains. Because Data ONTAP searches domains in an unspecified order, unless the order is specified by the `cifs.search_domains` option, the result becomes unpredictable.

What to do after you change usermap.cfg: If you make changes to the /etc/usermap.cfg file, you must turn quotas off and then turn quotas back on for the changes to take effect. For more information about turning quotas on and off, see “Activating or reinitializing quotas” on page 350.

How disk space owned by default users is counted

Disk space used by the default UNIX user

For a Windows name that does not map to a specific UNIX name, Data ONTAP uses the default UNIX name defined by the `wafl.default_unix_user` option when calculating disk space. Files owned by the Windows user without a specific UNIX name are counted against the default UNIX user name if either of the following conditions applies:

- The files are in qtrees with UNIX security style.
- The files do not have ACLs in qtrees with mixed security style.

Disk space used by the default Windows user

For a UNIX name that does not map to a specific Windows name, Data ONTAP uses the default Windows name defined by the `wafl.default_nt_user` option when calculating disk space. Files owned by the UNIX user without a specific Windows name are counted against the default Windows user name if the files have ACLs in qtrees with NTFS security style or mixed security style.

Activating or reinitializing quotas

About activating or reinitializing quotas

You use the `quota on` command to activate or reinitialize quotas. The following list outlines some facts you should know about activating or reinitializing quotas:

- You activate or reinitialize quotas for only one volume at a time.
- Your /etc/quotas file does not need to be free of all errors to activate quotas. Invalid entries are reported and skipped. If the /etc/quotas file contains any valid entries, quotas are activated.
- Reinitialization causes the quota file to be scanned and all quotas for that volume to be recalculated.
- Changes to the /etc/quotas file do not take effect until either quotas are reinitialized or the `quota resize` command is issued.
- Quota reinitialization can take some time, during which storage system data is available, but quotas are not enforced for the specified volume.
- Quota reinitialization is performed asynchronously by default; other commands can be performed while the reinitialization is proceeding in the background.

  Note

  This means that errors or warnings from the reinitialization process could be interspersed with the output from other commands.

- Quota reinitialization can be invoked synchronously with the `-w` option; this is useful if you are reinitializing from a script.
- Errors and warnings from the reinitialization process are logged to the console as well as to /etc/messages.

Note

For more information about when to use the `quota resize` command versus the `quota on` command after changing the quota file, see "Modifying quotas" on page 353.

CIFS requirement for activating quotas

If the /etc/quotas file contains user quotas that use Windows IDs as targets, CIFS must be running before you can activate or reinitialize quotas.

Quota initialization terminated by upgrade

Any quota initialization running when the storage system is upgraded is terminated and must be manually restarted from the beginning. For this reason, you should allow any running quota initialization to complete before upgrading your storage system.

Activating quotas

To activate quotas, complete the following step.

Step	Action
1	Enter the following command: **`quota on [-w] vol_name`** The `-w` option causes the command to return only after the entire /etc/quotas file has been scanned (synchronous mode). This is useful when activating quotas from a script. **Example:** The following example turns on quotas on a volume named cad: `quota on cad`

Reinitializing quotas

To reinitialize quotas, complete the following steps.

Step	Action
1	If quotas are already on for the volume you want to reinitialize quotas on, enter the following command: **`quota off vol_name`**
2	Enter the following command: **`quota on vol_name`**

Deactivating quotas

To deactivate quotas, complete the following step.

Step	Action
1	Enter the following command: **`quota off vol_name`** **Example:** The following example turns off quotas on a volume named cad: `quota off cad` **Note** 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

To cancel a quota initialization that is in progress, complete the following step.

Step	Action
1	Enter the following command: **`quota off vol_name`** **Note** If a quota initialization is almost complete, the `quota off` command can fail. In this case, the initialization scan is already complete.

Modifying quotas

About modifying quotas

When you want to change how quotas are being tracked on your storage system, you first need to make the required change to your /etc/quota file. Then, you need to request Data ONTAP to read the /etc/quota file again and incorporate the changes. You can do this using one of the two following methods:

- Resize quotas

  Resizing quotas is faster than a full reinitialization; however, some quota file changes may not be reflected.

- Reinitialize quotas

  Performing a full quota reinitialization reads and recalculates the entire quota file. This may take some time, but all quota file changes are guaranteed to be reflected after the initialization is complete.

Note

Your storage system functions normally while quotas are being initialized; however, quotas remain off until the initialization is complete.

When you can use resizing

Because quota resizing is faster than quota initialization, you should use resizing whenever possible. You can use quota resizing for the following types of changes to the /etc/quota file:

- You changed an existing quota file entry, including adding or removing fields.
- You added a quota file entry for a quota target that was already covered by a default or default tracking quota.
- You deleted an entry from your /etc/quota file for which a default or default tracking quota entry is specified.

Note

After you have made extensive changes to the /etc/quota file, you should perform a full reinitialization to ensure that all of the changes become effective.

Resizing example 1: Consider the following sample /etc/quota file:

```
#Quota Target type                disk  files  thold  sdisk  sfile
#------------ ----                ----  -----  -----  -----  -----
*             user@/vol/cad        50M   15K
*             group@/vol/cad      750M   85K
*             tree@vol/cad          -     -
jdoe          user@/vol/cad/      100M   75K
kbuck         user@/vol/cad/      100M   75K
```

Suppose you make the following changes:

- Increase the number of files for the default user target.
- Added a new user quota for a new user that needs more than the default user quota.
- Deleted the kbuck user's explicit quota entry; the kbuck user now only needs the default quota limits.

These changes result in the following /etc/quota file:

```
#Quota Target type                disk  files  thold  sdisk  sfile
#------------ ----                ----  -----  -----  -----  -----
*             user@/vol/cad        50M   25K
*             group@/vol/cad      750M   85K
*             tree@vol/cad          -     -
jdoe          user@/vol/cad/      100M   75K
bambi         user@/vol/cad/      100M   75K
```

All of these changes can be made effective using the `quota resize` command; a full quota reinitialization is not necessary.

Resizing example 2: Your quotas file did not contain the default tracking tree quota, and you want to add a tree quota to the sample quota file, resulting in this /etc/quota file:

```
#Quota Target  type               disk  files  thold  sdisk  sfile
#------------  ----               ----  -----  -----  -----  -----
*              user@/vol/cad       50M   25K
*              group@/vol/cad     750M   85K
jdoe           user@/vol/cad/     100M   75K
bambi          user@/vol/cad/     100M   75K
/vol/cad/proj1 tree               500M  100K
```

In this case, using the `quota resize` command does *not* cause the newly added entry to be effective, because there is no default entry for tree quotas already in effect. A full quota initialization is required.

Note

If you use the `resize` command and the /etc/quota file contains changes that will not be reflected, 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. A quota in the quota report indicates that the storage system is tracking the disk space and the number of files owned by the quota target. For more information about quota reports, see "Understanding quota reports" on page 362.

Resizing quotas

To resize quotas, complete the following step.

Step	Action
1	Enter the following command: **`quota resize vol_name`** *vol_name* is the name of the volume you want to resize quotas for.

Deleting quotas

About quota deletion

You can remove quota restrictions for a quota target in two ways:

- Delete the /etc/quotas entry pertaining to the quota target.

  If you have a default or default tracking quota entry for the target type you deleted, you can use the `quota resize` command to update your quotas. Otherwise, you must reinitialize quotas.
- Change the /etc/quotas entry so that there is no restriction on the amount of disk space or the number of files owned by the quota target. After the change, Data ONTAP continues to keep track of the disk space and the number of files owned by the quota target but stops imposing the restrictions on the quota target. The procedure for removing quota restrictions in this way is the same as that for resizing an existing quota.

  You can use the `quota resize` command after making this kind of modification to the quotas file.

Deleting a quota by removing restrictions

To delete a quota by removing the resource restrictions for the specified target, complete the following steps.

Step	Action

1	Open the /etc/quotas file and edit the quotas file entry for the specified target so that the quota entry becomes a tracking quota.
	Example: Your quota file contains the following entry for the jdoe user:
	`jdoe  user@/vol/cad/  100M  75K`
	To remove the restrictions on jdoe, edit the entry as follows:
	`jdoe  user@/vol/cad/  -  -`

2	Enter the following command to update quotas: **`quota resize`** ***`vol_name`***

Deleting a quota by removing the quota file entry

To delete a quota by removing the quota file entry for the specified target, complete the following steps.

Step	Action	
1	Open the /etc/quotas file and remove the entry for the quota you want to delete.	
2	**If…**	**Then…**
	You have a default or default tracking quotas in place for users, groups and qtrees	Enter the following command to update quotas: **`quota resize vol_name`**
	Otherwise	Enter the following commands to reinitialize quotas: **`quota off vol_name`** **`quota on vol_name`**

Turning quota message logging on or off

About turning quota message logging on or off

You can turn quota message logging on or off for a single volume or for all volumes. You can optionally specify a time interval during which quota messages will not be logged.

Turning quota message logging on

To turn quota message logging on, complete the following step.

Step	Action
1	Enter the following command: **`quota logmsg on [`*`interval`*`] [-v `*`vol_name`*` | all]`** *interval* is the time period during which quota message logging is disabled. The interval is a number followed by `d`, `h`, or `m` for days, hours, and minutes, respectively. Quota messages are logged after the end of each interval. If no interval is specified, Data ONTAP logs quota messages every 60 minutes. For continuous logging, specify `0m` for the interval. `-v` *vol_name* specifies a volume name. `all` applies the interval to all volumes in the storage system. **Note** If you specify a short interval, less than five minutes, quota messages might not be logged exactly at the specified rate because Data ONTAP buffers quota messages before logging them.

Turning quota message logging off

To turn quota message logging off, complete the following step.

Step	Action
1	Enter the following command: **`quota logmsg off`**

Displaying settings for quota message logging

To display the current settings for quota message logging, complete the following step.

Step	Action
1	Enter the following command: `quota logmsg`

Effects of qtree changes on quotas

Effect of deleting a qtree on tree quotas

When you delete a qtree, all quotas that are applicable to that qtree, whether they are explicit or derived, are automatically deleted.

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. If a default tree quota exists, Data ONTAP creates new derived quotas for the new qtree. However, explicit quotas in the /etc/quotas file do not apply until you reinitialize quotas.

Effect of renaming a qtree on tree quotas

When you rename a qtree, Data ONTAP keeps the same ID for the tree. As a result, all quotas applicable to the qtree, whether they are explicit or derived, continue to be applicable.

Effects of changing qtree security style on user quota usages

Because ACLs apply in qtrees using NTFS or mixed security style but not in qtrees using UNIX security style, changing the security style of a qtree through the `qtree security` command might affect how a UNIX or Windows user's quota usages for that qtree are calculated.

Example: If NTFS security is in effect on qtree A and an ACL gives Windows user Windows/joe ownership of a 5-MB file, then user Windows/joe is charged 5 MB of quota usage on qtree A.

If the security style of qtree A is changed to UNIX, and Windows user Windows/joe is default mapped to UNIX user joe, the ACL that charged 5 MB of diskspace against the quota of Windows/joe is ignored when calculating the quota usage of UNIX user joe.

Attention

To make sure quota usages for both UNIX and Windows users are properly calculated after you use the `qtree security` command to change the security style, turn quotas for the volume containing that qtree off and then back on again using the `quota off` *vol_name* and `quota on` *vol_name* commands.

If you change the security style from UNIX to either mixed or NTFS, previously hidden ACLs become visible, any ACLs that were ignored become effective again, and the NFS user information is ignored. If no ACL existed before, then the NFS information is used in the quota calculation.

Note

Only UNIX group quotas apply to qtrees. Changing the security style of a qtree, therefore, does not affect the quota usages that groups are subject to.

Understanding quota reports

About this section

This section provides information about quota reports.

Detailed information

The following sections provide detailed information about quota reports:

- "Types of quota reports" on page 363
- "Overview of the quota report format" on page 364
- "Quota report formats" on page 366
- "Displaying a quota report" on page 370

Types of quota reports

Types of quota reports

You can display these types of quota reports:

- A quota report for all volumes that have quotas turned on. It contains the following types of information:
  - Default quota information, which is the same information as that in the /etc/quotas file
  - Current disk space and the number of files owned by a user, group, or qtree that has an explicit quota in the /etc/quotas file
  - Current disk space and the number of files owned by a user, group, or qtree that is the quota target of a derived quota, if the user, group, or qtree currently uses some disk space
- A quota report for a specified path name. It contains information about all the quotas that apply to the specified path name.

  For example, in the quota report for the /vol/cad/specs path name, you can see the quotas to which the disk space used by the /vol/cad/specs path name is charged. If a user quota exists for the owner of the /vol/cad/specs path name and a group quota exists for the cad volume, both quotas appear in the quota report.

Overview of the quota report format

Contents of the quota report

The following table lists the fields displayed in the quota report and the information they contain.

Heading	Information
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 (Disk field).
S-Limit	Maximum amount of disk space that can be used by the quota target before a warning is issued (Soft Disk field). This column is displayed only when you use the `-s` option for the `quota report` command.
T-hold	Disk space threshold (Threshold field). 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. If a soft files limit is specified for the quota target, you can also display the soft files limit in this field.

Heading	Information
Limit	Maximum number of files allowed for the quota target (Files field).
S-Limit	Maximum number of files that can be used by the quota target before a warning is issued (Soft Files field). This column is displayed only when you use the `-s` option for the `quota report` command.
VFiler	Displays the name of the vFiler for this quota entry. This column is displayed only when you use the `-v` option for the `quota report` command, which is available only on storage systems that have MultiStore licensed.
Quota Specifier	For an explicit quota, it shows how the quota target is specified in the /etc/quotas file. For a derived quota, the field is blank.

Quota report formats

Available report formats

Quota reports are available in these formats:

- A default format generated by the `quota report` command

  For more information, see "Default format" on page 367.
- Target IDs displayed in numeric form using the `quota report -q` command

  For more information, see "Report format with quota report -q" on page 368.
- Soft limits listed using the `quota report -s` command
- Threshold values listed using the `quota report -t` command
- vFiler names included using the `quota report -v` command

  This option is valid only if MultiStore is licensed.
- Two enhanced formats for quota targets with multiple IDs:
  - IDs listed on different lines using the `quota report -u` command

    For more information, see "Report format with quota report -u" on page 368.
  - IDs listed in a comma separated list using the `quota report -x` command

    For more information, see "Report format with quota report -x" on page 369.

Factors affecting the contents of the fields

The information contained in the ID and Quota Specifier fields can vary according to these factors:

- Type of user—UNIX or Windows—to which a quota applies
- The specific command used to generate the quota report

Contents of the ID field

In general, the ID field of the quota report displays a user name instead of a UID or SID; however, the following exceptions apply:

- For a quota with a UNIX user as the target, the ID field shows the UID instead of a name if no user name for the UID is found in the password database, or if you specifically request the UID by including the `-q` option in 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.
  - The storage system 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.

Default format

The `quota report` command without options generates the default format for the ID and Quota Specifier fields.

The ID field: If a quota target contains only one ID, the ID field displays that ID. Otherwise, the ID field displays one of the IDs from the list.

The ID field displays information in 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: The Quota Specifier field displays an ID that matches the one in the ID field. The ID is displayed the same way the quota target is specified in the /etc/quotas file.

Examples: The following table shows what is displayed in the ID and Quota Specifier fields based on the quota target in the /etc/quotas file.

Quota target in the /etc/quotas file	ID field of the quota report	Quota Specifier field of the quota report
CORP\john_smith	\john_sm	CORP\john_smith
CORP\john_smith,NT\js	\john_sm or \js	CORP\john_smith or NT\js
S-1-5-32-544	5-32-544	S-1-5-32-544

Report format with quota report -q

The `quota report -q` command displays the quota target's UNIX UID or GID in numeric form. Data ONTAP does not perform a lookup of the name associated with the target ID.

For Windows IDs, the textual form of the SID is displayed.

UNIX UIDs and GIDs are displayed as numbers. Windows SIDs are displayed as text.

Report format with quota report -s

The format of the report generated using the `quota report -s` command is the same as the default format, except that the soft limit columns are included.

Report format with quota report -t

The format of the report generated using the `quota report -t` command is the same as the default format, except that the threshold column is included.

Report format with quota report -v

The format of the report generated using the `quota report -v` command is the same as the default format, except that the vFiler column is included. This format is available only if MultiStore is licensed.

Report format with quota report -u

The `quota report -u` command is useful if you have quota targets that have multiple IDs. It provides more information in the ID and Quota Specifier fields than the default format.

If a quota target consists of multiple IDs, the first ID is listed on the first line of the quota report for that entry. The other IDs are listed on the lines following the first line, one ID per line. Each ID is followed by its original quota specifier, if any. Without this option, only one ID is displayed for quota targets with multiple IDs.

Note

You cannot combine the `-u` and `-x` options.

The ID field: 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 on a separate line in its entirety.

The Quota Specifier field: The Quota Specifier field displays the same list of IDs as specified in the quota target.

Example: The following table shows what is displayed in the ID and Quota Specifier fields based on the quota target in the /etc/quotas file. In this example, the SID maps to the user name NT\js.

Quota target in /etc/quotas	ID field of the quota report	Quota Specifier field of the quota report
CORP\john_smith,S-1-5-21-123456-7890-1234-1166	\john_sm NT\js	CORP\john_smith,S-1-5-21-123456-7890-1234-1166

Report format with quota report -x

The `quota report -x` command report format is similar to the report displayed by the `quota report -u` command, except that `quota report -x` displays all the quota target's IDs on the first line of that quota target's entry, as a comma separated list. The threshold column is included.

Note

You cannot combine the `-x` and `-u` options.

Displaying a quota report

Displaying a quota report for all quotas

To display a quota report for all quotas, complete the following step.

Step	Action
1	Enter the following command: **`quota report [-q] [-s] [-t] [-v] [-u|-x]`** For complete information on the `quota report` options, see "Quota report formats" on page 366.

Displaying a quota report for a specified path name

To display a quota report for a specified path name, complete the following step.

Step	Action
1	Enter the following command: **`quota report [-s] [-u | -x] [-t] [-q]`** ***`path_name`*** *path_name* is a complete path name to a file, directory, or volume, such as /vol/vol0/etc. For complete information on the `quota report` options, see "Quota report formats" on page 366.

Glossary

ACL Access control list. A list that contains the users' or groups' access rights to each share.

adapter card See host adapter.

aggregate A manageable unit of RAID-protected storage, consisting of one or two plexes, that can contain one traditional volume or multiple FlexVol volumes.

ATM Asynchronous transfer mode. A network technology that combines the features of cell-switching and multiplexing to offer reliable and efficient network services. ATM provides an interface between devices, such as workstations and routers, and the network.

authentication A security step performed by a domain controller for the storage system's domain, or by the storage system itself, using its /etc/passwd file.

AutoSupport A storage system daemon that triggers e-mail messages from the customer site to NetApp, or to another specified e-mail recipient, when there is a potential storage system problem.

CIFS Common Internet File System. A file-sharing protocol for networked PCs.

client A computer that shares files on a storage system.

cluster A pair of storage systems connected so that one storage system can detect when the other is not working and, if so, can serve the failed storage system's data. For more information about managing clusters, see the *Cluster Installation and Administration Guide*.

cluster interconnect Cables and adapters with which the two storage systems in a cluster are connected and over which heartbeat and WAFL log information are transmitted when both storage systems are running.

cluster monitor Software that administers the relationship of storage systems in the cluster through the `cf` command.

console A terminal that is attached to a storage system's serial port and is used to monitor and manage storage system operation.

continuous media scrub A background process that continuously scans for and scrubs media errors on the storage system disks.

DAFS Direct Access File System protocol.

degraded mode The operating mode of a storage system when a disk is missing from a RAID4 array, when one or two disks are missing from a RAID-DP array, or when the batteries on the NVRAM card are low.

disk ID number A number assigned by a storage system to each disk when it probes the disks at boot time.

disk sanitization A multiple write process for physically obliterating existing data on specified disks in such a manner that the obliterated data is no longer recoverable by known means of data recovery.

disk shelf A shelf that contains disk drives and is attached to a storage system.

Ethernet adapter An Ethernet interface card.

expansion card See host adapter.

expansion slot The slots on the system board into which you insert expansion cards.

GID Group identification number.

group A group of users defined in the storage system's /etc/group file.

host adapter (HA) A SCSI card, an FC-AL card, a network card, a serial adapter card, or a VGA adapter that plugs into a storage system expansion slot.

hot spare disk A disk installed in the storage system that can be used to substitute for a failed disk. Before the disk failure, the hot spare disk is not part of the RAID disk array.

hot swap The process of adding, removing, or replacing a disk while the storage system is running.

hot swap adapter An expansion card that makes it possible to add or remove a hard disk with minimal interruption to file system activity.

inode A data structure containing information about files on a storage system and in a UNIX file system.

mail host The client host responsible for sending automatic e-mail to NetApp when certain storage system events occur.

maintenance mode An option when booting a storage system from a system boot disk. Maintenance mode provides special commands for troubleshooting your hardware and your system configuration.

MultiStore An optional software product that enables you to partition the storage and network resources of a single storage system so that it appears as multiple storage systems on the network.

NVRAM cache Nonvolatile RAM in a storage system, used for logging incoming write data and NFS requests. Improves system performance and prevents loss of data in case of a storage system or power failure.

NVRAM card An adapter card that contains the storage system's NVRAM cache.

NVRAM mirror A synchronously updated copy of the contents of the storage system NVRAM (nonvolatile random access memory) kept on the partner storage system.

panic A serious error condition causing the storage system to halt. Similar to a software crash in the Windows system environment.

parity disk The disk on which parity information is stored for a RAID4 disk drive array. In RAID groups using RAID-DP protection, two parity disks store parity and double-parity information. Used to reconstruct data in failed disk blocks or on a failed disk.

PCI Peripheral Component Interconnect. The bus architecture used in newer storage system models.

pcnfsd A storage system daemon that permits PCs to mount storage system file systems. The corresponding PC client software is called (PC)NFS.

qtree A special subdirectory of the root of a volume that acts as a virtual subvolume with special attributes.

RAID Redundant array of independent disks. A technique that protects against disk failure by computing parity information based on the contents of all the disks in an array. Storage systems use either RAID Level 4, which stores all parity information on a single disk, or RAID-DP, which stores parity information on two disks.

RAID disk scrubbing The process in which a system reads each disk in the RAID group and tries to fix media errors by rewriting the data to another disk area.

SCSI adapter An expansion card that supports SCSI disk drives and tape drives.

SCSI address The full address of a disk, consisting of the disk's SCSI adapter number and the disk's SCSI ID, such as 9a.1.

SCSI ID The number of a disk drive on a SCSI chain (0 to 6).

serial adapter An expansion card for attaching a terminal as the console on some storage system models.

serial console An ASCII or ANSI terminal attached to a storage system's serial port. Used to monitor and manage storage system operations.

share A directory or directory structure on the storage system that has been made available to network users and can be mapped to a drive letter on a CIFS client.

SID Security identifier.

snapshot An online, read-only copy of an entire file system that protects against accidental deletions or modifications of files without duplicating file contents. Snapshots enable users to restore files and to back up the storage system to tape while the storage system is in use.

system board A printed circuit board that contains a storage system's CPU, expansion bus slots, and system memory.

trap An asynchronous, unsolicited message sent by an SNMP agent to an SNMP manager indicating that an event has occurred on the storage system.

tree quota A type of disk quota that restricts the disk usage of a directory created by the `quota qtree` command. Different from user and group quotas that restrict disk usage by files with a given UID or GID.

UID User identification number.

Unicode A 16-bit character set standard. It was designed and is maintained by the nonprofit consortium Unicode Inc.

vFiler A virtual storage system you create using MultiStore, which enables you to partition the storage and network resources of a single storage system so that it appears as multiple storage systems on the network.

VGA adapter Expansion card for attaching a VGA terminal as the console.

volume A file system.

WAFL Write Anywhere File Layout. The WAFL file system was designed for the storage system to optimize write performance.

WebDAV Web-based Distributed Authoring and Versioning protocol.

workgroup A collection of computers running Windows NT or Windows for Workgroups that is grouped for browsing and sharing.

WORM

Write Once Read Many. WORM storage prevents the data it contains from being updated or deleted. For more information about how Data ONTAP provides WORM storage, see the *Data Protection Online Recovery and Backup Guide*.

Index

Symbols

/etc/messages file 145, 146
/etc/messages, automatic checking of 145
/etc/quotas file
    character coding 335
    Disk field 338
    entries for mapping users 345
    example entries 334, 342
    file format 333
    Files field 339
    order of entries 334
    quota_perform_user_mapping 346
    quota_target_domain 345
    Soft Disk field 340
    Soft Files field 341
    Target field 336
    Threshold field 339
    Type field 337
/etc/sanitized_disks file 115

A

ACL 371
adapter. *See also* disk adapter and host adapter
aggr commands
    aggr copy 252
    aggr create 188
    aggr offline 195
    aggr online 196
    aggr restrict 196
aggregate and volume operations compared 36
aggregate overcommitment 289
aggregates
    adding disks to 36, 199, 201
    aggr0 24
    bringing online 196
    changing states of 37
    changing the RAID level of 152
    changing the size of 36
    copying 37, 196
    creating 29, 38, 188
    described 3, 14
    destroying 204, 206
    determining state of 194
    displaying as FlexVol container 40
    displaying disk space of 202
    hot spare disk planning 199
    how to use 14, 184
    maximum limit per storage system 26
    mirrored 4, 185
    new storage system configuration 24
    operations 36
    overcommitting 289
    physically moving between storage systems 208
    planning considerations 24
    RAID, changing level 152
    renaming 197
    restoring a destroyed aggregate 206
    rules for adding disks to 198
    states of 193
    taking offline 195
    taking offline, when to 194
    undestroy 206
    when to put in restricted state 196
ATM 371
automatic shutdown conditions 146
Autosupport message, about disk failure 146

B

backup
    planning considerations 27
    using qtrees for 300
    with snapshots 10
block checksum disks 2, 49

C

cache hit 272
checksum type 220
    block 49, 220
    rules 187
    zoned 49, 220
CIFS

commands, options cifs.oplocks.enable (enables and disables oplocks) 309
oplocks
changing the settings (options cifs.oplocks.enable) 309
definition of 308
setting for volumes 219, 227
setting in qtrees 300
clones *See* FlexClone volumes
cloning FlexVol volumes 231
commands
disk assign 61
options raid.reconstruct.perf_impact (modifies RAID data reconstruction speed) 162
options raid.reconstruct_speed (modifies RAID data reconstruction speed) 163, 169
options raid.resync.perf_impact (modifies RAID plex resynchronization speed) 164
options raid.scrub.duration (sets duration for disk scrubbing) 169
options raid.scrub.enable (enables and disables disk scrubbing) 169
options raid.verify.perf_impact (modifies RAID mirror verification speed) 165
See also aggr commands, qtree commands, quota commands, RAID commands, storage commands, volume commands
containing aggregate, displaying 40
continuous media scrub
adjusting maximum time for cycle 175
checking activity 177
data disks 179
description 175
disabling 175, 176
enabling on spare disks 177
spare disks 179
converting directories to qtrees 313
converting volumes 35
create_reserved option 292

D

data disks
removing 102
replacing 148
stopping replacement 148
Data ONTAP, upgrading 16, 19, 24, 27, 33, 35
data reconstruction
after disk failure 147
description of 162
data sanitization
planning considerations 25
See also disk sanitization
data storage, configuring 29
dd command, when not available 106
degraded mode 102, 146
deleting qtrees 316
destroying
aggregates 39, 204
FlexVol volumes 39
traditional volumes 39
volumes 39, 263
directories, converting to qtrees 313
directory size, setting maximum 41
disk
assign command
modifying 62
use on the FAS270 and 270c systems 61
commands
aggr show_space 202
aggr status -s (determines number of hot spare disks) 95
df (displayes free disk space) 94
df (reports discrepancies) 94
disk scrub (starts and stops disk scrubbing) 167
disk show 59
storage 124
sysconfig -d 86
displaying disk space usage on an aggregate 202
failure
reconstrution of single-disk failure 180
failures
data reconstruction after 147
predicting 144

RAID reconstruction after 145
without hot spare 146
ownership
automatically erasing information 65
erasing prior to removing disk 64
modifying assignments 62
software-based 58
undoing accidental conversion to 66
viewing 59
ownership assignment
description 58
modifying 62
sanitization
description 105
licensing 106
limitations 105
log files 115
selective data sanitization 110
starting 107
stopping 110
sanitization, easier on traditional volumes 33
scrubbing
description of 166
enabling and disabling (options raid.scrub.enable) 169
manually running it 170
modifying speed of 163, 169
scheduling 167
setting duration (options raid.scrub.duration) 169
starting/stopping (disk scrub) 167
toggling on and off 169
space, report of discrepancies (df) 94
swap command, cancelling 104
disk speed, overriding 189
disks
adding new to a storage system 98
adding to a RAID group other than the last RAID group 201
adding to a storage system 98
adding to an aggregate 199
adding to storage systems 98
addressing 50
assigning 60
assigning ownership of of FAS270 and FAS270c systems 58
available space on new 48
data, removing 102
data, stopping replacement 148
description of 13, 45
determining number of hot spares (sysconfig) 95
failed, removing 100
FC addressing 51
forcibly adding 201
hot spare, removing 101
hot spares, displaying number of 95
how initially configured 2
how to use 13
ownership of on FAS270 and FAS270c systems 58
portability 27
reasons to remove 100
removing 100
replacing
replacing data disks 148
re-using 63
rules for adding disks to an aggregate 198
software-based ownership 58
speed matching 188
viewing information about 88
when to add 97
double-disk failure
avoiding with media error thresholds 180
RAID-DP protection against 138
without hot spare disk 146
dumpblock, when not available 106
duplicate volume names 252

E

effects of oplocks 308

F

failed disk, removing 100
failure, data reconstruction after disk 147
FAS250 system, default RAID4 group size 157
FAS270 system, assigning disks to 61
FAS270c system, assigning disks to 61
FC disk addresses 51

Fibre Channel, Multipath I/O 69
file grouping, using qtrees 300
files
    as storage containers 18
    space reservation for 292
files, how used 12
FlexCache volumes
    about 269
    attribute cache timeouts 270
    cache consistency 270
    cache hits and misses 272
    cache objects 269
    creating 277
    description 268
    forward proxy deployment 275
    license requirement 269
    limitations of 272
    reverse proxy deployment 275
    sample deployments 275
    statistics, viewing 281
    volume options 271
    write operation proxy 272
FlexClone volumes
    creating 39, 231
    splitting 237
flexible volumes
    See FlexVol volumes
FlexVol volumes
    about creating 225
    bringing online in an overcommitted aggregate 290
    changing states of 37, 256
    changing the size of 36
    cloning 231
    co-existing with traditional 10
    copying 37
    creating 29, 38, 225
    defined 9
    definition of 212
    described 16
    displaying containing aggregate 242
    how to use 16
    migrating to traditional volumes 244
    operations 224
    resizing 229
    space guarantees, planning 27
fractional reserve, about 294

G

group quotas 320, 325

H

host adapter
    2202 70
    2212 70
    changing state of 132
    storage command 124
    viewing information about 126
hot spare disks
    displaying number of 95
    removing 101
hot swappable ESH controller modules 83
hub, viewing information about 127

I

inodes 265

L

language
    displaying its code 40
    setting for volumes 41
    specifying the character set for a volume 27
LUNs
    in a SAN environment 17
    with V-Series systems 18
LUNs, how used 11

M

maintenance center 117
maintenance mode 66, 195
maximum files per volume 265
media error failure thresholds 180
media scrub
    adjusting maximum time for cycle 175
    continuous 175
    continuous. *See also* continuous media scrub

disabling 176
displaying 40
migrating volumes with SnapMover 33
mirror verification, description of 165
mixed security style, description of 304
mode, degraded 102, 146
Multipath I/O
enabling 70
host adapters 70
preventing adapter single-point-of-failure 69
understanding 69

N

naming conventions for volumes 216, 225
NTFS security style, description of 304

O

oplocks
definition of 308
disabling 309
effects when enabled 308
enabling 309
enabling and disabling (options cifs.oplocks.enable) 309
setting for volumes 219, 227
options command, setting system automatic shutdown 146
overcommitting aggregates 289
overriding disk speed 189

P

parity disks, size of 199
permissions for new qtrees 302
physically transferring data 33
planning
for maximum storage 24
for RAID group level 25
for RAID group size 25
for SyncMirror replication 24
planning considerations 27
backup 27
data sanitization 25
FlexVol space guarantees 27
language 27
qtrees 27, 28
quotas 28
root volume sharing 25
SnapLock volume 25
traditional volumes 27
plex, synchronization 164
plexes
defined 3
described 14
how to use 14
snapshots of 10

Q

qtree commands
qtree create 302
qtree security (changes security style) 306
qtrees
changing security style 306
CIFS oplocks in 299
converting from directories 313
creating 33, 302
definition of 11
deleting 316
described 17, 298
displaying statistics 312
grouping criteria 300
grouping files 300
how to use 11, 17
maximum number 298
permissions for new 302
planning considerations 27, 28
quotas and changing security style 360
quotas and deleting 360
quotas and renaming 360
reasons for using in backups 300
reasons to create 298
renaming 316
security styles for 304
security styles, changing 306
stats command 312
status, determining 311
understanding 298
qtrees and volumes

changing security style in 306
comparison of 298
security styles available for 304
quota commands
quota logmsg (displays message logging settings) 359
quota logmsg (turns quota message logging on or off) 358
quota off (deactivates quotas) 352
quota off(deactivates quotas) 352
quota off/on (reinitializes quota) 351
quota on (activates quotas) 351
quota on (enables quotas) 351
quota report (displays report for quotas) 370
quota resize (resizes quota) 355
quota reports
contents 364
formats 366
ID and Quota Specifier fields 366
types 363
quota_perform_user_mapping 346
quota_target_domain 345
quotas
/etc/quotas file. *See* /etc/quotas file in the "Symbols" section of this index
/etc/rc file and 323
activating (quota on) 351
applying to multiple IDs 329
canceling initialization 352
changing 353
CIFS requirement for activating 350
conflicting, how resolved 344
console messages 331
deactivating 352
default
advantages of 327
description of 324
examples 342
overriding 324
scenario for use of 324
where applied 324
default UNIX name 349
default Windows name 349
deleting 356
derived 325
disabling (quota off) 352
Disk field 338
displaying report for (quota report) 370
enabling 351
example quotas file entries 334, 342
explicit quota examples 342
explicit, description of 321
Files field 339
group 320
group drived from tree 326
group quota rules 334
hard versus soft 321
initialization
canceling 352
description 323
message logging
display settings (quota logmsg) 359
turning on or off (quota logmsg) 358
modifying 353
notification when exceeded 331
order of entries in quotas file 334
overriding default 324
planning considerations 28
prerequisite for working 323
qtree
deletion and 360
renaming and 360
security style changes and 360
quota_perform_user_mapping 346
quota_taraget_domain 345
quotas file *See also* /etc/quotas file in the "Symbols" section of this index
reinitializing (quota on) 351
reinitializing versus resizing 353
reports
contents 364
formats 366
types 363
resizing 353, 355
resizing versus reinitializing 353
resolving conflicts 344
root users and 330
SNMP traps when exceeded 331
Soft Disk field 340
Soft Files field 341

soft versus hard 321
Target field 336
targets, description of 320
Threshold field 339
thresholds, description of 321, 339
tree 320
Type field 337
types of reports available, description of 363
types, description of 320
UNIX IDs in 328
UNIX names without Windows mapping 349
user and group, rules for 334
user derived from tree 326
user quota rules 334
Windows
group IDs in 329
IDs in 328
IDs, mapping 345
names without UNIX mapping 349

R

RAID
automatic group creation 138
changing from RAID4 to RAID-DP 152
changing from RAID-DP to RAID4 154
changing group size 157
changing RAID level 152
changing the group size option 158
commands
aggr create (specifies RAID group size) 149
aggr status 149
vol volume (changes RAID group size) 152, 158
data reconstruction speed, modifying (options raid.reconstruct.perf_impact) 162
data reconstruction speed, modifying (options raid.reconstruct_speed) 163, 169
data reconstruction, description of 162
description of 135
group size
changing (vol volume) 152, 158
comparison of larger versus smaller groups 142
default size 149
maximum 157
planning 25
specifying at creation (vol create) 149
group size changes
for RAID4 to RAID-DP 153
for RAID-DP to RAID4 154
groups
about 13
level, planning considerations 25
size, planning considerations 25
level
changing 152
descriptions of 136
verifying 156
maximum and default group sizes
RAID4 157
RAID-DP 157
media errors during reconstruction 174
mirror verification speed, modifying (options raid.verify.perf_impact) 165
operations
effects on performance 161
types you can control 161
options
setting for aggregates 42
setting for traditional volumes 42
parity checksums 2
plex resynchronization speed, modifying (options raid.resync.perf_impact) 164
reconstruction
media error encountered during 173
reconstruction of disk failure 145
status displayed 181
throttling data reconstruction 162
verifying RAID level 156
verifying the group size option 159
RAID groups
adding disks 201
RAID4
maximum and default group sizes 157
See also RAID
RAID-DP
maximum and default group sizes 157
See also RAID

RAID-level scrub performing
on aggregates 41
on traditional volumes 41
rapid RAID recovery 144
reallocation, running after adding disks for LUNs 203
reconstruction after disk failure, data 147
reliability, improving with MultiPath I/O 69
renaming
aggregates 41
flexible volumes 41
traditional volumes 41
volumes 41
renaming qtrees 316
resizing FlexVol volumes 229
restoring
with snapshots 10
restoring data with snapshots 298
restoring data, using qtrees for 300
root volume, setting 42
rooted directory 313

S

security styles
changing of, for volumes and qtrees 301, 306
for volumes and qtrees 303
mixed 304
NTFS 304
setting for volumes 219, 227
types available for qtrees and volumes 304
UNIX 304
setflag wafl_metadata_visible, when not available 106
SharedStorage
description of 77
displaying initiators in the community 82
how it works 78
hubs, benefits of 83
installing a community of 79
managing disks with 80
preventing disruption of service when downloading firmware 83
requiring Multipath I/O 71
requiring software-based disk ownership 58
requiring traditional volumes 33
requirments 79
supporing SyncMirror 79
using software-based disk ownership with 80
using vFiler no-copy migration 25
shutdown conditions 146
single-disk failure
without hot spare disk 137, 146
single-disk failure reconstruction 180
SnapLock
volumes
creating 39
planning considerations 25
SnapMirror software 10
SnapMover
described 58, 76
volume migration, easier with traditional volumes 33
snapshot 10
software-based disk ownership 58
space guarantees
about 286
changing 289
setting at volume creation time 288
space management
about 283
how to use 284
traditional volumes and 287
space management policy
applying 241
definition 295
space reservations
about 292
enabling for a file 293
querying 293
speed matching of disks 188
splitting FlexClone volumes 237
status
displaying aggregate 40
displaying FlexVol 40
displaying traditional volume 40
storage commands
changing state of host adapter 132
disable 132, 133
displaying information about

disks 88
primary and secondary paths 88
enable 132, 133
managing host adapters 124
reset tape drive statistics 131
viewing information about
host adapters 126
hubs 127
media changers 129
supported tape drives 130
switch ports 130
switches 129
tape drives 130
storage systems
adding disks to 98
automatic shutdown conditions 146
determining number of hot spare disks in (sysconfig) 95
running in degraded mode 146
when to add disks 97
storage, maximizing 24
swap disk command
cancelling 104
SyncMirror replica, creating 39, 42
SyncMirror replica, splitting 42
SyncMirror, planning for 24

T

thin provisioning. *See* aggregate overcommitment
traditional volumes
adding disks 36
changing states of 37, 256
changing the size of 36
copying 37
creating 33, 38, 216
definition of 16, 212
how to use 16
migrating to FlexVol volumes 244
operations 215
planning considerations, transporting disks 27
reasons to use 33
See also volumes
space management and 287
transporting 27
transporting between storage systems 221
upgrading to Data ONTAP 7.0 or later 27
transporting disks, planning considerations 27
tree quotas 320

U

undestroy an aggregate 206
UNICODE options, setting 42
UNIX security style, description of 304
uptime, improving with MultiPath I/O 69

V

volume and aggregate operations compared 36
volume commands
maxfiles (displays or increases number of files) 266, 288, 293
vol create (creates a volume) 190, 217, 225
vol create (specifies RAID group size) 149
vol destroy (destroys an off-line volume) 229, 233, 237, 242, 263
vol lang (changes volume language) 255
vol offline (takes a volume offline) 260
vol online (brings volume online) 261
vol rename (renames a volume) 262
vol restrict (puts volume in restricted state) 196, 261
vol status (displays volume language) 254
vol volume (changes RAID group size) 158
volume names, duplicate 252
volume operations 36, 213, 243
volume-level options, configuring 43
volumes
aggregates as storage for 7
as a data container 6
attributes 6, 26
bringing online 196, 261
bringing online in an overcommitted aggregate 290
cloning FlexVol 231
common attributes 15
conventions of 187
converting from one type to another 35
creating (vol create) 187, 190, 217, 225
creating FlexVol volumes 225

creating traditional 216
destroying (vol destroy) 229, 233, 237, 242, 263
destroying, reasons for 229, 263
displaying containing aggregate 242
duplicate volume names 252
flexible. *See* FlexVol volumes
growing automatically 240
how to use 15
increasing number of files (maxfiles) 266, 288, 293
language
changing (vol lang) 255
choosing of 253
displaying of (vol status) 254
planning 27
limits on number 213
maximum limit per storage system 26
maximum number of files 265
migrating between traditional and FlexVol 244
mirroring of, with SnapMirror 10
naming conventions 216, 225
number of files, displaying (maxfiles) 266
operations for FlexVol 224
operations for traditional 215
operations, general 243
post-creation changes 219, 227
renaming 262
renaming a volume (vol rename) 197
resizing FlexVol 229
restricting 261
root, planning considerations 25
root, setting 42
security style 219, 227
SnapLock, creating 39
SnapLock, planning considerations 25
specifying RAID group size (vol create) 149
taking offline (vol offline) 260
traditional. *See* traditional volumes
volume state, definition of 256
volume state, determining 259
volume status, definition of 256
volume status, determining 259
when to put in restricted state 260
volumes and qtrees
changing security style 306
comparison of 298
security styles available 304
volumes, traditional
co-existing with FlexVol volumes 10
V-Series LUNs 18
V-Series systems
and LUNs 11, 12

W

wafl.default_qtree_mode option 302

Z

zoned checksum disks 2, 49
zoned checksums 220